PLOS | COLLECTIONS
MEDICINE

www.ploscollections.org/measuringcoverageinmnch

# Measuring Coverage in Maternal, Newborn, and Child Health

Produced with support from the Child Health Epidemiology Reference Group (CHERG). Financial support for CHERG is provided by The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF.

## About this Collection

*Measuring Coverage in Maternal, Newborn and Child Health*, a *PLOS Medicine* Collection, presents innovative assessments of the validity of measuring population coverage for interventions in this field. Coverage indicators are widely used to assess whether interventions are reaching women and children in low and middle income countries, particularly through population-based household surveys. This collection of original research articles and reviews shows that while some indicators can be measured accurately, others may not provide valid results and therefore need further investigation and development.

The collection explores issues around measurement using household surveys, assessment of health inequalities, consideration of survey error, and the generation of indicators for global monitoring, and presents recommendations for transforming the findings of these and previous studies into better measurement, reporting and interpretation of coverage estimates.

Original research includes evaluating prevention of mother-to-child transmission of HIV in four African countries with estimates of community coverage for this indicator, and establishing whether mothers and caregivers can accurately recall diagnosis and treatment of childhood pneumonia in Pakistan and Bangladesh, and of malaria in Zambia. Other research looks at interventions delivered in pregnancy and around delivery in Africa, Latin America and China, including recall of emergency caesarean section, an important metric in maternity care. Review articles summarise the challenges of measuring coverage for a wide variety of indicators with validity and reliability, including those relating to immediate newborn care, childhood vaccinations, severity and treatment of diarrhea and community case management.

This collection was produced with support from the Child Health Epidemiology Reference Group (CHERG). Financial support for CHERG is provided by The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The Collection Editor is Dr. Lucy Chappell. The *PLOS Medicine* Editors have sole editorial responsibility for the content of this Collection.

http://www.ploscollections.org/measuringcoverageinmnch

# Table of Contents

## Overview



## Measuring Coverage - Maternal and Newborn Health



## Measuring Coverage - Child Health

# Measuring Coverage - Methodology

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: New Findings, New Strategies, and Recommendations for Action

**Jennifer Bryce[1]*, Fred Arnold[2], Ann Blanc[3], Attila Hancioglu[4], Holly Newby[4], Jennifer Requejo[1], Tessa Wardlaw[4], the CHERG Working Group on Improving Coverage Measurement[¶]**

**1** Institute for International Programs, Department of International Health, The Johns Hopkins University, Baltimore, Maryland, United States of America, **2** MEASURE DHS, ICF International, Calverton, Maryland, United States of America, **3** Population Council, New York, New York, United States of America, **4** Statistics and Monitoring Section, Division of Policy and Strategy, United Nations Children's Fund, New York, New York, United States of America

**Abstract:** Considerable progress has been made in reducing maternal, newborn, and child mortality worldwide, but many more deaths could be prevented if effective interventions were available to all who could benefit from them. Timely, high-quality measurements of intervention coverage—the proportion of a population in need of a health intervention that actually receives it—are essential to support sound decisions about progress and investments in women's and children's health. The *PLOS Medicine* "Measuring Coverage in MNCH" Collection of research studies and reviews presents systematic assessments of the validity of health intervention coverage measurement based on household surveys, the primary method for estimating population-level intervention coverage in low- and middle-income countries. In this overview of the Collection, we discuss how and why some of the indicators now being used to track intervention coverage may not provide fully reliable coverage measurements, and how a better understanding of the systematic and random error inherent in these coverage indicators can help in their interpretation and use. We draw together strategies proposed across the Collection for improving coverage measurement, and recommend continued support for high-quality household surveys at national and sub-national levels, supplemented by surveys with lighter tools that can be implemented every 1–2 years and by complementary health-facility-based assessments of service quality. Finally, we stress the importance of learning more about coverage measurement to strengthen the foundation for assessing and improving the progress of maternal, newborn, and child health programs.

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection.*

## Introduction

Despite dramatic declines in child mortality over the past decade, in 2011 there were about 7 million deaths among children less than five years of age [1]. Moreover, declines have been slower for newborn [1] and maternal deaths [2]. Effective interventions are available to prevent most of these deaths [3], but are not reaching all those who need them. Global monitoring of the proportion of women and children in need of these interventions who actually receive them (referred to here as "intervention coverage") shows both progress and missed opportunities to save lives [4]. Gaps in coverage are concentrated in poor countries, and within countries among the most vulnerable—the poorest and the least educated [5].

Monitoring of coverage levels for maternal, newborn, and child health (MNCH) interventions is central to assessing progress toward national and international health goals. Coverage data are also needed at national and sub-national levels to identify underserved populations and to monitor the effectiveness of strategies to reach them. The UN Secretary-General's global strategy, *Every Woman, Every Child*, calls for the scale-up of high-impact interventions, with oversight by an independent Expert Review Group. The independent Expert Review Group's 2012 report found that only 11 of the 75 countries that together account for over 95% of deaths among women and children had recent data on all eight coverage indicators recommended for global monitoring [6]. This finding reflects experience gained in tracking progress toward the Millennium Development Goals [7], and points to the challenges that must be overcome to improve the use of information for action. Examples of these challenges include obtaining adequate sample sizes for disaggregated analysis and reporting—and here new work on small area estimation techniques holds promise—and the need for temporally specific and recent measurements (i.e., data available within 12 months of collection).

Most data on intervention coverage in low- and middle-income countries are generated through the United States Agency for

**Citation:** Bryce J, Arnold F, Blanc A, Hancioglu A, Newby H, et al. (2013) Measuring Coverage in MNCH: New Findings, New Strategies, and Recommendations for Action. PLoS Med 10(5): e1001423. doi:10.1371/journal.pmed.1001423

**Academic Editor:** Lucy Chappell, Clinical Senior Lecturer in Maternal & Fetal Medicine and Honorary Consultant in Obstetrics, Women's Health Academic Centre, Kings College London, United Kingdom

**Published** May 7, 2013



**Funding:** This work was conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF, with financial support from The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** CHERG, Child Health Epidemiology Reference Group; DHS, Demographic and Health Survey/s; MICS, Multiple Indicator Cluster Survey/s; MNCH, maternal, newborn, and child health.

* E-mail: jbrycedanby@aol.com.

¶ Membership of the CHERG Working Group on Improving Coverage Measurement is provided in the Acknowledgments.

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

International Development-supported Demographic and Health Surveys (DHS) program [8], the United Nations Children's Fund-supported Multiple Indicator Cluster Surveys (MICS) program [9], and other national surveys modeled on these standards. DHS and MICS surveys send well-trained fieldworkers to interview preselected sample households (a probability sample in which each household has a known, non-zero chance of selection). Fieldworkers then conduct interviews with mothers and other caregivers about their health needs and interventions received as a basis for estimating coverage at the population level [10]. Household surveys—especially DHS and MICS surveys, which have refined their methods through more than 500 nationally representative surveys over more than two decades—are the primary tools available at present to track population-based trends in MNCH intervention coverage [11]. In most low- and middle-income countries, reports produced by routine health information systems include individuals in contact with the health system but miss those who are not and are often incomplete, late, or inaccurate. These reports do not, therefore, currently produce data adequate to support programmatic decisions, with occasional exceptions (e.g., in some countries with high vaccination coverage) [12]. Moreover, household surveys, unlike health information system reports, permit the analysis of coverage by equity variables such as gender, wealth, geography, and ethnicity.

Few studies have assessed the validity of coverage indicators for MNCH interventions measured through household surveys, or explored ways to improve their measurement. The aim of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection is to bring together cutting-edge scientific work in this area. The Collection includes the results of a three-year program of research and reviews carried out by a diverse group of experts working under the auspices of the Child Health Epidemiology Reference Group (CHERG) [13], as well as work supported by other groups.

The CHERG work on improving coverage measurement began in 2009, guided by a Core Group that includes the technical directors of DHS and MICS and others with expertise in measurement issues (the named authors of this paper). We defined the scope of work (Box 1) and assessed the evidence base on the validity of coverage indicators for each proven intervention along the continuum of care for MNCH tracked by the Countdown to 2015 for Maternal, Newborn and Child Survival [14]. We established collaborative relationships with groups addressing measurement issues for particular indicators. For indicators with specific definitional or measurement issues that could be readily improved, we addressed them directly with the DHS and MICS leadership, and in most cases were able to resolve the problem. We prioritized indicators for which we were unable to find evidence that validity had been assessed, and commissioned research studies to assess the validity of these indicators and to identify ways to improve their measurement. Papers on methodological issues (e.g., uncertainty and interpretation [15] and equity [16]) were also commissioned.

We adopted the definition of coverage presented above ("the proportion of women and children in need of interventions who actually receive them") in preference to measures of "effective coverage" that include estimates of intervention effectiveness, access, utilization, and service quality [17,18]. Composite "effective coverage" metrics can have limited usefulness in national or global monitoring because they usually require a great deal of data that are rarely available in low-income countries and because they sometimes rely on modeling that produces "black box" statistics that must be unpacked to guide program decisions. Indeed, perhaps for these reasons, composite "effective coverage" metrics have generally not been adopted for use in monitoring progress toward global goals, including the Millennium Development Goals.

**Box 1. Scope of the Collection**

The *PLOS Medicine* "Measuring Coverage in MNCH" Collection includes nine reviews, six research papers, and an overview of the measurement of population coverage for interventions of proven benefit to women's and children's health. It focuses on the following:

- Coverage measurement for interventions of proven efficacy that are feasible for implementation at scale in low- and middle-income countries to reduce maternal, newborn, and under-five mortality
- Global consensus coverage indicators as defined in the Countdown to 2015 for Maternal, Newborn and Child Survival
- Indicators that are or could be measured through population-based household surveys, with a specific focus on the core DHS and MICS surveys

The Core Group oversaw the commissioned research studies, conceived of the Collection as a means of disseminating the findings, and invited groups working in related areas to contribute. Authors of the papers (the CHERG Working Group on Improving Coverage Measurement) met in Baltimore in June 2012 to review draft manuscripts and to ensure that the Collection addressed key themes and was internally consistent. Each manuscript was reviewed by at least two members of the Working Group and the Technical Deputy Director of DHS (Fred Arnold) prior to submission for publication. The full Collection is available at http://www.ploscollections.org/measuringcoverageinmnch.

The Collection presents new evidence drawn from research studies and reviews of existing literature. In this overview we synthesize the findings of the Collection papers under three broad topics: (1) validity of coverage estimates based on respondents' reports, (2) potential strategies for improving coverage measurement through household surveys, and (3) crosscutting methodological issues in coverage measurement. We close with a set of action recommendations directed to those who conduct, support, or use household survey results on measurement of coverage for MNCH interventions.

## The Validity of Coverage Estimates Based on Respondents' Reports

At the core of the Collection is a set of research studies that assesses the validity of respondents' reports on interventions that they and the children under their care have received. The basic design of these studies was to first establish an accurate record of the need for interventions (e.g., standard or emergency obstetric services, treatment for pneumonia or malaria), and then to either conduct direct observations in health facilities or obtain high-quality health records (the "reference standard") to determine which interventions were provided to address this need. The next step involved visiting each woman in her community to administer survey questions using time intervals and question sequences similar to those in the DHS and MICS household surveys. These data were then used to estimate the sensitivity, specificity, and accuracy of the mothers' individual reports, and to determine whether indicator values would be overestimated or underestimated when measured in a household survey (the true/actual positives ratio or its mathematical equivalent) [19]. This design may produce results that are positively biased for interventions

that can be offered either at home or at a health facility because the samples are drawn from health service settings. However, even studies based on determinations of need conducted in purposively selected teaching hospitals offer important evidence about the research question on mothers' recall of health-related events.

We conducted studies of this type for emergency cesarean sections in Ghana and the Dominican Republic [20], a broad range of interventions delivered around the time of birth (the peripartum period) in Mozambique [21], diagnosis and antibiotic treatment of childhood pneumonia in Pakistan and Bangladesh [22], diagnosis and treatment of malaria in Zambia [23], and selected services across the MNCH continuum of care in a rural population in China [24]. Table 1 presents findings on sensitivity, specificity, and accuracy for a subset of the indicators assessed in the research studies, focusing on global consensus coverage indicators currently in use and indicators closely related to these accepted coverage indicators. These results are not strictly comparable across studies owing to differences in what is being measured, how it is being measured, and in what context. We believe for the same reason that applying standard cutoff values for acceptable levels of accuracy is not appropriate in this situation. True/actual positives ratios or their mathematical equivalents are presented in the individual papers [20,21,23,24].

The studies show that the sensitivity and specificity of coverage indicators are highly variable across interventions, with some suggestion of more accurate reporting for events related to care-seeking behaviors (e.g., place of delivery) or invasive interventions (e.g., finger/heel stick or cesarean section performed). Women reported less accurately about interventions that occurred immediately after childbirth, such as whether the newborn was dried [21], or for interventions requiring recognition of a complex disease syndrome such as pneumonia [22]. Also noteworthy are large differences in the sensitivity and specificity of individual indicators within and across countries, such as urban–rural differences in Bangladesh [22], and in the sensitivity of several indicators in different settings [20,22]. The findings of the studies must be generalized with caution, however, because the sensitivity and specificity of the indicators may vary by characteristics of the host and the pathogen, the broader epidemiological setting (e.g., low levels of falciparum malaria in Bangladesh and Pakistan), and cultural and educational differences in interpreting and responding to survey questions.

Another set of challenges is more structural, and relates to the difficulty of measuring coverage using respondents' reports for interventions that address relatively rare events. Even if mothers can report accurately, obtaining adequate denominators to support coverage measurement for low prevalence events requires very large samples—as does the determination of reasonably precise disaggregated estimates—and, for some indicators, requires attention to be paid to seasonality issues [25].

The research paper on using community surveys to estimate HIV survival [26] uses a different design, but makes an important contribution to the Collection as a whole. The authors demonstrate that in four African countries (Cameroon, Côte D'Ivoire, South Africa, and Zambia), health-facility-based estimates of coverage for regimens to prevent mother-to-child HIV transmission consistently tend to overestimate true population coverage measured in the community. The authors also show that the health-facility-based estimates do not correlate well with infant HIV-free survival. These results serve as a wake-up call to those advocating for coverage measurement based solely on routine health-facility-based information systems.

Based on the research findings in this Collection, we can make concrete recommendations about which indicators perform well in household surveys, about which perform poorly and may produce spurious results, and about whether there are new indicators that are good candidates for inclusion in future surveys. Indicators that performed well in this set of validation studies include place of delivery (hospital, health center, home), placing the newborn skin-to-skin against the mother's chest, emergency cesarean section, and treatment of childhood malaria with artemisinin-based combination therapies. Coverage indicators for antibiotic treatment of childhood pneumonia showed poor results [22,25], consistent with previous research indicating that lists of illness signs are poor predictors of actual disease [27–31]. For malaria, presumptive treatment of fever with antimalarials is no longer recommended; malaria treatment should be given only to children with a malaria diagnosis confirmed by microscopy or rapid diagnostic test, where possible. Ideally the global indicator for malaria treatment would reflect this policy and be limited to children with laboratory-confirmed disease, but the relatively poor accuracy of correct recall of a malaria diagnosis [23] currently precludes this change.

These research results are buttressed and extended by the findings from the Collection reviews of previous research on the accuracy of respondents' reports of services received. Many factors can affect the reliability of a verbal history of service provision, including the information received or understood at the time the intervention was delivered, interviewer behaviors, the recall period, the characteristics and salience of the intervention itself, and the length of the questionnaire and resulting interview fatigue [32–35]. Previous studies have shown, for example, that women have difficulty reporting on interventions provided during labor and in the first hours after birth, especially in surveys asking about these events up to two years after they occur [36]. Similarly, although an early household survey in Mozambique found that mothers' reports of symptoms of dehydration in the last 24 hours among children with diarrhea corresponded reasonably well with the diagnosis of dehydration by trained interviewers, the Collection review on intervention coverage for diarrhea identifies challenges in the measurement of diarrhea treatment coverage that arise from difficulties in standardizing definitions of diarrhea severity and a lack of clarity about the types and quantities of fluids required to treat diarrhea [37].

## Potential Strategies for Improving Coverage Measurement through Household Surveys

The papers in this Collection propose many ways in which household surveys might be improved to produce more valid estimates of intervention coverage. We review them briefly here and direct readers to the full papers for further explanation.

### Use Aides Mémoires to Improve Accurate Reporting

Hazir and colleagues in Pakistan and Bangladesh used expanded lists of clinical signs of pneumonia and video clips of children with specific clinical presentations in caregiver interviews to increase the accuracy of the denominator for the indicator on antibiotic treatment of pneumonia, and showed mothers a selection of locally available antibiotics to increase the accuracy of the numerator [22]. Their results were mixed. The video showed much higher sensitivity in Pakistan (where it was developed) than in Bangladesh (62% versus 28%), whereas the use of visual prompts for antibiotic treatment, a common practice in DHS and MICS surveys, was associated with increased accuracy of reports of treatment regimens in both settings. Implementation of this strategy may be challenging, however, in settings with an active private sector and many potential

**Table 1.** Selected Collection research findings on respondents' reports of intervention coverage.

| Category | Intervention (Selected) | Collection Research Paper | | Selected Findings | | |
|---|---|---|---|---|---|---|
| | | Total Study Sample[a] (Reference Standard) | Reference | Sensitivity (CI) | Specificity (CI) | Accuracy/AUC (CI) |
| **Antenatal care** | First antenatal care visit <12 weeks of gestational age | 914 women aged 18–49 in China with at least one live birth in the preceding five years (home-based booklets and electronic service records) | Guo, et al. [24] | 90% (86–94) | 22% (19–26) | 56% (54–59) |
| | At least four antenatal care visits for last pregnancy | Same as previous | | 98% (96–99) | 25% (19–32) | 62% (58–65) |
| **Interventions delivered around the time of birth** | Woman delivered in a hospital (versus a health center) | 304 women in Mozambique who gave birth 8–10 months previously in government facility (direct observation by trained clinician) | Stanton, et al. [21] | 81% (75–87) | 94% (90–98) | 88% (84–91) |
| | Newborn placed skin-to-skin on mother's chest | Same as previous | | 60% (52–69) | 69% (62–76) | 65% (59–70) |
| | Newborn immediately dried | Same as previous | | 77% (72–82) | 31% (12–50) | 54% (45–63) |
| | Mothers' recall of emergency cesarean section[a] | 659 women in Ghana delivered in a hospital via cesarean section (facility-based data supplemented by information requested from the medical staff) | Tunçalp, et al. [20] | 79% (73–83) | 82% (78–85) | 80% (77–83) |
| | | 1,531 women in the Dominican Republic delivered in a hospital via cesarean section (facility-based data supplemented by information requested from the medical staff) | | 50% (47–53) | 80% (77–83) | 65% (62–67) |
| | Any cesarean section | 914 women aged 18–49 in China with at least one live birth in the preceding five years (home-based booklets and electronic service records) | Guo, et al. [24] | 96% (93–99) | 83% (80–86) | 90% (88–92) |
| **Vaccination** | Diphtheria-tetanus-pertussis vaccine | 914 women aged 18–49 in China with at least one live birth in the preceding 5 years (home-based booklets and electronic service record) | Guo, et al. [24] | 89% (86–92) | 70% (61–78) | 80% (75–84) |
| | Measles vaccine | Same as previous | | 95% (92–98) | 44% (38–49) | 69% (66–72) |
| **Treatment of childhood illness** | Correct treatment of pneumonia (using DHS algorithms on symptoms of acute respiratory infection and detailed enquiry) | 672 caregivers of children 0–59 months diagnosed pneumonia or ''no pneumonia'' in the out-patient department of an urban hospital in Islamabad, Pakistan (direct observation) | Eisele, et al. [22] | 67% (62–72) | 69% (64–74) | 0.66 (0.62–0.69) |
| | | 700 caregivers of children 0–59 months diagnosed pneumonia or ''no pneumonia'' in the out-patient department of an urban hospital in Dhaka, Bangladesh (direct observation) | | 24% (19–30) | 82% (77–87) | 0.53(0.49–0.57) |
| | | 478 caregivers of children 0–59 months diagnosed pneumonia or ''no pneumonia'' in rural Mirzapur, Bangladesh (direct observation) | | 72% (65–78) | 55% (47–62) | 0.63 (0.59–0.68) |

**Table 1.** Cont.

| Category | Intervention (Selected) | Collection Research Paper | | Selected Findings | | |
|---|---|---|---|---|---|---|
| | | Total Study Sample[a] (Reference Standard) | Reference | Sensitivity (CI) | Specificity (CI) | Accuracy/AUC (CI) |
| | Child with fever | 601 caregivers at least 18 years old of children under five years old presenting for treatment for fever in five health centers in the previous two weeks, Zambia (recording by trained clinician) | Eisele, et al. [23] | 96% (87–100) | 100% (—) | 96% (87–100) |
| | Finger/heel stick performed | Same as previous | | 63% (18–100) | 90% (86–94) | 72% (41–100) |
| | Malaria diagnosis made | Same as previous | | 77% (55–99) | 76% (48–100) | 76% (55–98) |
| | Artemisinin combination therapy given | Same as previous | | 81% (51–100) | 92% (80–100) | 85% (73–98) |

[a]This is the total number of completed interviews with respondents; sample sizes for specific indicators may be smaller and readers are referred to the original article.
AUC, area under the receiver operating characteristic curve; CI, confidence interval.
doi:10.1371/journal.pmed.1001423.t001

alternative sources of medicine. Small-scale trials of similar techniques may yield important findings about how recall accuracy can be improved for interventions that address a broader range of conditions.

## Refine the Survey Questionnaire and/or Procedures

The papers in this Collection suggest that attention should be refocused away from indicators that are not producing valid measurements to those that do, even if these alternative indicators do not provide a full answer to the question about whether a specific intervention was received. For example, indicators related to whether care was sought for a child with signs of pneumonia—although not yet validated—offer promise as measures of access to correct pneumonia treatment and could be coupled with assessments of correct management collected in health service sites [25]. Another way forward might be to assess care seeking for children with signs of any illness, avoiding the pitfall of constructing differential diagnoses based on respondent reports. Alternative question formulations also hold promise for improved maternal reporting on interventions received around the time of birth [20,21]. Moreover, previous research on cognitive aspects of recalling health-related events in developed country settings (e.g., [38]) may hold additional untapped potential for improving MNCH coverage measurement in low- and middle-income countries.

In their papers, the Collection authors make numerous other suggestions for changes in the core DHS/MICS questionnaires and/or procedures. Some are relatively easy to implement and have already been adopted based on preliminary findings. For example, core questionnaires for DHS and MICS surveys now include response options for questions about where a child was taken for care that include the workers responsible for providing community case management [39]. Other suggestions are more ambitious, such as adding questions on the severity of diarrhea [37] or adding HIV testing of mothers and infants in more surveys [26].

Several of the Collection studies report new findings that compare the accuracy of mothers' reports of intervention receipt for different recall periods. Understanding the effect of the recall period on the accuracy of reports is important because where longer recall periods do not jeopardize accuracy by introducing bias they can yield an increased sample size for the measurement of specific indicators without increasing the number of households surveyed. Hazir et al. found no differences in the accuracy of mothers' reports related to childhood pneumonia at two and four weeks [22], which suggests that in future surveys more children with signs of pneumonia can be included in the estimation of indicators for care seeking, if not treatment. Eisele et al. found no drop-off in the accuracy of recalling malaria diagnosis and treatment interventions two weeks versus one week after the clinic visit [23]. These results warrant confirmation, as earlier work has suggested that reporting accuracy decreases with longer recall periods for duration of breastfeeding [40,41], signs of respiratory illness in children [27,42], and a broader range of health symptoms [43].

## Link Household Surveys to Other Sources of Information about Service Provision

Our understanding of services received by mothers and children can be improved by linking information on the sources of care obtained through household surveys to facility assessments of the extent and quality of the interventions delivered. In this way, for example, coverage estimates for peripartum interventions that women who report delivering in a health facility are unable to recall could be linked to temporally specific information about the standard practices and quality of care at that facility. This approach was used in an evaluation in Bangladesh that linked detailed information on care seeking for childhood illness obtained from mothers during a household survey with observation-based assessments of the quality of child health care offered by those providers [44]. Similarly, vaccination records held at health facilities can be sought when a home-based record is not available [32]. The malaria research community is also considering the use of a set of indicators that measure aspects of treatment-seeking practices, diagnosis, and treatment from a combination of sources including household and health facility surveys, rather than the single indicator of receipt of artemisinin combination therapy among children with fever [45]; results presented in this Collection provide further justification for that decision [23]. Tools for assessing service provision [46] and quality of care [47–49] exist for most MNCH interventions; the methodological challenge is to link them to household surveys in ways that produce valid population-based coverage estimates at reasonable cost.

### Incorporate Information Technology

As digital technology becomes more widespread, the potential for quick and accurate electronic recording and transmission of intervention data will improve rapidly. Digital registries offer opportunities to improve monitoring of target populations and of interventions, such that accurate denominators and numerators will become increasingly feasible to obtain via routine administrative data in low- and middle-income countries. Advances in the use of electronic and telecommunications processes in health (e-health) and the use of mobile devices to collect health information in surveys and patient care (m-health) also have important implications for coverage measurement and monitoring. Finally, technology can improve data quality in household surveys and is increasingly being used in real-time measurement of child health program indicators (e.g., [50]). However, further research and evaluation is needed, with strong coordination, to develop and roll out the effective use of digital records and to avoid a proliferation of different, often incompatible, e-health projects.

### Increase the Salience of Intervention Delivery

Health personnel may also be able to contribute to improved coverage measurement. Health workers should explain to mothers the importance of keeping the home-based child health or vaccination record carefully and bringing it to every health facility visit. Reinforcement of the importance of home-based records should increase their availability during household surveys. Careful explanations by service providers about the interventions being delivered, and why, are also likely to increase recall. Explanations need not be limited to the patient contact during which the service is provided. For example, telling a pregnant woman during an antenatal care visit that she should receive an injection immediately following birth may increase her recall of the event. In addition, the salience of intervention delivery may be increased by having an outreach or community health worker wear a highly visible, unique article of clothing (e.g., community health workers providing case management to sick children in Bangladesh carried a bright pink bag to help mothers differentiate them from other types of community workers). Advance planning and creativity on the part of program personnel can yield benefits for coverage monitoring.

### Use Measures That Do Not Rely on Respondents' Reports

A final strategy for improving coverage measurement is to find ways to replace mothers' recall with more objective measures. The use of home-based records of services received offers an attractive alternative to mothers' recall, if the records contain complete and accurate information. However, the review of experience with home-based vaccination records conducted as part of this Collection highlights multiple potential sources of error, including incomplete, inaccurate, or outdated records, as well as errors in transcription [32]. A review of DHS and MICS surveys since 2000 found that home-based vaccination records were available for less than 70% of children in 21 of the 33 least developed countries [51]. The report from China in this Collection, which used a combination of home-based records and an electronic system as the reference standard, had high proportions of missing data for all the coverage indicators assessed [24].

It may also be possible to expand the use of examinations and biological testing to assess coverage for certain indicators and to generate evidence that an intervention has actually been received. Although experience to date has largely been limited to testing for the presence of infection during a household interview (e.g., rapid diagnostic tests for anemia and malaria and testing of mothers and infants for HIV antibodies), work has begun on methods to determine whether specific medicines have been received, and there is growing interest and experience in using medically trained teams to conduct household surveys that include physical examinations and the collection of biological specimens [52]. Measurement of tetanus and measles antibodies in serum or oral fluid to determine receipt of diphtheria-tetanus-pertussis and measles vaccines is also feasible, although each currently has caveats [32]. The potential of such measures and their practicality and cost should be explored.

## Crosscutting Methodological Issues in Coverage Measurement

Three methodological themes appear consistently throughout the Collection papers. First, household surveys based on probability sampling, and conducted with careful attention to quality in sampling, interviewer training, fieldwork, data management and cleaning, and analysis are the bedrock of coverage monitoring at the population level. Although there may be cheaper alternatives to conducting high-quality, nationally representative surveys to ascertain intervention coverage, the use of these alternatives should be treated with caution unless the same rigor is used in them for sampling and quality control as in the larger surveys [15,32]. Facility-based assessments are not based on probability samples of the population, will often overestimate population coverage [26], and cannot replace household surveys for the measurement of intervention coverage as defined here.

Second, no matter how carefully intervention coverage is measured through household surveys, sampling and non-sampling error is always present and must be accounted for when interpreting results for decision-making and program evaluation purposes [15]. To ensure that sampling error is considered, we recommend that all surveys publish appropriately calculated confidence intervals for key coverage indicators, as is already done in DHS and MICS reports, and refer to them consistently when presenting and interpreting the results. Countries implementing nonstandard surveys should ensure that they measure global coverage indicators comparable to those produced by DHS and MICS, and reports should describe the sampling design in

**Box 2. What Needs to Be Done to Improve MNCH Coverage Measurement?**

- Efforts to conduct high-quality household surveys at national and sub-national levels must be sustained to provide essential information on coverage trends and inequities, even as routine health information systems improve.
- These large surveys need to be supplemented with lighter tools that can be implemented every 1–2 years to produce high-quality estimates of MNCH intervention coverage.
- Further investment is needed in complementary assessments of service quality in health service settings, including the delivery of specific interventions during service contacts; these assessments should be synchronized in time and linked geographically to population-based household surveys that measure coverage for the same interventions.
- Efforts to learn more about coverage measurement using innovative designs to assess validity must be recognized and supported.

sufficient detail to allow determination of whether a true probability sample was drawn and whether the survey design was appropriately accounted for in the calculation of the standard errors. To account for non-sampling error such as information and selection bias, survey reports should also include a limitations section that explicitly lists probable sources of non-sampling error, and authors should speculate about the direction and magnitude of error where possible.

Third, disaggregated reporting makes data on intervention coverage more useful to policy and program decision makers. Disaggregation of national coverage estimates by wealth, geographic region, and other relevant stratifiers helps identify groups that are not being reached. A paper in this Collection provides basic guidance on the measurement and interpretation of inequalities in health coverage data from household surveys [16]. As targets are set for universal health coverage, governments and their partners must actively seek out and demand data that help them develop effective, local delivery strategies to reach those who are currently not receiving services, and must seek ways to generate community demand for essential MNCH services.

## Action Recommendations

We hope this Collection will serve as a vehicle for advancing the field of intervention coverage measurement in maternal and child health by providing a strong justification for increased attention to the quality and precision of coverage estimates. One aim of this Collection is to inform those who use coverage indicators at global and national levels, so that they can make sound choices about the selection of indicators and can interpret the results intelligently by recognizing their uncertainty bounds [15,53]. A second aim is to highlight actions needed by care providers and researchers in the MNCH community to improve coverage measurement and the use of coverage results (Box 2).

High-quality household survey programs will continue to be the primary source of data on MNCH intervention coverage for the foreseeable future, even as routine health information systems improve. These surveys are needed to validate and calibrate data produced from other sources, to investigate variations in coverage in specific subgroups, and to assess equity by gender or income. Careful thought needs to be given to how we can ensure household surveys continue to produce the best and most relevant information that is needed by public health decision makers. The survey protocols must be continuously adapted to capture new interventions and delivery strategies for which coverage is not currently measured and to incorporate new evidence about coverage measurement, such as that presented in this Collection. Supplemental survey tools that are tailored to the need for more frequent measurement of MNCH intervention coverage may be required as country and global interest in accountability grows. These supplemental survey tools might focus on what surveys do well—the assessment of coverage for interventions that are clearly defined and highly salient (the numerator) and needed by all members of a specific population subgroup (the denominator)—and leave more challenging measurements for the full and more complex surveys every 3–5 years.

Methodological work is also needed to link survey data on sources of health care to rigorous, comparable assessments of the extent and quality of interventions being delivered in those settings. As routine health information systems improve, especially in middle-income countries, there may be opportunities to calibrate them with data collected from representative samples of the population to increase their usefulness in coverage measurement.

### Key Points

- Regular, high-quality measurement of the proportion of women, newborns, and children in need of life-saving interventions who actually receive them ("intervention coverage") is essential to support sound decisions at local, national, and global levels.
- Standardized household surveys based on probability sampling are the cornerstone of coverage monitoring and provide a wealth of important background information to support interpretation and equity analyses.
- Some of the indicators now being used to track intervention coverage may not provide fully accurate or reliable results for a variety of reasons, among them limitations of respondent recall and of using symptoms as a basis for defining specific diseases.
- A better understanding of the systematic and random error inherent in these coverage indicators—and approaches to mitigate that error—can help in indicator interpretation and use.
- Measurement of intervention coverage can be improved through focused operational research on household survey techniques, supplemented by more frequent assessments using less sophisticated or routine methods, and extended through links to assessments of service quality.

Perhaps the most important message of this Collection is that much remains to be learned about how best to measure MNCH intervention coverage through household surveys. The questions are clear and can be answered through well-designed studies building on and extending the work started here. Options for improving existing metrics need to be systematically evaluated, including alternative question formulations, strategies to aid recall, and use of biomarkers and technologies. New indicators and modules, including those for neonatal interventions [34] and postnatal visits [35], need to be validated, as do indicators for additional intervention areas such as young child feeding that are not addressed in this Collection. Existing findings need to be confirmed among more representative populations in additional countries. Demographic surveillance sites might be engaged to conduct further tests, recognizing that they are not always representative of larger populations and may not currently be able to link household survey data with those generated through facility assessments. This learning agenda must be implemented urgently as a foundation for producing better evidence for stronger programs and improved accountability.

## Acknowledgments

Members of the CHERG Working Group on Improving Coverage Measurement contributed to the authorship of this work: Shams El Arifeen, Aluisio J. D. Barros, Khadija Begum, Harry Campbell, Felicity Cutts, Thomas P. Eisele, Christa Fischer-Walker, Elizabeth Hazel, Tabish Hazir, Amira Khan, Li Liu, Allisyn C. Moran, Barbara Rawlins, Dale A. Rhoda, Jim Ricca, Cynthia K. Stanton, Jeffrey Stringer, and Özge Tunçalp.

## Author Contributions

Wrote the first draft of the manuscript: JB. Contributed to the writing of the manuscript: JB FA AB AH HN JR TW SA AB KB HC FC TE CW EH TH AK LL AM BR DR JR CS JS OT. ICMJE criteria for authorship read and met: JB FA AB AH HN JR TW SA AB KB HC FC TE CW EH TH AK LL AM BR DR JR CS JS OT. Agree with manuscript results and conclusions: JB FA AB AH HN JR TW SA AB KB HC FC TE CW EH TH AK LL AM BR DR JR CS JS OT.

## References


1. United Nations Children's Fund, World Health Organization, World Bank, United Nations (2012) Levels and trends in child mortality: report 2012. New York: United Nations Children's Fund. Available: http://www.unicef.org/videoaudio/PDFs/UNICEF_2012_child_mortality_for_web_0904.pdf. Accessed 20 January 2013.
2. World Health Organization, United Nations Children's Fund, United Nations Population Fund, World Bank (2012) Trends in maternal mortality: 1990 to 2012. Geneva: World Health Organization. Available; http://www.who.int/reproductivehealth/publications/monitoring/9789241503631/en/index.html. Accessed 20 January 2013.
3. World Health Organization, Aga Khan University, Partnership for Maternal, Newborn and Child Health (2011) Essential interventions, commodities and guidelines for reproductive, maternal, newborn and child health: a global review of the key interventions related to reproductive, maternal, newborn and child health. Geneva: World Health Organization. Available: http://www.who.int/pmnch/topics/part_publications/essential_interventions_18_01_2012.pdf. Accessed 11 March 2013.
4. Requejo J, Bryce J, Victora C, the Countdown to 2015 Report Writing Team (2012) Building a future for women and children: the 2012 report. Geneva: World Health Organization. Available: http://www.countdown2015mnch.org/reports-and-articles/2012-report. Accessed 18 October 2012.
5. Barros AJD, Ronsmans C, Axelson H, Loaiza E, Bertoldi AD, et al. (2012) Equity in maternal, newborn and child interventions in Countdown to 2015: a retrospective review of survey data from 54 countries. Lancet 379: 1225–1233.
6. Independent Expert Review Group on Information and Accountability for Women's and Children's Health (2012) Every woman, every child: from commitments to action. Geneva: World Health Organization. http://www.who.int/woman_child_accountability/ierg/reports/2012/IERG_report_low_resolution.pdf. Accessed 20 January 2013.
7. United Nations (2012) The Millennium Development Goals report 2012. New York: United Nations. http://www.un.org/millenniumgoals/pdf/MDG%20Report%202012.pdf. Accessed 20 January 2013.
8. MEASURE DHS Demographic and Health Surveys. Available: http://www.measuredhs.com/. Accessed 20 October 2012.
9. UNICEF. Statistics and Monitoring: Multiple Indicator Cluster Survey. Available: http://www.unicef.org/statistics/index_24302.html. Accessed 11 March 2013.
10. Hancioglu A, Arnold F (2013) Measuring coverage in MNCH: Tracking progress in health for women and children using DHS and MICS household surveys. PLoS Med 10: e1001391. doi:10.1371/journal.pmed.1001391
11. Commission on Information and Accountability for Women's and Children's Health (2011) Keeping promises; tracking results. New York: United Nations. http://whqlibdoc.who.int/publications/2011/9789241564328_eng.pdf. Accessed 20 January 2013.
12. Lim SS, Stein DB, Charrow A, Murray CJL (2008) Tracking progress toward universal childhood immunization and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus and pertussis immunization coverage. Lancet 372: 2031–2046.
13. Child Health Epidemiology Reference Group. Available: http://cherg.org/main.html. Accessed 20 October 2012.
14. Countdown to 2015 for Maternal, Newborn and Child Survival. Available: http://www.countdown2015mnch.org/. Accessed 20 October 2012.
15. Eisele T, Rhoda DA, Cutts FT, Keating J, Ren R, et al. (2013) Measuring coverage in MNCH: Total survey error and the interpretation of intervention coverage estimates from household surveys. PLoS Med 10: e1001386. doi:10.1371/journal.pmed.1001386
16. Barros AJD, Victora CG (2013) Measuring coverage in MNCH: Determining and interpreting inequalities in coverage of maternal, newborn, and child health interventions. PLoS Med 10: e1001390. doi:10.1371/journal.pmed.1001390
17. Tanahashi T (1978) Health service coverage and its evaluation. Bull World Health Organ 56: 295–303.
18. Shengelia B, Tandon A, Adams OB, Murray CJL (2005) Access, utilization, quality, and effective coverage: an integrated conceptual framework and measurement strategy. Soc Sci Med 61: 97–109.
19. Campbell H, Biloglav Z, Rudan I (2008) Reducing bias from test misclassification in burden of disease studies: use of test to actual positive ratio—new test parameter. Croat Med J 49: 402–414.
20. Tunçalp Ö, Stanton C, Castro A, Adanu RK, Heymann M, et al. (2013) Measuring coverage in MNCH: Validating women's self-report of emergency cesarean sections in Ghana and the Dominican Republic. PLoS ONE 8: e60761. doi:10.1371/journal.pone.0060761
21. Stanton CK, Rawlins B, Drake M, Anjos M, Cantor D, et al. (2013) Measuring coverage in MNCH: Testing the validity of women's self-report of key maternal and newborn health Interventions during the peripartum period in Mozambique. PLoS ONE 8: e60694. doi:10.1371/journal.pone.0060694
22. Hazir T, Bequem K, el Arifeen S, Khan AM, Huque MH, et al. (2013) Measuring coverage in MNCH: A prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia. PLoS Med 10: e1001422. doi:10.1371/journal.pmed.1001422
23. Eisele T, Silumbe K, Yukich J, Hamainza B, Keating J, et al. (2013) Measuring coverage in MNCH: Accuracy of measuring diagnosis and treatment of childhood malaria from household surveys in Zambia. PLoS Med 10: e1001417. doi:10.1371/journal.pmed.1001417
24. Guo Y, Liu L, Li M, Yang L, Ju L, et al. (2013) Measuring coverage in MNCH: A validation study linking population survey derived coverage to maternal, newborn, and child health care records in rural China. PLoS ONE 8: e60762. doi:10.1371/journal.pone.0060762
25. Campbell H, el Arifeen S, Hazir H, O'Kelly J, Bryce J, et al. (2013) Measuring coverage in MNCH: Challenges in monitoring the proportion of young children with pneumonia who receive antibiotic treatment. PLoS Med 10: e1001421. doi:10.1371/journal.pmed.1001421
26. Stringer J, Stinson K, Tih PM, Giganti M, Ekouevi D, et al. (2013) Measuring coverage in MNCH: Population HIV-free survival among children under two years of age in four African countries. PLoS Med 10: e1001424. doi:10.1371/journal.pmed.1001424
27. Campbell H, Byass P, Greenwood BM (1990) Acute lower respiratory infections in Gambian children: maternal perception of illness. Ann Trop Paediatr 10: 45–51.
28. Mull DS, Mull JD, Kundi MZ, Anjum M (1994) Mothers' perceptions of severe pneumonia in their own children: a controlled study in Pakistan. Soc Sci Med 38: 973–987.
29. Harrison LH, Moursi S, Guinena AH, Gadomski AM, el-Ansary KS, et al. (1995) Maternal reporting of acute respiratory infection in Egypt. Int J Epidemiol 24: 1058–1063.
30. Lanata CF, Quintanilla N, Verastegui HA (1994) Validity of a respiratory questionnaire to identify pneumonia in children in Lima, Peru. Int J Epidemiol 23: 827–834.
31. Muhe L (1996) Mothers' perceptions of signs and symptoms of acute respiratory infections in their children and their assessment of severity in an urban community of Ethiopia. Ann Trop Paediatr 16: 129–135.
32. Cutts FT, Izurieta H, Rhoda DA (2013) Measuring coverage in MNCH: Design, implementation, and interpretation challenges associated with tracking vaccination coverage using household surveys. PLoS Med 10: e1001404. doi:10.1371/journal.pmed.1001404
33. Gareaballah ET, Loevinsohn BP (1989) The accuracy of mother's reports about their children's vaccination status. Bull World Health Organ 67: 669.
34. Langsten R, Hill K (1998) The accuracy of mothers' reports of child vaccination: evidence from rural Egypt. Soc Sci Med 46: 1205–1212.
35. Valadez JJ, Weld LH (1992) Maternal recall error of child vaccination status in a developing nation. Am J Public Health 82: 120–122.
36. Moran A, Kerber K, Sitrin D, Guenther T, Morrissey CS, et al. (2013) Measuring coverage in MNCH: Indicators for global tracking of newborn care. PLoS Med 10: e1001415. doi:10.1371/journal.pmed.1001415
37. Fischer Walker CL, Fontaine O, Black RE (2013) Measuring coverage in MNCH: Current indicators for measuring coverage of diarrhea treatment interventions and opportunities for improvement. PLoS Med 10: e1001385. doi:10.1371/journal.pmed.1001385
38. Warnecke RB, Sudman S, Johnson TP, O'Rourke D, Davis AM, et al. (1997) Cognitive aspects of recalling and reporting health-related events: Papanicolaou smears, clinical breast examinations, and mammograms. Am J Epidemiol 146: 982–992.
39. Hazel E, Requejo J, David J, Bryce J (2013) Measuring coverage in MNCH: Evaluation of community-based treatment of childhood illnesses through household surveys. PLoS Med 10: e1001384. doi:10.1371/journal.pmed.1001384
40. Bland RM, Rollins NC, Solarsh G, Van den Broeck J, Coovadia HM, et al. (2003) Maternal recall of exclusive breast feeding duration. Arch Dis Child 88: 778–783.
41. Li R, Scanlon KS, Serdula MK (2005) The validity and reliability of maternal recall of breastfeeding practice. Nutr Rev 63: 103–110.
42. Feikin DR, Audi A, Olack B, Bigogo GM, Polyak C, et al. (2010) Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. Int J Epidemiol 39: 450–458.
43. Schmier JK, Halpern MT (2004) Patient recall and recall bias of health state and health status. Expert Rev Pharmacoecon Outcomes Res 4: 159–163.
44. Arifeen SE, Hoque DME, Akter T, Rahman M, Hoque ME, et al. (2009) Effect of the Integrated Management of Childhood Illness strategy on childhood mortality and nutrition in a rural area in Bangladesh: a cluster randomised trial. Lancet 374: 393–403.
45. Roll Back Malaria (2012) Monitoring and Evaluation Reference Group. Available: http://www.rollbackmalaria.org/mechanisms/merg.html. Accessed 3 November 2012.
46. MEASURE DHS (2012) SPA overview. Available: http://www.measuredhs.com/What-We-Do/Survey-Types/SPA.cfm. Accessed 27 October 2012.
47. Gilroy KE, Callaghan-Koru JA, Cardemil CV, Nsona H, Amouzou A, et al. (2012) Quality of sick child care delivered by Health Surveillance Assistants in Malawi. Health Policy Plan. E-pub ahead of print. doi:10.1093/heapol/czs095
48. World Health Organization (2003) Health facility survey: tool to evaluate the quality of care delivered to sick children attending outpatients facilities. Geneva: Department of Child and Adolescent Health and Development, World Health Organization. Available: http://whqlibdoc.who.int/publications/2003/9241545860.pdf. Accessed 11 March 2013.

49. Maternal and Child Health Integrated Program (2013) Maternal and newborn quality of care surveys. Available: http://www.mchip.net/QoCsurveys. Accessed 11 March 2013.
50. Rajatonirina S, Heraud JM, Randrianasolo L, Orelle A, Razanajatovo NH, et al. (2012) Short message service sentinel surveillance of influenza-like illness in Madagascar, 2008–2012. Bull World Health Organ 90: 385–389.
51. Brown DW (2012) Child immunization cards: essential yet underutilized in national immunization programmes. Open Vaccine J 5: 1–7.
52. World Health Organization (2012) SAGE: WHO study on global ageing and adult health. Available: http://www.who.int/healthinfo/systems/sage/en/index2.html. Accessed 20 October 2012.
53. Requejo JH, Newby H, Bryce J (2013) Measuring coverage in MCNH: Challenges and opportunities in the selection of coverage indicators for global monitoring. PLoS Med 10: e1001416. doi:10.1371/journal.pmed.1001416

PLOS | MEDICINE

# Measuring Coverage in MNCH: Population HIV-Free Survival among Children under Two Years of Age in Four African Countries

**Jeffrey S. A. Stringer[1,2]*, Kathryn Stinson[3], Pius M. Tih[4], Mark J. Giganti[1], Didier K. Ekouevi[5,6], Tracy L. Creek[7], Thomas K. Welty[4], Benjamin H. Chi[1,2], Catherine M. Wilfert[8], Nathan Shaffer[9], Elizabeth M. Stringer[1,2], Francois Dabis[5,6], David Coetzee[3]**

**1** Centre for Infectious Disease Research in Zambia, Lusaka, Zambia, **2** Department of Obstetrics and Gynecology, University of North Carolina School of Medicine, Chapel Hill, North Carolina, United States of America, **3** Centre for Infectious Disease Epidemiology and Research, School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa, **4** Cameroon Baptist Health Convention Health Board, Bamenda, Cameroon, **5** Programme PAC-CI, Abidjan, Côte D'Ivoire, France, **6** Institute National de la Santé et de la Recherche Médicale, Université Victor Segalen, Bordeaux, France, **7** Children's Healthcare of Atlanta, Atlanta, Georgia, United States of America, **8** Elizabeth Glaser Pediatric AIDS Foundation, Washington, District of Columbia, United States of America, **9** World Health Organization, Geneva, Switzerland

## Abstract

***Background:*** Population-based evaluations of programs for prevention of mother-to-child HIV transmission (PMTCT) are scarce. We measured PMTCT service coverage, regimen use, and HIV-free survival among children ≤24 mo of age in Cameroon, Côte D'Ivoire, South Africa, and Zambia.

***Methods and Findings:*** We randomly sampled households in 26 communities and offered participation if a child had been born to a woman living there during the prior 24 mo. We tested consenting mothers with rapid HIV antibody tests and tested the children of seropositive mothers with HIV DNA PCR or rapid antibody tests. Our primary outcome was 24-mo HIV-free survival, estimated with survival analysis. In an individual-level analysis, we evaluated the effectiveness of various PMTCT regimens. In a community-level analysis, we evaluated the relationship between HIV-free survival and community PMTCT coverage (the proportion of HIV-exposed infants in each community that received any PMTCT intervention during gestation or breastfeeding). We also compared our community coverage results to those of a contemporaneous study conducted in the facilities serving each sampled community. Of 7,985 surveyed children under 2 y of age, 1,014 (12.7%) were HIV-exposed. Of these, 110 (10.9%) were HIV-infected, 851 (83.9%) were HIV-uninfected, and 53 (5.2%) were dead. HIV-free survival at 24 mo of age among all HIV-exposed children was 79.7% (95% CI: 76.4, 82.6) overall, with the following country-level estimates: Cameroon (72.6%; 95% CI: 62.3, 80.5), South Africa (77.7%; 95% CI: 72.5, 82.1), Zambia (83.1%; 95% CI: 78.4, 86.8), and Côte D'Ivoire (84.4%; 95% CI: 70.0, 92.2). In adjusted analyses, the risk of death or HIV infection was non-significantly lower in children whose mothers received a more complex regimen of either two or three antiretroviral drugs compared to those receiving no prophylaxis (adjusted hazard ratio: 0.60; 95% CI: 0.34, 1.06). Risk of death was not different for children whose mothers received a more complex regimen compared to those given single-dose nevirapine (adjusted hazard ratio: 0.88; 95% CI: 0.45, 1.72). Community PMTCT coverage was highest in Cameroon, where 75 of 114 HIV-exposed infants met criteria for coverage (66%; 95% CI: 56, 74), followed by Zambia (219 of 444, 49%; 95% CI: 45, 54), then South Africa (152 of 365, 42%; 95% CI: 37, 47), and then Côte D'Ivoire (3 of 53, 5.7%; 95% CI: 1.2, 16). In a cluster-level analysis, community PMTCT coverage was highly correlated with facility PMTCT coverage (Pearson's $r = 0.85$), and moderately correlated with 24-mo HIV-free survival (Pearson's $r = 0.29$). In 14 of 16 instances where both the facility and community samples were large enough for comparison, the facility-based coverage measure exceeded that observed in the community.

***Conclusions:*** HIV-free survival can be estimated with community surveys and should be incorporated into ongoing country monitoring. Facility-based coverage measures correlate with those derived from community sampling, but may overestimate population coverage. The more complex regimens recommended by the World Health Organization seem to have measurable public health benefit at the population level, but power was limited and additional field validation is needed.

*Please see later in the article for the Editors' Summary*.

**Citation:** Stringer JSA, Stinson K, Tih PM, Giganti MJ, Ekouevi DK, et al. (2013) Measuring Coverage in MNCH: Population HIV-Free Survival among Children under Two Years of Age in Four African Countries. PLoS Med 10(5): e1001424. doi:10.1371/journal.pmed.1001424

**Academic Editor:** Carla AbouZahr, Independent consultant: Health statistics and policy, Switzerland

**Received** September 1, 2012; **Accepted** March 4, 2013; **Published** May 7, 2013



**Funding:** The Zambia, South Africa, and Côte d'Ivoire work was supported by contract T0906150021 from the US Centers for Disease Control and Prevention (CDC) Global AIDS Program. The Cameroon work was supported by a grant from The Bill & Melinda Gates Foundation (351-07), which was awarded through the Elizabeth Glaser Pediatric AIDS Foundation. The CDC was not involved in data collection or management, but through coauthors and the CDC manuscript clearance process, the CDC was substantially involved in the preparation, review, and approval of the manuscript. The findings and conclusions of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the CDC. The Bill & Melinda Gates Foundation was not involved in the design or conduct of the study, collection, management, analysis, or interpretation of the data, nor was it involved in preparation, review, or approval of the manuscript. The Elizabeth Glaser Pediatric AIDS Foundation was not involved in data collection, management, or analysis, but through a coauthor was involved in design of the study, data interpretation, and manuscript preparation, review, and approval.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** ARV, antiretroviral; DHS, Demographic and Health Surveys; HAART, highly active antiretroviral therapy; HR, hazard ratio; IQR, interquartile range; NVP, nevirapine; PEARL Study, PMTCT Effectiveness in Africa: Research and Linkages to Care Study; PMTCT, prevention of mother-to-child HIV transmission.

* E-mail: jeff_stringer@unc.edu

*This paper is part of the* PLOS Medicine "*Measuring Coverage in MNCH" Collection.*

## Introduction

While transmission of HIV from mother to infant has been virtually eliminated in high-income countries, substantial challenges remain in low- and middle-income settings, where almost 1,000 children are thought to acquire HIV each day [1–3]. The efficacy of antiretroviral (ARV) regimens during pregnancy and postpartum to prevent mother-to-child transmission is well established, and prevention of mother-to-child transmission (PMTCT) programs have been implemented in most countries in Africa [4,5]. However, objective measures of program effectiveness and public health impact are largely absent from the refereed literature [6,7]. Despite calls for the virtual elimination of pediatric HIV, global coverage of PMTCT services remains far below what will be required to meet such an aggressive goal [8]. Current ARV guidelines for resource-limited settings recommend earlier and more effective regimens at higher CD4+ T-lymphocyte thresholds in pregnancy, as well as postpartum treatment or prophylaxis to the mother or child to prevent transmission through breastfeeding. More recently, the initiation of lifelong highly active antiretroviral therapy (HAART) has been proposed for all pregnant, HIV-infected women [9–11].

The performance of national PMTCT programs is commonly assessed by aggregating routinely collected service coverage indicators, which are often incomplete and inaccurate, and typically lack the site-level nuance needed to drive meaningful quality improvement efforts [12–14]. Recent service-based evaluations have used umbilical cord blood surveillance [15,16] and infant PCR testing at immunization clinics [17,18] to improve data quality, but since these methods are facility-based they may not represent the entire community [6].

## Methods

The PMTCT Effectiveness in Africa: Research and Linkages to Care Study (PEARL Study) was a multi-country evaluation of the effectiveness of PMTCT services at the community and facility level. The study's methodology and facility-based components have been reported elsewhere [6,15,19–21]. The community survey component of the study was conducted in 26 communities in Cameroon, Côte D'Ivoire, South Africa, and Zambia between April 2007 and May 2009.

### Household Selection

Each country employed a two-stage sampling technique. The first stage involved random selection of facilities providing PMTCT services and has been described in detail elsewhere [15]. Once the health facilities were selected in the first stage of sampling, the catchment area of each facility was defined. In all countries, this was done at the local level by reviewing available district data and consulting with local staff. In South Africa and Zambia, we used publically available satellite images (http://www.google.com/earth); in Cameroon and Côte D'Ivoire, we relied upon local maps. Once the catchment boundaries were defined, we subdivided each evaluation area by overlaying a grid of 1 km by 1 km squares. We then estimated the approximate number of households in each square. If the number of households exceeded 50, the square was further subdivided into four smaller squares. All squares of at least five households were then ranked in random order using a "rank probability proportional to size" technique and were then visited in sequence, with all households in each successive grid approached until at least 250 households had been surveyed. In cases where no one was home to participate in the interview, we made two additional attempts to return, after which we counted the household as unavailable.

Each country team employed different ways to educate the communities prior to the surveys, but in each case we met with local community leaders to introduce the survey. Drama performances at public markets and targeted radio announcement were also used. Households were eligible to participate if a child was born in the home during the 24 mo prior to the study visit (whether the mother or infant was currently alive or dead). If the child's mother was not available for interview, we made up to two additional attempts to find her. Failing that, or in the case of a deceased mother, a caregiver was interviewed.

### Study Questionnaire

The survey instrument was adapted from the Zambia Demographic and Health Surveys (DHS) [22] and comprised three parts. Team members used an eligibility form to enumerate all living and deceased household members, their sex, and their age. If there was a child born in the household within the 2 y prior to the survey, the household was deemed eligible for participation in the study and informed consent was obtained. The primary survey respondent was the mother of the eligible child. If the mother was not available, the child's primary caregiver was asked to participate. The maternal questionnaire focused on household characteristics; demographic information; HIV knowledge, attitudes, and perceptions; and previous obstetrical history. The infant module covered specific information about the index pregnancy, including access to and utilization of PMTCT, delivery and postnatal care, and infant feeding. Teams of at least two members, including at least one registered nurse, were trained to administer the survey in local languages at each site.

### Specimen Collection and Testing

In all countries we tested all participating HIV-exposed infants under 24 mo of age for HIV infection. Our procedures for household testing and results disclosure varied by country. In Cameroon, mothers and children over 1 y of age were antibody tested in the household, where pre- and post-test counseling was performed, and results were provided during the visit. Infant dried blood spots for HIV PCR testing were not collected from children whose mothers were seronegative or from those whose mothers were seropositive, but who were over 12 mo of age, no longer breastfeeding, and themselves HIV antibody negative. In the other countries, specimens were obtained from all consenting mothers and children, with HIV testing performed off-site. In Zambia, we provided a card with a linked identification number that allowed participants to access their results (and/or those of their child) at the local health facility. In South Africa and Côte D'Ivoire, test results were not provided to participants. This was a requirement

of local ethics review boards, since we were unable to do confirmatory testing in the home. Women were encouraged to access routine counseling and testing services if they so desired, and the study team provided information about where free services could be accessed. In Zambia, we also offered infant hemoglobin testing with referral for children found to be anemic.

All countries used the Determine HIV-1 rapid antibody test (Abbott Laboratories) for initial testing. Confirmatory tests differed by country. We tested infant dried blood spots for HIV DNA PCR using Roche Amplicor version 1.5 with manual extraction.

### Informed Consent and Ethical Approvals

In each country setting, informed consent was obtained prior to administering the questionnaire and blood draws. The informed consent information was customized to suit the procedure and research ethics requirements in the respective countries. In cases of participant illiteracy, a non-biased literate witness joined the informed consent procedure to ensure that objective information was given about the study and that the patient understood. Ethical approval was obtained from the institutional review boards at the University of Alabama at Birmingham (Birmingham, Alabama, US) and the US Centers for Disease Control and Prevention (Atlanta, Georgia, US), as well as the local research ethics review bodies in each of the participating countries.

### Analysis

Our primary outcome was HIV-free survival among HIV-exposed children. Our sample size was chosen to deliver a half-confidence interval of 5% around the primary outcome in each country; we assumed a binomial distribution. The actual number of houses approached per community also took into account country-specific estimates of maternal HIV seroprevalence and baseline child survival. Statistical analyses were performed using SAS statistical software version 9.1.4 (SAS Institute) and R software version 2.4.1 (http://www.r-project.org). Background characteristics were described using medians and interquartile ranges (IQRs) for continuous variables and percentages for categorical variables. Children were stratified according to their HIV-exposure status, HIV-status, and mortality. The present analysis is limited to those children who were known to be exposed to HIV, either through the detection of HIV antibodies in the infant him- or herself, or through a positive maternal HIV diagnosis. HIV-exposed children who were alive and tested HIV negative by DNA PCR were right-censored at the time of testing in time-to-event analyses. HIV-infected children were left-censored at the time of testing, because transmission was known to have occurred prior to the test date. Twenty-four-month HIV-free survival was estimated using a Weibull regression model and a parametric accelerated failure time model with corresponding 95% confidence intervals. Hazard ratios (HRs) for various covariates of interest were fitted in multivariate Weibull regression models and were restricted to children with responses for each of the covariates of interest. Children born in a particular community or a particular country may be similar to each other. We accounted for this potential correlation by including a separate fixed effect variable for both community and country in our multivariable model. In cases where one mother contributed more than one child to the analysis (e.g., twin gestation or short birth interval), we accounted for this correlation by including a "mother" variable as a random effect. To understand the potential for variable co-linearity, we calculated the variance inflation factors of all variables included in the multivariate model. The variance inflation factor for each variable was less than 1.5, which indicates variable multi-co-linearity is not an issue in our model. The effect of infant feeding on HIV-free survival was limited to those children who were at least 6 mo of age.

In a planned secondary analysis, we compared coverage of PMTCT services at the facility level ("facility PMTCT coverage") to coverage of PMTCT services at the community level ("community PMTCT coverage"). We also compared facility PMTCT coverage to HIV-free survival among HIV-exposed infants. Facility PMTCT coverage was ascertained through an anonymous cord blood surveillance exercise that has been described elsewhere [15]. In the present report, we compare this facility-based coverage measure to the results of our community surveys. Community PMTCT coverage is defined as the proportion of children in our survey born to HIV-positive mothers in whom any ARV drug is reported to have been used during pregnancy. We weighted our estimates according to the number of HIV-exposed infants in a given service cluster and restricted the analysis to clusters with at least 15 exposed infants.

## Results

The household survey was conducted in six areas in Cameroon, six in Côte D'Ivoire, six in South Africa, and eight in Zambia. Collectively, these areas represent the entire catchment population of the 43 facilities included in the sampling scheme [15]. Of 28,942 households visited between 16 May 2008 and 20 May 2009, 9,348 (32%) were eligible (Figure 1). In total, 10,236 children had been born to 9,606 mothers in the previous 2 y. Some children could not be analyzed: 1,465 (14%) mothers or caregivers refused to participate in the survey, while 588 (6%) agreed to be interviewed but refused to provide specimens. In addition, we lost or could not analyze 198 specimens (2%).

Of 7,985 surveyed and analyzable children under 2 y of age, 1,014 (12.7%) were HIV-exposed. Of these, 110 (10.9%) were HIV-infected, 851 (83.9%) were HIV-uninfected, and 53 (5.2%) were dead (Figure 1). Overall response rates—defined as complete child data available—varied by country, with South Africa having the lowest (64%), followed by Côte D'Ivoire (73%), Zambia (88%), and Cameroon (90%).

### Participant Characteristics

Household characteristics are shown in Table 1. The proportion of households reporting access to a flush toilet or pit latrine ranged from 41% in Côte D'Ivoire to 98% in Cameroon. More than 60% of all households in each of the four countries had finished floors (e.g., cement, tile, or wood). Access to electricity was highest in South Africa and Côte D'Ivoire (95%), and lower in Cameroon (66%) and Zambia (33%). South Africa had the highest proportion of households with a refrigerator (75%) and television (81%). The proportion of homes with mobile phones ranged from 55% in Zambia to 87% in South Africa.

The median age of women participating in the survey was 25 y (IQR: 21–30) and the median parity was 2 (IQR: 1–3). With the exception of South African mothers (43%), the majority of women were married or cohabiting. Employment was highest in Cameroon (67%) and lowest in South Africa (20%). In South Africa 83% of mothers had completed at least secondary level school, but only 13% had done so in Cameroon (Table 2).

Reported utilization of antenatal and delivery services was high across all countries (Table 3). More than 90% of women had attended at least one antenatal visit during their pregnancy with the index child, and the proportion reporting a facility delivery ranged from 87% in Côte D'Ivoire to 96% in Cameroon. The median gestational age at which women initiated antenatal care was 4 mo (IQR: 3–6) in Côte D'Ivoire and South Africa and 5 mo

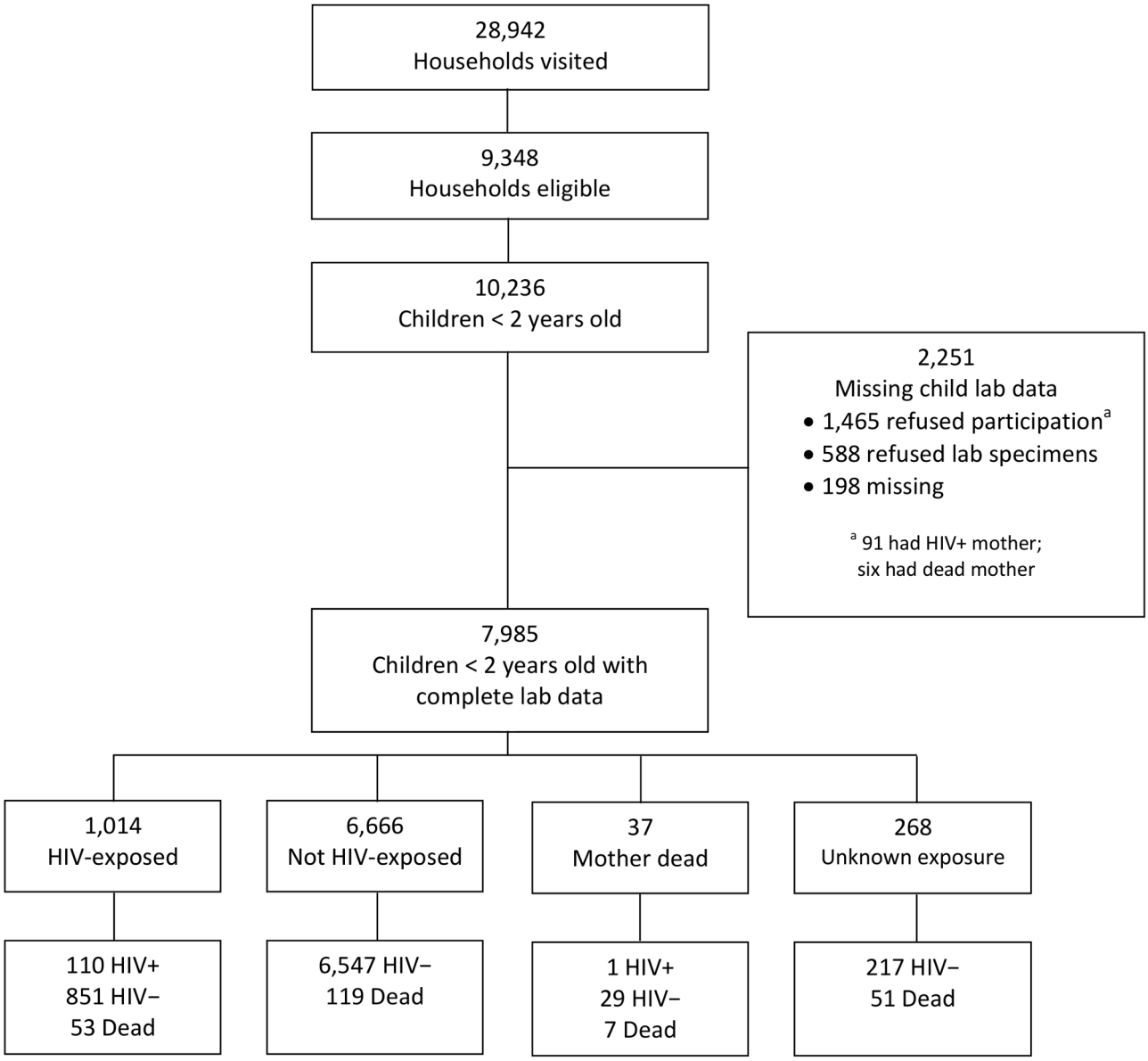


**Figure 1. Description of the cohort of children born in the previous 2 y ($n=10,236$) from all eligible households ($n=9,348$) visited from May 2008 to May 2009.**
doi:10.1371/journal.pmed.1001424.g001

(IQR: 4–6) in Cameroon and Zambia. The proportion of respondents who reported having undergone HIV testing during the index pregnancy was high for Zambia (91%), Cameroon (93%), and South Africa (94%), but much lower in Côte D'Ivoire (39%). Among women who reported having been tested for HIV during the index pregnancy, the rates of self-reported HIV-positivity were significantly lower than the actual HIV seroprevalence of mothers from blood collected at the time of the survey (Côte D'Ivoire [0.3% versus 3.9%; $p<0.01$], Cameroon [4.9% versus 7.4%; $p<0.01$], South Africa [9.1% versus 19.8%; $p<0.01$], and Zambia [12.6% versus 23.6%; $p<0.01$]).

The median birth weight of the index pregnancy was similar across countries (total cohort median 3,100 g; IQR: 2,800–3,500). Measles vaccination coverage by 10 mo of age ranged between 61% (Côte D'Ivoire) and 90% (Cameroon). In Côte D'Ivoire, Cameroon, and Zambia, more than 75% of children more than 6 mo of age were reported to have been exclusively breastfed for 6 mo, while in South Africa, only 40% of children were reported to have been exclusively breastfed, and mixed feeding was common (47%). The proportion of HIV-exposed infants in each country who received no PMTCT prophylaxis at all varied from 48% in Cameroon to 61% in South Africa. The highest proportion of infants whose mothers received HAART during pregnancy was found in Côte D'Ivoire (46%), followed by 14% in Zambia, 9% in South Africa, and 4% in Cameroon (Table 2).

## HIV-Free Survival

HIV-free survival at 24 mo of age among HIV-exposed children was 79.7% (95% CI: 76.4, 82.6). These rates differed by country, but not significantly so: Cameroon (72.6%; 95% CI:

**Table 1.** Characteristics of households surveyed with a child under 2 y of age in the PEARL Study.

| Characteristic | Côte D'Ivoire | | Cameroon | | Zambia | | South Africa | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | Value | *N* | Value | *N* | Value | *N* | Value | *N* | Value |
| **Participation in survey** | 1,682 | 1,579 (93.9%) | 2,425 | 2,403 (99.1%) | 2,302 | 2,089 (90.7%) | 2,939 | 1,924 (65.5%) | 9,348 | 7,995 (85.6%) |
| **Water supply to household** | 1,561 | | 2,385 | | 2,083 | | 1,915 | | 7,944 | |
| Piped water into house | | 409 (26.2%) | | 97 (4.1%) | | 305 (14.6%) | | 762 (39.8%) | | 1,573 (19.8%) |
| Piped water outside but available within plot | | 749 (48.0%) | | 302 (12.7%) | | 279 (13.4%) | | 937 (48.9%) | | 2,267 (28.5%) |
| Public tap | | 130 (8.3%) | | 1,472 (61.7%) | | 488 (23.4%) | | 177 (9.2%) | | 2,267 (28.5%) |
| Other | | 273 (17.5%) | | 514 (21.6%) | | 1,011 (48.5%) | | 39 (2.0%) | | 1,837 (23.1%) |
| **Toilet facilities in household** | 1,556 | | 2,394 | | 2,086 | | 1,921 | | 7,957 | |
| Flush toilet or pit latrine | | 637 (40.9%) | | 2,352 (98.2%) | | 1,844 (88.4%) | | 1,577 (82.1%) | | 6,410 (80.6%) |
| No facility | | 919 (59.1%) | | 42 (1.8%) | | 242 (11.6%) | | 344 (17.9%) | | 1,547 (19.4%) |
| **Does your household have electricity?** | 1,566 | | 2,400 | | 2,086 | | 1,924 | | 7,976 | |
| Yes | | 1,491 (95.2%) | | 1,597 (66.5%) | | 685 (32.8%) | | 1,826 (94.9%) | | 5,599 (70.2%) |
| No | | 75 (4.8%) | | 803 (33.5%) | | 1,401 (67.2%) | | 98 (5.1%) | | 2,377 (29.8%) |
| **Does your household have a refrigerator?** | 1,563 | | 2,398 | | 2,085 | | 1,923 | | 7,969 | |
| Yes | | 303 (19.4%) | | 296 (12.3%) | | 458 (22.0%) | | 1,444 (75.1%) | | 2,501 (31.4%) |
| No | | 1,260 (80.6%) | | 2,102 (87.7%) | | 1,627 (78.0%) | | 479 (24.9%) | | 5,468 (68.6%) |
| **Main material of floor** | 1,557 | | 2,351 | | 2,055 | | 1,917 | | 7,880 | |
| Finished floor (cement/tiles/wood planks) | | 1,375 (88.3%) | | 1,593 (67.8%) | | 1,257 (61.2%) | | 1,675 (87.4%) | | 5,900 (74.9%) |
| Natural floor (earth/mud/dung/sand) | | 182 (11.7%) | | 758 (32.2%) | | 798 (38.8%) | | 242 (12.6%) | | 1,980 (25.1%) |
| **Does your household have a television?** | 1,564 | | 2,396 | | 2,059 | | 1,914 | | 7,933 | |
| Yes | | 1,188 (76.0%) | | 1,243 (51.9%) | | 797 (38.7%) | | 1,558 (81.4%) | | 4,786 (60.3%) |
| No | | 376 (24.0%) | | 1,153 (48.1%) | | 1,262 (61.3%) | | 356 (18.6%) | | 3,147 (39.7%) |
| **Does your household have a cell phone?** | 1,563 | | 2,393 | | 2,075 | | 1,912 | | 7,943 | |
| Yes | | 1,205 (77.1%) | | 1,583 (66.2%) | | 1,139 (54.9%) | | 1,664 (87.0%) | | 5,591 (70.4%) |
| No | | 358 (22.9%) | | 810 (33.8%) | | 936 (45.1%) | | 248 (13.0%) | | 2,352 (29.6%) |

doi:10.1371/journal.pmed.1001424.t001

62.3, 80.5), South Africa (77.7%; 95% CI: 72.5, 82.1), Zambia (83.1%; 95% CI: 78.4, 86.8), and Côte D'Ivoire (84.4%; 95% CI: 70.0, 92.2) (Figure 2A). HIV-free survival was highest in children born to mothers who received dual ARV prophylaxis or HAART (88.5%; 95% CI: 82.4, 92.6), followed by children whose mothers received single-dose nevirapine (NVP) (82.2%; 95% CI: 73.7, 88.2), followed by children who received no PMTCT intervention (78.3%; 95% CI: 73.9, 82.0), although these results did not differ significantly (Figure 2B). In adjusted analyses, risk of death or HIV infection among children receiving a more complex regimen of either dual ARV prophylaxis or HAART was lower (adjusted HR: 0.60; 95% CI: 0.34, 1.06; $p = 0.067$), compared to those receiving no prophylaxis, but not significantly so. In a pairwise comparison of children receiving a more complex regimen of either dual ARV prophylaxis or HAART versus those receiving only single-dose NVP prophylaxis, we did not observe additional benefit (adjusted HR: 0.88; 95% CI: 0.45, 1.44), but power to detect a difference was limited. No other parameters studied were associated with improved HIV-free survival (Table 4).

### PMTCT Intervention Coverage

Of 976 HIV-exposed infants in the study for whom complete laboratory and survey data are available, 449 reportedly received at least some ARV drug prophylaxis during pregnancy or the postnatal period. This represents an unadjusted community PMTCT coverage of 46% (95% CI: 43, 49). Community PMTCT coverage differed significantly by country. It was highest in Cameroon, where 75 of 114 HIV-exposed infants met criteria for coverage (66%; 95% CI: 56, 74), followed by Zambia (219 of 444, 49%; 95% CI: 45, 54), then South Africa (152 of 365, 42%; 95% CI: 37, 47), and then Côte D'Ivoire (3 of 53, 5.7%; 95% CI: 1.2, 16).

In a cluster-level analysis limited to the 16 facility–community pairs that had at least 15 HIV-exposed infants, we found community PMTCT coverage to be highly correlated with facility PMTCT coverage [15] (Pearson's correlation coefficient $r = 0.85$). In every instance except for two, the facility-based coverage measure exceeded that observed in the community (Table 5; Figure 3A). While community PMTCT coverage was moderately correlated with HIV-free survival ($r = 0.29$; Figure 3B), we did not observe an association between facility PMTCT coverage and community HIV-free survival ($r = 0.02$; Figure 3C).

## Discussion

The PEARL Study community survey sampled representative households in four African countries to estimate 24-mo HIV-free survival among HIV-exposed children. Despite high reported HIV testing rates and utilization of delivery services, overall community PMTCT coverage was low, with less than half of HIV-exposed infants having received any ARV prophylaxis during gestation or the postnatal period. We think there is an important lesson to be learned in comparing the previously reported facility-based coverage estimates [15] to the community-based estimates

**Table 2.** Characteristics of mothers surveyed with a child under 2 y of age in the PEARL Study.

| Characteristic | Côte D'Ivoire | | Cameroon | | Zambia | | South Africa | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | Value | *N* | Value | *N* | Value | *N* | Value | *N* | Value |
| **Participation in survey** | 1,869 | 1,751 (93.7%) | 2,451 | 2,423 (98.9%) | 2,347 | 2,116 (90.2%) | 2,939 | 2,029 (69.0%) | 9,606 | 8,319 (86.6%) |
| **Caregiver interviewed instead of mother** | 1,731 | 2 (0.1%) | 2,412 | 25 (1.0%) | 2,110 | 18 (0.9%) | 2,013 | 121 (6.0%) | 8,266 | 166 (2.0%) |
| **Age at time of survey (median [IQR])** | 1,730 | 26 (21, 30) | 2,404 | 25 (22, 30) | 2,112 | 25 (21, 31) | 2,026 | 25 (21, 32) | 8,272 | 25 (21, 30) |
| 12–19 y | | 227 (13.1%) | | 276 (11.5%) | | 285 (13.5%) | | 312 (15.4%) | | 1,100 (13.3%) |
| 20–29 y | | 995 (57.5%) | | 1,487 (61.9%) | | 1,211 (57.4%) | | 1,074 (53.0%) | | 4,767 (57.6%) |
| 30+ y | | 508 (29.4%) | | 639 (26.6%) | | 615 (29.1%) | | 640 (31.6%) | | 2,402 (29.0%) |
| **Parity (median [IQR])** | 1,558 | 2 (1, 3) | 2,394 | 2 (1, 3) | 2,084 | 2 (1, 4) | 1,918 | 2 (1, 3) | 7,954 | 2 (1, 3) |
| 1 | | 638 (40.9%) | | 812 (33.9%) | | 681 (32.7%) | | 861 (44.9%) | | 2,992 (37.6%) |
| 2 | | 500 (32.1%) | | 575 (24.0%) | | 461 (22.1%) | | 534 (27.8%) | | 2,070 (26.0%) |
| 3+ | | 420 (27.0%) | | 1,007 (42.1%) | | 942 (45.2%) | | 523 (27.3%) | | 2,892 (36.4%) |
| **Marital status** | 1,726 | | 2,414 | | 2,103 | | 2,019 | | 8,262 | |
| Married/cohabitating | | 1,513 (87.7%) | | 1,878 (77.8%) | | 1,655 (78.7%) | | 870 (43.1%) | | 5,916 (71.6%) |
| Other | | 213 (12.3%) | | 536 (22.2%) | | 448 (21.3%) | | 1,149 (56.9%) | | 2,346 (28.4%) |
| **Education** | 1,723 | | 2,401 | | 2,100 | | 2,019 | | 8,243 | |
| Secondary or higher | | 221 (12.8%) | | 980 (40.8%) | | 872 (41.5%) | | 1,682 (83.3%) | | 3,755 (45.6%) |
| No schooling or primary | | 1,502 (87.2%) | | 1,421 (59.2%) | | 1,228 (58.5%) | | 337 (16.7%) | | 4,488 (54.4%) |
| **Mother currently employed** | 1,562 | | 2,393 | | 2,071 | | 1,921 | | 7,947 | |
| Yes | | 640 (41.0%) | | 1,600 (66.9%) | | 1,136 (54.9%) | | 372 (19.4%) | | 3,748 (47.2%) |
| No | | 922 (59.0%) | | 793 (33.1%) | | 935 (45.1%) | | 1,549 (80.6%) | | 4,199 (52.8%) |
| **Maternal HIV status (lab data)** | 1,325 | | 2,247 | | 1,968 | | 1,917 | | 7,457 | |
| Positive | | 52 (3.9%) | | 167 (7.4%) | | 464 (23.6%) | | 379 (19.8%) | | 1,062 (14.2%) |
| Negative | | 1,265 (95.5%) | | 2,076 (92.4%) | | 1,499 (76.2%) | | 1,515 (79.0%) | | 6,355 (85.2%) |
| Mother dead | | 8 (0.6%) | | 4 (0.2%) | | 5 (0.3%) | | 23 (1.2%) | | 40 (0.5%) |
| **PMTCT regimen** | 48 | | 160 | | 450 | | 363 | | 1,021 | |
| None | | 25 (52.1%) | | 77 (48.1%) | | 240 (53.3%) | | 222 (61.2%) | | 564 (55.2%) |
| NVP only | | 1 (2.1%) | | 37 (23.1%) | | 48 (10.7%) | | 62 (17.1%) | | 148 (14.5%) |
| NVP+zidovudine | | 0 (0.0%) | | 40 (25.0%) | | 98 (21.8%) | | 46 (12.7%) | | 184 (18.0%) |
| HAART | | 22 (45.8%) | | 6 (3.8%) | | 64 (14.2%) | | 33 (9.1%) | | 125 (12.2%) |

Data are number (percent) unless otherwise indicated.

doi:10.1371/journal.pmed.1001424.t002

reported here. While the estimates correlate highly ($r = 0.85$), we found the facility-based coverage estimate to be higher than the community-based coverage estimate in every facility–community pair except for two (Figure 3A). This suggests to us that the facility population can be thought of as a subset of the community, and a coverage estimate derived from it will be valid only in settings where most patients come to the facility for services. There are mother–infant pairs in the community who are simply not sampled by a facility-based estimate, and, not surprisingly, the PMTCT coverage rates among these women were generally lower. This finding is consistent with the observed associations between the two PMTCT coverage estimates and HIV-free survival (Figure 3B and 3C). While there is a moderate association between community PMTCT coverage and HIV-free survival, the association weakens considerably when the coverage estimate is derived from the facility population only.

The risk of death was lower among children using a potent ARV regime for PMTCT, but due to limited power, the association was not significant. An effect of drug potency is suggested by a dose response in the adjusted analysis and in Figure 2B. While this phenomenon has been known for some time in the controlled environment of clinical trials, its demonstration at the population level provides support for World Health Organization recommendations to implement more complex PMTCT ARV regimens on a wide scale.

In aggregate, we think our findings suggest a critical role for both facility- and community-based monitoring of PMTCT programs. The facility-based estimates—particularly those capable of recreating the PMTCT "cascade" [15,17]—can provide important, practical information about how to best improve service delivery at the site level. Facility-based estimates are also easier to obtain, and are thus suited for more routine performance monitoring. Since some patients do not interact with the healthcare system at the level of the facility, however, periodic evaluation of community PMTCT coverage must also be considered. This aspect of monitoring is most useful when utilization of facility-based services is low, but is also subject to bias if sampling is not random. In our study, for instance, we used random sampling, but excluded those sites and communities known not to have any PMTCT services at all. Thus, both our facility and community samples are optimistic representations.

We observed an overall rate of HIV-free survival that was higher than anticipated by our original modeling [6]. This could

**Table 3.** Characteristics of antenatal and postnatal care for children under 2 y of age in the PEARL Study.

| Characteristic | Côte D'Ivoire | | Cameroon | | Zambia | | South Africa | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | Value | *N* | Value | *N* | Value | *N* | Value | *N* | Value |
| **Participation in survey among eligible children** | 1,934 | 1,797 (92.9%) | 2,601 | 2,523 (97.0%) | 2,441 | 2,173 (89.0%) | 2,145 | 2,108 (98.3%) | 9,121 | 8,601 (94.3%) |
| **Gestational age at which ANC was sought (median [IQR])** | 1,549 | 4 (3, 6) | 2,415 | 5 (4, 6) | 2,107 | 5 (4, 6) | 1,864 | 4 (3, 6) | 7,935 | 5 (4, 6) |
| **Place of delivery** | 1,746 | | 2,506 | | 2,143 | | 2,083 | | 8,478 | |
| Institutional | | 1,527 (87.5%) | | 2,412 (96.2%) | | 1,892 (88.3%) | | 1,988 (95.4%) | | 7,819 (92.2%) |
| Home | | 219 (12.5%) | | 94 (3.8%) | | 251 (11.7%) | | 95 (4.6%) | | 659 (7.8%) |
| **Consulted ANC (self report)** | 1,782 | | 2,513 | | 2,160 | | 2,064 | | 8,519 | |
| Yes | | 1,619 (90.9%) | | 2,466 (98.1%) | | 2,147 (99.4%) | | 1,960 (95.0%) | | 8,192 (96.2%) |
| No | | 163 (9.1%) | | 47 (1.9%) | | 13 (0.6%) | | 104 (5.0%) | | 327 (3.8%) |
| **HIV test during ANC**[a] | 1,563 | | 2,418 | | 2,118 | | 1,868 | | 7,967 | |
| Yes | | 607 (38.8%) | | 2,242 (92.7%) | | 1,924 (90.8%) | | 1,748 (93.6%) | | 6,521 (81.9%) |
| No | | 688 (44.0%) | | 148 (6.1%) | | 189 (8.9%) | | 103 (5.5%) | | 1,128 (14.2%) |
| Do not know | | 268 (15.5%) | | 28 (1.1%) | | 5 (0.2%) | | 17 (0.9%) | | 318 (3.8%) |
| **HIV test result (self report)**[b] | 606 | | 2,235 | | 1,919 | | 1,741 | | 6,501 | |
| HIV-negative | | 555 (91.6%) | | 2,088 (93.4%) | | 1,639 (85.4%) | | 1,531 (87.9%) | | 5,813 (89.4%) |
| HIV-positive | | 2 (0.3%) | | 110 (4.9%) | | 242 (12.6%) | | 158 (9.1%) | | 512 (7.9%) |
| Do not know | | 47 (7.8%) | | 35 (1.6%) | | 36 (1.9%) | | 39 (2.2%) | | 157 (2.4%) |
| Refused to answer | | 2 (0.3%) | | 2 (0.1%) | | 2 (0.1%) | | 13 (0.7%) | | 19 (0.3%) |
| **Birth weight (median [IQR])** | 1,356 | 3,000 (2,700, 3,350) | 2,073 | 3,300 (3,000, 3,600) | 1,819 | 3,100 (2,800, 3,400) | 1,914 | 3,020 (2,700, 3,360) | 7,162 | 3,100 (2,800, 3,500) |
| >2,500 g | | 1,130 (83.3%) | | 1,863 (89.9%) | | 1,592 (87.5%) | | 1,611 (84.2%) | | 6,196 (86.5%) |
| ≤2,500 g | | 226 (16.7%) | | 210 (10.1%) | | 227 (12.5%) | | 303 (15.8%) | | 966 (13.5%) |
| **Measles vaccination (children>10 mo only)** | 741 | | 1,086 | | 1,165 | | 1,143 | | 4,135 | |
| Yes | | 452 (61.0%) | | 975 (89.8%) | | 1,004 (86.2%) | | 931 (81.5%) | | 3,362 (81.3%) |
| No | | 289 (39.0%) | | 111 (10.2%) | | 161 (13.8%) | | 212 (18.5%) | | 773 (18.7%) |
| **Bacillus Calmette–Guerin vaccination** | 1,412 | | 2,391 | | 2,144 | | 2,049 | | 7,996 | |
| Yes | | 1,259 (89.2%) | | 2,286 (95.6%) | | 1,975 (92.1%) | | 2,000 (97.6%) | | 7,520 (94.0%) |
| No | | 153 (10.8%) | | 105 (4.4%) | | 169 (7.9%) | | 49 (2.4%) | | 476 (6.0%) |
| **Child feeding method during first 6 mo (limited to children >6 mo of age)** | 1,250 | | 1,579 | | 1,429 | | 1,515 | | 5,773 | |
| Exclusive breastfeeding | | 1,170 (93.6%) | | 1,112 (70.4%) | | 1,217 (85.2%) | | 589 (38.9%) | | 4,088 (70.8%) |
| Mixed | | 73 (5.8%) | | 449 (28.4%) | | 201 (14.1%) | | 713 (47.1%) | | 1,436 (24.9%) |
| Formula | | 7 (0.6%) | | 18 (1.1%) | | 11 (0.8%) | | 213 (14.1%) | | 249 (4.3%) |

Data are number (percent) unless otherwise indicated.
[a]Restricted to those who consulted ANC.
[b]Restricted to those who consulted ANC and had an HIV test.
ANC, antenatal care.
doi:10.1371/journal.pmed.1001424.t003

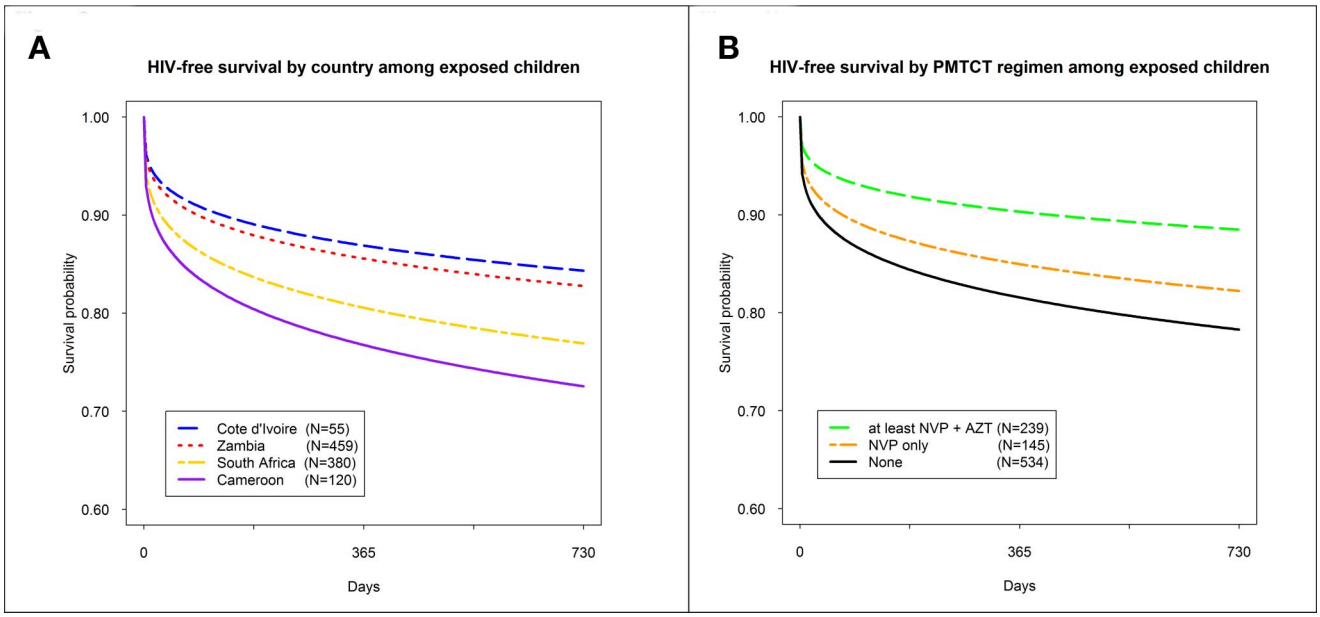


**Figure 2. HIV-free survival by country and by PMTCT regimen in the PEARL Study.** (A) HIV-free survival by country among exposes children; (B) HIV-free survival by PMTCT regimen among exposed children. AZT, zidovudine.
doi:10.1371/journal.pmed.1001424.g002

**Table 4.** Adjusted hazard of HIV infection or death among HIV-exposed children under 2 y of age in the PEARL Study ($n = 893$).

| Category | Characteristic | Adjusted HR (95%CI) |
|---|---|---|
| **Household** | **Toilet facilities in household** | |
| | Flush toilet or pit latrine | 1.0 |
| | No facility | 1.3 (0.8, 2.1) |
| | **Main material of floor** | |
| | Finished floor (cement/tiles/wood planks) | 1.0 |
| | Natural floor (earth/mud/dung/sand) | 0.9 (0.5, 1.5) |
| | **Does your household have a cell phone?** | |
| | Yes | 1.0 |
| | No | 1.2 (0.8, 1.8) |
| **Mother** | **Age at time of survey (years)** | 0.98 (0.94, 1.03) |
| | **School level** | |
| | No schooling or primary | 1.4 (0.9, 2.2) |
| | Secondary or higher | 1.0 |
| **Infant** | **PMTCT prophylactic regimen** | |
| | Zidovudine+NVP or HAART | 0.60 (0.34, 1.06) |
| | NVP only[a] | 0.68 (0.39, 1.19) |
| | None | 1.0 |
| | **Child feeding method during first 6 mo of life (restricted to children older than 6 mo; $n = 570$)** | |
| | Exclusive breastfeeding | 1.0 |
| | Mixed | 0.51 (0.19, 1.36) |
| | Formula | 0.42 (0.12, 1.44) |

Adjusted analyses were restricted to children with responses for each of the covariates of interest; all models adjust for country effects.
[a]Pairwise comparison of NVP only versus zidovudine+NVP or HAART adjusted HR: 0.88 (95% CI: 0.45, 1.72).
doi:10.1371/journal.pmed.1001424.t004

be explained by at least three sources of bias. First, all infant deaths in the communities may not have been identified in the survey. While every attempt was made to ensure complete ascertainment, including use of census enumerators from the local community and study instruments that queried child deaths in multiple ways, we cannot be certain that all deaths were captured. Second, because we chose to sample only those communities that had at least some access to PMTCT services, the included clusters may have not been representative of the population as a whole. Since better-resourced facilities will be the first to implement PMTCT services, our sampled patients may have had access to better services than others in their respective countries. This suspicion is supported by the finding that over 85% of respondents reported having delivered at a facility, which is higher than we expected from national figures [23–26] but which is perhaps explained by our selection of households based on their proximity to particular health facilities. A third potential bias of our study lies in its imperfect participant response rates. Although a response rate of 78% is typical for a community-based survey, households where occupants were not found or refused to participate may have been systematically different from those that were included. If a household with a death under 24 mo of age was more likely to refuse entry to study staff, this could have resulted in overestimation of HIV-free survival, especially if that refusal was more likely among households affected by HIV.

The questionnaire instruments and sampling methodology for this study drew heavily on the widely used DHS surveys [22]. The successful conduct of the PEARL Study, and others of similar design [27], demonstrates that with relatively minor modification, the DHS surveys could serve as a platform for widespread estimation of PMTCT program effectiveness. The most critical modification would require testing mothers for HIV antibodies (to ascertain infant HIV exposure) and, for those who are seropositive, testing their infants for infection. Some countries already perform HIV testing as part of their DHS surveys, but others do not, and this modification would have to be planned and resources for its implementation identified.

**Table 5.** Comparison of PMTCT service coverage in the community survey compared to results of a simultaneous survey in corresponding facilities [15].

| Country (Site) | Community | | Facility | |
|---|---|---|---|---|
| | *n*/*N* | Percent Coverage (95% CI) | *n*/*N* | Percent Coverage (95% CI) |
| Cameroon (1) | 23/41 | 56.1 (39.7–71.5) | 186/262 | 71.0 (65.1–76.4) |
| Cameroon (2) | 18/35 | 51.4 (34.0–68.6) | 61/80 | 76.3 (65.4–85.1) |
| Cameroon (3) | 20/28 | 71.4 (51.3–86.8) | 135/218 | 61.9 (55.1–68.4) |
| Côte D'Ivoire (1) | 0/18 | 0.0 (0.0–18.5) | 26/116 | 22.4 (15.2–31.1) |
| South Africa (1) | 51/124 | 41.1 (32.4–50.3) | 89/177 | 50.3 (42.7–57.9) |
| South Africa (2) | 53/91 | 58.2 (47.4–68.5) | 124/195 | 63.6 (56.4–70.3) |
| South Africa (3) | 12/35 | 34.3 (19.1–52.2) | 51/109 | 46.8 (37.2–56.6) |
| South Africa (4) | 34/116 | 29.3 (21.2–38.5) | 76/186 | 40.9 (33.7–48.3) |
| Zambia (1) | 14/30 | 46.7 (28.3–65.7) | 22/35 | 62.9 (44.9–78.5) |
| Zambia (2) | 57/105 | 54.3 (44.3–64.0) | 72/133 | 54.1 (45.3–62.8) |
| Zambia (3) | 6/14 | 42.9 (17.7–71.1) | 32/54 | 59.3 (45.0–72.4) |
| Zambia (4) | 22/53 | 41.5 (28.1–55.9) | 53/101 | 52.5 (42.3–62.5) |
| Zambia (5) | 25/49 | 51.0 (36.3–65.6) | 42/62 | 67.7 (54.7–79.1) |
| Zambia (6) | 38/85 | 44.7 (33.9–55.9) | 83/142 | 58.5 (49.9–66.7) |
| Zambia (7) | 44/77 | 57.1 (45.4–68.4) | 115/167 | 68.9 (61.2–75.8) |
| Zambia (8) | 23/49 | 46.9 (32.5–61.7) | 83/164 | 50.6 (42.7–58.5) |
| Crude[a] mean | 440/950 | 46.3 (43.2–49.5) | 1,250/2,201 | 56.8 (54.7–58.9) |

Coverage defined as the proportion of children born to HIV-positive mothers in whom any ARV drug is reported to have been used during pregnancy.
[a]Not adjusted to account for clustering within communities or countries.
doi:10.1371/journal.pmed.1001424.t005

Monitoring PMTCT programs remains a formidable challenge. Given the continually increasing emphasis on program scale-up, the unprecedented donor investment in affected countries, and the recent call by the World Health Organization and others for the virtual elimination of HIV in children by 2015, it is imperative that the international research and policy communities agree upon standard methodologies to measure the effectiveness of programs over time. Population-based evaluations of PMTCT effectiveness placed within periodic national health surveys could represent an important program monitoring resource.

## Acknowledgments

We would like to thank all the patients, doctors, nurses, community workers, and support staff who contributed to the completion of this study. We would also like to thank Andrew Boulle, Marc Bulterys, Latasha Treger, and Wendy Zijdel for their contributions to this study.

**Members of the PEARL Study Team:**

*Cameroon:* Pius Tih (Cameroon Baptist Health Convention Health Board [CBCHB]), Thomas Welty (CBCHB), Allison Spensley (Elizabeth Glaser Pediatric AIDS Foundation [EGPAF]), Christophe Grundmann (EGPAF), Catherine Wilfert (EGPAF and Duke University).

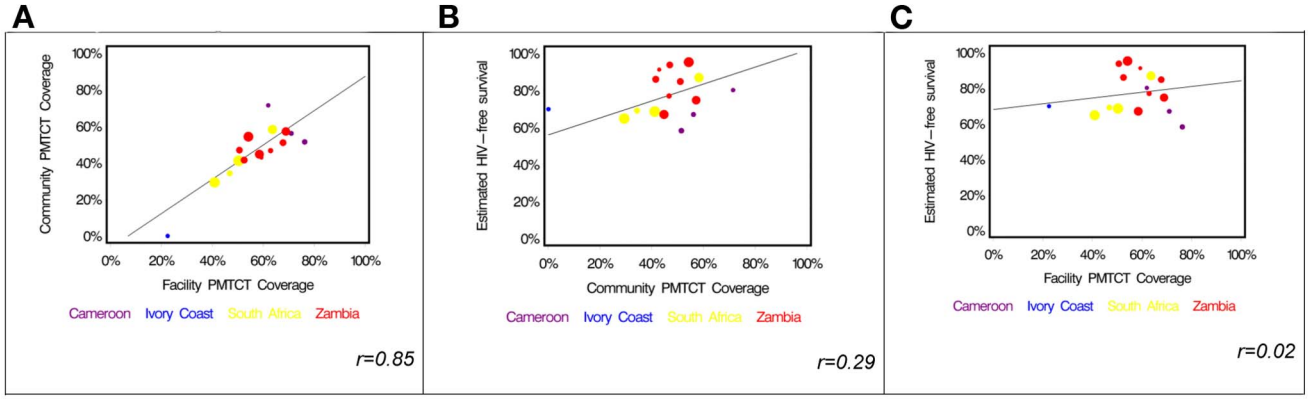


**Figure 3. Relationship among service coverage at the facility level, service coverage at the community level, and HIV-free survival in children under 2 y of age in the PEARL Study.** (A) Facility-based coverage versus community-based coverage; (B) community-based coverage versus HIV-free survival; (C) facility-based coverage versus HIV-free survival.
doi:10.1371/journal.pmed.1001424.g003

*Côte D'Ivoire:* Didier Ekouevi (Programme PAC-CI), Francois Dabis (Université Victor Segalen), Serge Kahon (Côte D'Ivoire Ministry of Health).

*South Africa:* David Coetzee (University of Cape Town [UCT]), Kathryn Stinson (UCT), Peter Smith (UCT), Andrew Boulle (UCT), Felicity Gopolang (UCT).

*Zambia:* Elizabeth M. Stringer; Jeffrey S. A. Stringer (Centre for Infectious Disease Research in Zambia [CIDRZ]), Benjamin H. Chi (CIDRZ), Namwinga Chintu (CIDRZ), Mark Giganti (CIDRZ), Maximilian Bweupe (Zambia Ministry of Health), Nande Putta (CIDRZ), Alain DeGroot (CIDRZ), Humphrey Mulenga (CIDRZ), Wendy Z. Mazimba (CIDRZ), Andrew Westfall (CIDRZ), Marc Bulterys (US Centers for Disease Control and Prevention–Zambia), Lawrence H. Marum (US Centers for Disease Control and Prevention–Zambia), Charles Cowan (Analytic Focus).

*US Centers for Disease Control and Prevention:* Tracy Creek (National Center for HIV, STD, and TB Prevention, Global AIDS Program).

*World Health Organization:* Nathan Shaffer (Department of HIV/AIDS).

## Author Contributions

Conceived and designed the experiments: JSAS KS PMT MJG DKE TLC TKW BHC CMW NS EMS FD DC. Performed the experiments: JSAS KS PMT MJG DKE TLC TKW BHC CMW NS EMS FD DC. Analyzed the data: JSAS MJG KS DC BC. Contributed reagents/materials/analysis tools: JSAS KS PMT MJG DKE TLC TKW BHC CMW NS EMS FD DC. Wrote the first draft of the manuscript: KS MG DC JSAS. Contributed to the writing of the manuscript: JSAS KS PMT MJG DKE TLC TKW BHC CMW NS EMS FD DC. ICMJE criteria for authorship read and met: JSAS KS PMT MJG DKE TLC TKW BHC CMW NS EMS FD DC. Agree with manuscript results and conclusions: JSAS KS PMT MJG DKE TLC TKW BHC CMW NS EMS FD DC. Enrolled patients: KS PMT DKE EMS.

## References


1. Joint United Nations Programme on HIV/AIDS (2010) Report on the global AIDS epidemic 2010. New York: Joint United Nations Programme on HIV/AIDS.
2. Joint United Nations Programme on HIV/AIDS (2011) Global plan towards the elimination of new HIV infections among children by 2015 and keeping their mothers alive. New York: Joint United Nations Programme on HIV/AIDS.
3. Lallemant M, Chang S, Cohen R, Pecoul B (2011) Pediatric HIV—A neglected disease? N Engl J Med 365: 581–583.
4. Cooper ER, Charurat M, Mofenson L, Hanson C, Pitt J, et al. (2002) Combination antiretroviral strategies for the treatment of pregnant HIV-1-infected women and prevention of perinatal HIV-1 transmission. J Acquir Immune Defic Syndr 29: 484–494.
5. Spensley A, Sripipatana T, Turner A, Hoblitzelle C, Robinson J, et al. (2009) Preventing mother-to-child transmission of HIV in resource-limited settings: The Elizabeth Glaser Pediatric AIDS Foundation experience. Am J Public Health 99: 631–337.
6. Stringer EM, Chi BH, Chintu N, Creek TL, Ekouevi DK, et al. (2008) Monitoring effectiveness of programmes to prevent mother-to-child HIV transmission in lower-income countries. Bull World Health Organ 86: 57–62.
7. Reithinger R, Megazzini K, Durako SJ, Harris DR, Vermund SH (2007) Monitoring and evaluation of programmes to prevent mother to child transmission of HIV in Africa. BMJ 334: 1143–1146.
8. World Health Organization (2011) Towards the elimination of mother-to-child transmission of HIV: Report of a WHO technical consultation. Available: http://www.who.int/hiv/pub/mtct/elimination_report/en/index.html. Accessed 20 November 2012.
9. Schouten EJ, Jahn A, Midiani D, Makombe SD, Mnthambala A, et al. (2001) Prevention of mother-to-child transmission of HIV and the health-related Millennium Development Goals: Time for a public health approach. Lancet 378: 282–284.
10. World Health Organization (2010) Antiretroviral drugs for treating pregnant women and preventing HIV infection in infants: recommendations for a public health approach. Geneva: World Health Organization.
11. World Health Organization (2012) Programmatic update: use of antiretroviral drugs for treating pregnant women and preventing HIV infection in infants. Geneva: World Health Organization.
12. Mate KS, Bennett B, Mphatswe W, Barker P, Rollins N (2009) Challenges for routine health system data management in a large public programme to prevent mother-to-child HIV transmission in South Africa. PLoS ONE 4: e5483. doi:10.1371/journal.pone.0005483
13. Ekouevi DK, Karcher S, Coffie PA (2011) Strengthening health systems through HIV monitoring and evaluation in sub-Saharan Africa. Curr Opin HIV AIDS 6: 245–250.
14. Hayashi C (2009) International guidance on methods to measure PMTCT impact [abstract]. 6th International AIDS Society Conference on HIV Pathogenesis, Treatment and Prevention; 17–20 July 2011; Rome, Italy.
15. Stringer EM, Ekouevi DK, Coetzee D, Tih P, Creek TL, et al. (2010) Coverage of nevirapine-based services to prevent mother-to-child transmission in four African countries. JAMA 304: 293–302.
16. Stringer JSA, Sinkala M, Goldenberg R, Vermund S, Acosta E (2003) Monitoring nevirapine-based programmes for prevention of mother-to-child transmission of HIV-1. Lancet 362: 667–667.
17. Rollins N, Mzolo S, Moodley T, Esterhuizen T, van Rooyen H (2009) Universal HIV testing of infants at immunization clinics: An acceptable and feasible approach for early infant diagnosis in high HIV prevalence settings. AIDS 23: 1851–1857.
18. Goga A, Dinh T-H, Jackson D, The SAPMTCTE Study Group (2010) Evaluation of the effectiveness of the national prevention of mother-to-child transmission (PMTCT) programme on infant HIV measured at six weeks postpartum in South Africa, 2010. Cape Town: South African Medical Research Council. Available: http://www.doh.gov.za/docs/reports/2012/pmtcteffectiveness.pdf. Accessed 16 November 2012.
19. Coffie PA, Kanhon SK, Tour H, Hapsatou T, Ettiegne-Traoré V, et al. (2011) Nevirapine for the prevention of mother-to-child transmission of HIV: A nationwide coverage survey in Côte d'Ivoire. J Acquir Immune Defic Syndr 57 (Suppl 1): S3–S8.
20. Ekouevi DK, Stringer E, Coetzee D, Tih P, Creek T, et al. (2012) Health facility characteristics and their relationship to coverage of PMTCT of HIV services across four African countries: The PEARL Study. PLoS ONE 7: e29823. doi:10.1371/journal.pone.0029823
21. Stinson K, Boulle A, Smith PJ, Stringer EM, Stringer JSA, et al. (2012) Coverage of the prevention of mother-to-child transmission programme in the Western Cape, South Africa using cord blood surveillance. J Acquir Immune Defic Syndr 60: 199–204.
22. Central Statistical Office, Ministry of Health in Zambia, Tropical Diseases Research Centre, University of Zambia, Macro International (2009) Zambia Demographic and Health Survey 2007. Calverton (Maryland): Central Statistical Office and Macro International. Available: http://www.measuredhs.com/pubs/pdf/FR211/FR211[revised-05-12-2009].pdf. Accessed 20 November 2012.
23. United Nations Children's Fund (2013) Statistics and monitoring: South Africa—delivery care coverage (%), institutional delivery (2006–2010). New York: United Nations Children's Fund.
24. United Nations Children's Fund (2013) Statistics and monitoring: Zambia—delivery care coverage (%), institutional delivery (2006–2010). New York: United Nations Children's Fund.
25. United Nations Children's Fund (2013) Statistics and monitoring: Cameroon—delivery care coverage (%), institutional delivery (2006–2010). New York: United Nations Children's Fund.
26. United Nations Children's Fund (2013) Statistics and monitoring: Côte d'Ivoire—delivery care coverage (%), institutional delivery (2006–2010). New York: United Nations Children's Fund.
27. Ruton H, Mugwaneza P, Shema N, Lyambabaje A, de Dieu Bizimana J, et al. (2012) HIV-free survival among nine- to 24-mo-old children born to HIV-positive mothers in the Rwandan national PMTCT programme: a community-based household survey. J Int AIDS Soc 15: 4.

## Editors' Summary

**Background.** For a pregnant woman who is HIV-positive, the discrepancy across the world in outlook for mother and child is stark. Mother-to-child transmission of HIV during pregnancy is now less than 1% in many high-income settings, but occurs much more often in low-income countries. Three interventions have a major impact on transmission of HIV to the baby: antiretroviral drugs, mode of delivery, and type of infant feeding. The latter two are complex, as the interventions commonly used in high-income countries (cesarean section if the maternal viral load is high; exclusive formula feeding) have their own risks in low-income settings. Minimizing the risks of transmitting HIV through effective drug regimes therefore becomes particularly important. Monitoring progress on reducing the incidence of mother-to-child HIV transmission is essential, but not always easy to achieve.

**Why Was This Study Done?** A research group led by Stringer and colleagues recently reported a study from four countries in Africa: Cameroon, Côte D'Ivoire, South Africa, and Zambia. The study showed that even in the health facility setting (e.g., hospitals and clinics), only half of infants whose mothers were HIV-positive received the minimum recommended drug treatment (one dose of nevirapine during labor) to prevent HIV transmission. Across the population of these countries, it is possible that fewer receive antiretroviral drugs, as the study did not include women who did not access health facilities. Therefore, the next stage of the study by this research group, reported here, involved going into the communities around these health facilities to find out how many infants under two years old had been exposed to HIV, whether they had received drugs to prevent transmission, and what proportion were alive and not infected with HIV at two years old.

**What Did the Researchers Do and Find?** The researchers tested all consenting women who had delivered a baby in the last two years in the surrounding communities. If the mother was found to be HIV-positive, then the infant was also tested for HIV. The researchers then calculated how many of the infants would be alive at two years and free of HIV infection. Most mothers (78%) agreed to testing for themselves and their infants. There were 7,985 children under two years of age in this study, of whom 13% had been born to an HIV-positive mother. Less than half (46%) of the HIV-positive mothers had received any drugs to prevent HIV transmission. Of the children with HIV-positive mothers, 11% were HIV-infected, 84% were not infected with HIV, and 5% had died. Overall, the researchers estimated that around 80% of these children would be alive at two years without HIV infection. This proportion differed non-significantly between the four countries (ranging from 73% to 84%). The researchers found higher rates of infant survival than they had expected and knew that they might have missed some infant deaths (e.g., if households with infant deaths were less likely to take part in the study).

The researchers found that their estimates of the proportion of HIV-positive mothers who received drugs to prevent transmission were fairly similar between their previous study, looking at health facilities, and this study of the surrounding communities. However, in 14 out of 16 comparisons, the estimate from the community was lower than that from the facility.

**What Do These Findings Mean?** This study shows that it would be possible to estimate how many infants are surviving free of HIV infection using a study based in the community, and that these estimates may be more accurate than those for studies based in health facilities. There are still a large proportion of HIV-positive mothers who are not receiving drugs to prevent transmission to the baby. The authors suggest that using two or three drugs to prevent HIV may help to reduce transmission.

There are already community surveys conducted in many low-income countries, but they have not included routine infant testing for HIV. It is now essential that organizations providing drugs, money, and infrastructure in this field consider more accurate means of monitoring incidence of HIV transmission from mother to infant, particularly at the community level.

**Additional Information** Please access these websites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.1001424.

- The World Health Organization has more information on mother-to-child transmission of HIV
- The United Nations Children's Fund has more information on the status of national PMTCT responses in the most affected countries

OPEN ACCESS Freely available online

PLOS ONE

# Measuring Coverage in MNCH: Testing the Validity of Women's Self-Report of Key Maternal and Newborn Health Interventions during the Peripartum Period in Mozambique

**Cynthia K. Stanton[1]*, Barbara Rawlins[2], Mary Drake[2], Matias dos Anjos[3], David Cantor[2], Lidia Chongo[4], Leonardo Chavane[3], Maria da Luz Vaz[3], Jim Ricca[2]**

**1** Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, **2** Maternal and Child Health Integrated Program, Washington, DC, United States of America, **3** Maternal and Child Health Integrated Program, Maputo, Mozambique, **4** Mozambique Ministry of Health, Maputo, Mozambique

## Abstract

***Background:*** As low-income countries strive to meet targets for Millennium Development Goals 4 and 5, there is growing need to track coverage and quality of high-impact peripartum interventions. At present, nationally representative household surveys conducted in low-income settings primarily measure contact with the health system, shedding little light on content or quality of care. The objective of this study is to validate the ability of women in Mozambique to report on facility-based care they and their newborns received during labor and one hour postpartum.

***Methods and Findings:*** The study involved household interviews with women in Mozambique whose births were observed eight to ten months previously as part of a survey of the quality of maternal and newborn care at government health facilities. Of 487 women whose births were observed and who agreed to a follow-up interview, 304 were interviewed (62.4%). The validity of 34 indicators was tested using two measures: area under receiver operator characteristic curve (AUC) and inflation factor (IF); 27 indicators had sufficient numbers for robust analysis, of which four met acceptability criteria for both (AUC $>0.6$ and $0.75<IF<1.25$). Two of these indicators are considered high demand and are recommended for incorporation into international survey programs: presence of a support person during labor/delivery and placement of the newborn skin to skin against the mother. Nine indicators met acceptability criteria for one of the validity measures. All 13 indicators are recommended for use in in-depth maternal/newborn health surveys.

***Conclusions:*** Women are able to report on some aspects of peripartum care. Larger studies may be able to validate some indicators that this study could not assess due to the sample size. Future qualitative research may assist in improving question formulation for some indicators. Studies of similar design in other low-income settings are needed to confirm these results.

**Citation:** Stanton CK, Rawlins B, Drake M, dos Anjos M, Cantor D, et al. (2013) Measuring Coverage in MNCH: Testing the Validity of Women's Self-Report of Key Maternal and Newborn Health Interventions during the Peripartum Period in Mozambique. PLoS ONE 8(5): e60694. doi:10.1371/journal.pone.0060694

**Editor:** Dr. Lucy Chappell, Kings College London, United Kingdom, in consultation with Carla AbouZahr, independent consultant, health statistics and policy.

**Received** October 16, 2012; **Accepted** February 25, 2013; **Published** May 7, 2013



**Funding:** This study was funded by the Maternal and Child Health Integrated Program, led by Jhpiego, via their funding from the United States Agency for International Development in Washington, D.C. Technical assistance was provided by the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF, with financial support from The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** AUC, area under receiver operating characteristic curve; DHS, Demographic and Health Survey/s; IF, inflation factor; MICS, Multiple Indicator Cluster Survey/s; MNCH, maternal, newborn, and child health.

* E-mail: cstanton@jhsph.edu

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection.*

## Introduction

As low-income countries strive to meet their targets for Millennium Development Goals 4 and 5, there is an urgent need to increase use and quality of maternal, newborn, and child health (MNCH) care services. To better monitor global trends, there is a strong demand to improve MNCH coverage indicators, spurred mainly by efforts in response to the Commission on Information and Accountability for Women's and Children's Health [1] and the Countdown to 2015 initiative [2]. Such indicators are also critical at the national level to provide actionable information regarding the adequacy of the content and quality of MNCH care to achieve desired health outcomes.

Presently, nationally representative household surveys conducted in low-income settings, such as the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS) track few indicators that measure care during the intrapartum and immediate postpartum periods. Typically, these are limited to: location of birth, qualification of birth attendant, and cesarean

section. The first two quantify contact with the health system, but provide no information on content of care. The validity of even these commonly reported survey indicators has not been assessed. However, national and international monitoring relies heavily on DHS and MICS survey data due to the inadequate state of routine health information system data in many low-income countries resulting from incomplete and irregular reporting, errors in manual calculation, incomplete or irregular data transmission from lower to higher levels of the health system, and compilation of subsets of indicators at irregular intervals by multiple divisions within the ministries of health.

A review of published literature from the past 30 years yielded multiple studies validating women's report of MNCH-related outcomes and health care in high-resource settings [3–20] but few from low-resource settings [21–26]. All studies used clinical records as the reference standard. The six studies from low-resource countries [21–26] all validated women's self-reports of obstetric complications comparing follow-up interviews with women who delivered in hospitals against clinical records. In the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, of which this paper is a part, two other articles address indicators for care in the peripartum period [27,28].

The objective of this study was to assess the validity of women's self-reports of selected health facility–based, peripartum MNCH interventions in Mozambique in two complementary ways: (1) via calculation of sensitivity and specificity, and area under receiver operator characteristic curve (AUC); and (2) estimation of the inflation factor (IF), which is the ratio of the prevalence of these interventions that would be obtained from a population-based survey, given the sensitivity and specificity from this study, and the indicator's true prevalence.

Mozambique was chosen for this study because it is a priority country for the Countdown to 2015 initiative and is typical of the other 74 countries which represent 95% of maternal and child deaths globally. Mozambique is similar particularly to other sub-Saharan African countries: maternal mortality ratio is high (500 maternal deaths per 100,000 live births) [29], the institutional birth rate is 54% [29], there are few births in private or non-governmental facilities, approximately 85% of women of reproductive age have achieved at most primary school education, and just over one third (38%) of its population resides in an urban area [30].

## Methods

### Ethical Review

This study was approved by the Mozambique National Bioethics Committee and the Institutional Review Board of the Johns Hopkins Bloomberg School of Public Health.

### Study Design

The study involved face-to-face interviews with women in Mozambique whose births were observed and documented as part of a government health facility survey of the Quality of MNCH Care (referred to subsequently as Quality of Care study) which was conducted from September to November 2011; results are available elsewhere [31]. Eight to ten months later, study participants were interviewed in their homes regarding the care they received during labor, delivery, and up to approximately one hour following birth. Data from the follow-up interviews were compared against data from the observations, which served as the reference standard. Because this validation study focused on facility-based care, the sample consisted of women who delivered in a health facility.

Indicators included in this study were selected based on three criteria: (1) evidence-based interventions during the peripartum period identified in the WHO Integrated Management of Pregnancy and Childbirth (IMPAC) manuals [32,33]; (2) important elements of the Mozambique humanization of birth program; and (3) events considered feasible for a woman to report on (e.g., inquiries or physical interventions by birth attendants versus the conduct of laboratory tests). Where questions of interest already exist in the DHS and/or MICS surveys, we tested the same or similar formulations of the questions.

The list of indicators selected for validity testing is presented in Table 1. Given constraints on the length of large-scale survey questionnaires, this list differentiates between high-demand indicators that are potential candidates for inclusion into the DHS/MICS questionnaires and indicators appropriate for more in-depth surveys of maternal and newborn health. The table also identifies indicators as beneficial or harmful. To note, some indicators are neither. These indicators represent two types of practice: (1) interventions that would require subjective information as to whether the intervention was medically indicated; for example, augmentation of labor or cesarean section; such judgments were beyond the scope of this study; and (2) indicators that are by their nature neither harmful or beneficial, such as choosing to deliver at a hospital versus a health center.

High-demand, evidence-based intervention indicators that were identified as potential candidates for inclusion in MICS/DHS surveys include some based on one question: HIV status checked, blood pressure measurement, urine testing (interventions performed during initial assessment of a woman in labor), presence of support person during labor or delivery, together with indicators related to newborn thermal care and active management of the third stage of labor. Three composite indicators (based on two or more questions, as defined in Table 1) were identified: thermal care for the newborn defined in two ways: and active management of the third stage of labor. In one of the papers in this Collection, Moran et al. also place importance on two of the indicators above, endorsing the use of an indicator on immediate drying of the newborn and recommending testing the skin to skin indicator [28].

### Sample Size

Mozambique has a total of 650 public maternity units. In the Quality of Care observational Study [31], 46 government health facilities (20 hospitals and 26 health centers) with an average of at least three births per day were selected randomly from a group of 122 government health facilities included in a national initiative to improve the quality of maternity care. This larger group of health facilities is responsible for 50%–60% of institutional births nationally. At the time of the observations, government-sponsored training programs targeting evidence-based interventions during the peripartum period were underway in 19 of the 46 hospitals sampled. A total of 525 births were observed, including women whose births resulted in a stillbirth or early neonatal death, and some cases of women who eventually delivered via cesarean. There were no refusals in the Quality of Care Study [31].

For the validation study, the anticipated prevalence of study indicators ranged from 20% to 80%, given that some reflect preventive interventions and should be nearly universal, whereas others represent harmful practices and should rarely, if ever, be performed. Assumptions required for sample size calculation included: 50% prevalence for all indicators, 60% sensitivity (with precision at $\pm$7.0%), 70% specificity (or 30% [1 $-$ specificity] with precision at $\pm$6.4%), and joint 90% confidence intervals [34] of 53%–67% for sensitivity and 23.6%–36.4% for (1 $-$ specificity). Variation across facilities in the prevalence of indicators potentially

**Table 1.** List of 34 indicators for which validation was attempted in this study.

| Question Already Exists in DHS/MICS Surveys | High-Demand Indicators Potentially Suggested for Use in DHS/MICS-Type Surveys | Indicator of Beneficial or Harmful Care | Indicator |
|---|---|---|---|
| | | | **Initial Assessment of a Woman in Labor** |
| | Yes | Beneficial | Percent of women asked about their HIV status |
| | | Beneficial | AMONG WOMEN WITH UNKNOWN HIV STATUS: Percent of women offered HIV test |
| | Yes | Beneficial | Percent of women who had their blood pressure taken |
| | Yes | Beneficial | Percent of women who were asked for a urine sample upon arrival at the health facility |
| | | | **Intra-partum Care** |
| | | Beneficial | Percent of women who were encouraged to have a companion present during labor/delivery |
| | Yes | Beneficial | Percent of women who had a companion present during labor or delivery |
| | | Beneficial | Percent of women who were encouraged to ambulate or move around during labor |
| | | Beneficial | Percent of women who were draped for privacy during labor |
| | | Harmful | Percent of women who were slapped, physically mistreated |
| | | Harmful | Percent of women who were shouted at or otherwise verbally mistreated |
| | | Neither | Percent of women who had their labor augmented with injection |
| | | Neither | Percent of births with cephalic presentation |
| | | Neither | Percent of women who had more than one health care provider assisting during birth |
| Yes | | Neither | Percent of women who delivered in a hospital (versus a health center) |
| Yes | | Neither | Percent of women who delivered by cesarean section |
| | | Neither | Percent of women with an instrumental birth (forceps, vacuum extraction) |
| | | Neither | Percent of women who delivered on their backs (lithotomy position) |
| | | Neither | Percent of women who received an episiotomy |
| | | | **Immediate Postpartum/Postnatal Care** |
| | Yes | Beneficial | *Element of postpartum hemorrhage prevention:* Percent of women who received a uterotonic within 3 (a few) minutes after birth of the baby |
| | Yes | Beneficial | *Element of postpartum hemorrhage prevention:* Percent of women who received controlled cord traction |
| | Yes | Beneficial | *Element of postpartum hemorrhage prevention:* Percent of women who received fundal massage after delivery of the placenta |
| | Yes | Beneficial | *Composite indicator of active management of the third stage of labor:* Percent of women who received uterotonic within a few minutes after birth of baby, controlled cord traction, AND fundal massage after delivery of placenta |
| | | Harmful | Percent of women who received fundal pressure before birth of baby |
| | | Harmful | Percent of women for whom the birth attendant manually explored uterus after birth of baby |
| | Yes | Beneficial | *Element of thermal care for the newborn:* Percent of newborns[a] who were immediately dried |
| | Yes | Beneficial | *Element of thermal care for the newborn:* Percent of newborns[a] placed skin to skin against the mother's chest |
| | Yes | Beneficial | *Element of thermal care for the newborn:* AMONG NEWBORNS PLACED SKIN TO SKIN: Percent of newborns[a] placed skin to skin on mother and covered with a cloth |
| | Yes | Beneficial | *Element of thermal care for the newborn:* AMONG NEWBORNS NOT PLACED SKIN TO SKIN: Percent of newborns[a] wrapped in a towel/cloth |
| | Yes | Beneficial | *Composite indicator 1 of thermal care for the newborn:* AMONG NEWBORNS NOT PLACED SKIN TO SKIN: Percent of newborns[a] immediately dried with a towel and wrapped with a cloth |
| | Yes | Beneficial | *Composite indicator 2 of thermal care for the newborn:* Percent of newborns[a] immediately dried with a towel, placed skin to skin on mother and covered with a cloth |
| Yes | | Beneficial | Percent of newborns[a] for whom breastfeeding was initiated within one hour of birth |
| | | Harmful | Percent of newborns[a] held upside down |
| | | Harmful | Percent of newborns[a] slapped |
| | | Harmful | Percent of newborns[a] bathed within one hour (i.e., bathing not delayed) |

[a]Any indicator referring to newborns or stillbirths refers to the second twin in cases of multiple birth.

doi:10.1371/journal.pone.0060694.t001

influenced by in-service training programs was accounted for by basing sample size on 50% prevalence. Based on these assumptions, a sample of 400 follow-up interviews was required, increased to 480 to allow for 20% loss to follow-up and refusals.

### Data Collection

In the health facilities selected for the Quality of Care Study [31], all deliveries were observed between 7am and 11pm over two to three days (depending on the volume of births in the facility), with the following caveats: a single observer could not observe more than two deliveries simultaneously, and women admitted into the emergency ward or taken immediately for cesarean section were not observed. Births were observed by nurses trained to observe maternity care using a standard validated checklist. Following birth and before hospital discharge, study participants who granted permission to be interviewed at home several months later were asked for their address and phone number and detailed landmarks to assist the interviewer in locating the household.

All women who provided a phone number were called during the interim to remind them of the upcoming interview. Twelve interviewers were recruited, 11 of whom had previously served as interviewers for the 2011 Mozambique DHS survey. They did not have a medical background and were not the observers in the Quality of Care Study [31]. Interviewer training included review of biological and health care-related events during the peripartum period. Interviewers attempted to relocate study participants at home. Interviews were conducted in Portuguese, and data were entered directly into Android platform tablet computers running Mobile Data Studio data entry software, with data entry validation checks. Figures S1 and S2 provide the wording in Portuguese and English, respectively, of the each question supporting each indicator. Supervisors reviewed data for consistency and completeness before transmitting it electronically to a central web server connected to a Structured Query Language server database. The database was monitored on an ongoing basis. Data were exported to Stata Version 11 for analysis.

### Analysis

Two-by-two tables were constructed. If any cell had fewer than five cases, the analysis was not performed. Sensitivity and specificity were estimated for each indicator for which there was adequate information, with uncertainty represented by 95% confidence intervals assuming a binomial distribution. AUC quantifies the performance of a diagnostic test (in this case, a woman's response to a survey question compared against the health facility–based reference standard). The Receiver Operating Characteristic Curve is produced when the sensitivity of a test is plotted against 1 − specificity of the test. The area under the curve can then be estimated. The most common usage of this statistic is to estimate AUC for multiple cut-off points resulting from a single diagnostic test or to compare results from different diagnostic tests. An AUC of 1.0 represents a perfect diagnostic test, whereas an AUC of 0.50 represents a random guess. For the purposes of this study, AUC was estimated based on dichotomous variables and is used as a means of comparing overall validity across multiple indicators [35]. These traditional measures are used to assess validity at the individual level.

The prevalence of an indicator that would be obtained from a population-based survey, given the sensitivity and specificity of that indicator resulting from this study, was estimated using the equation below from Vecchio [36].

$$\mathbf{Pr} = \mathbf{P} \times (\mathbf{SE} + \mathbf{SP} - \mathbf{1}) + (\mathbf{1} - \mathbf{SP})$$

In this equation, Pr is the estimate of survey-based prevalence, P is the hypothetical "true" prevalence in the population, SE is sensitivity, and SP is specificity. Results regarding estimated population-based prevalence rates for selected indicators are expressed in this paper as the inflation factor, that is, as an over- or under-estimation factor relative to the true rate. Several other papers in this supplement have utilized the ratio of Test to Actual Positives (TAP ratio) [37], which is the mathematical equivalent of estimated prevalence as calculated by Vecchio and the IF. The IF is the ratio of the estimated survey-based prevalence to the true population prevalence, as measured in the Quality of Care Study [31]. The IF is used to represent population-based validity.

For those indicators for which the analysis could be conducted, we defined acceptability criteria for validation as an AUC>0.60 or an IF between 0.75 and 1.25. There is no consensus on acceptable levels of any of the validation measures used in this study. The cut-offs for AUC and IF are subjective and were selected prior to data collection following discussion among the investigators taking into account the complexity of the questions, the lengthy recall period, and the fact that women were answering questions about events during labor or the immediate postpartum period. Due to intense constraints to lengthening the DHS or MICS questionnaires, our acceptability criteria for new indicators warranting incorporation into international survey programs are stricter and include only those identified as "high demand" in Table 1 and those which meet *both* acceptability criteria. Thus, indicators recommended for the DHS and MICS surveys are those that showed accurate reporting at both the individual and population levels.

## Results

Of the 525 women observed in the Quality of Care Study [31], 92.3% consented to a follow-up interview (Figure 1). Interviewers were able to locate the households of 64.7% of women who provided consent, and among those, succeeded in interviewing 96.5%. Thus, loss to follow-up was substantially higher than assumed (37.6% versus 20%), and the final sample of 304 interviewed women fell short of the target sample of 400 women. Of note, 0.8% of the sample of women (n = 4) died in the interim between birth and the follow-up interview and 1.4% refused the interview.

Table 2 presents the distribution of background characteristics for respondents to the follow-up study, participants in the Quality of Care Study [31], and a nationally representative sample of respondents of the 2008 MICS survey who delivered in a health facility. All provinces in Mozambique are represented in the Quality of Care [31] and follow-up studies. However, in the follow-up study, Maputo City and Manica are over-represented and Inhambane is under-represented relative to nationally representative data. Women in the follow-up study were somewhat more educated, urban, and younger than MICS survey respondents.

Tables 3 and 4 present validation results, the estimated survey-based prevalence of indicators, and the inflation factor for indicators with cell sizes sufficient for analysis. Table 3 includes recommended indicators based on one or both of our acceptability criteria. Table 4 includes indicators that cannot be recommended based these same criteria. In Tables 3 and 4 the estimated prevalence of indicators was based on their "true" prevalence as

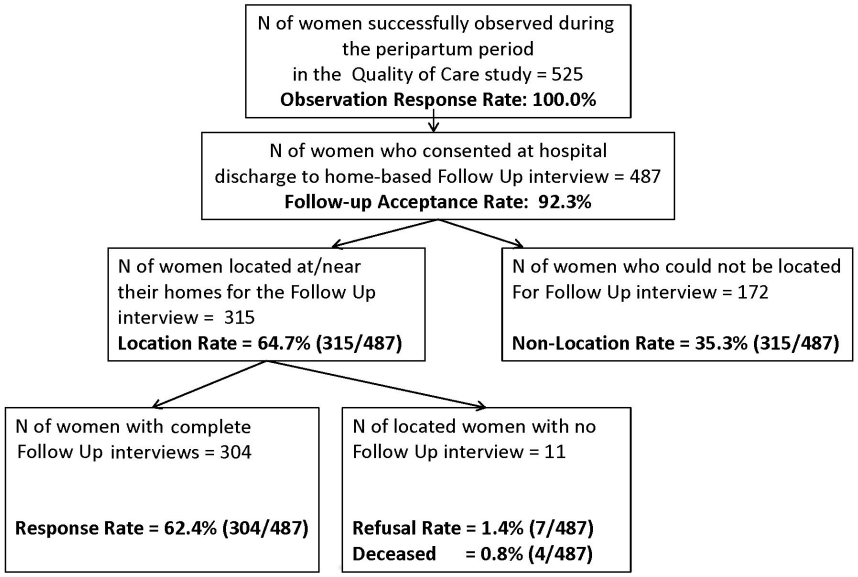


**Figure 1.** Response rates.
doi:10.1371/journal.pone.0060694.g001

measured in the entire sample of Quality of Care Study observations (i.e., not just those that were matched with follow-up interviews) and the sensitivity and specificity resulting from this study. "Don't know" responses constituted a small percentage of responses in the Quality of Care and follow-up Study and were treated as "No. " Most indicators were based on a single question, but some were composite indicators based on a combination of responses to two or more questions. Although 525 women consented to observation in the Quality of Care Study [31], differing numbers of women were observed during the different stages of the peripartum period, shown in the varying Ns for the true prevalence from the Quality of Care Study [31] and for AUC, which was based on cases of women observed during that indicator-specific period matched to their follow-up responses. In total, 27 of 34 indicators had sufficient numbers in all cells of the $2\times2$ table for the validation exercise. Table 5 includes indicators which could not be assessed due to small cell sizes.

Six indicators had AUC results of 0.60 or greater (Table 3). The most accurately reported responses were to the question on whether the woman delivered in a hospital versus a health center (0.88, 95% CI: 0.84–0.91); her labor was augmented (0.72, 95% CI: 0.51–0.93); fundal pressure was applied before birth of baby (0.67, 95% CI: 0.55–0.78); the newborn was placed skin to skin against the mother (0.65, 95% CI: 0.59–0.70); the woman was encouraged to have a companion during labor or birth (0.62, 95% CI: 0.55–0.69); and the woman had a support person present during labor or delivery (0.61, 95% CI: 0.55–0.66).

The other criterion of acceptable validity was an inflation factor between 0.75 and 1.25. Eleven indicators met this criterion. These were: the woman was encouraged to have a companion during labor/delivery (1.00); newborn was placed skin to skin against the mother (1.02); blood pressure was taken during initial assessment (0.98); the baby was cephalic presentation at birth (0.98); the woman delivered on her back (0.95); the woman had a support person present during labor or delivery (0.93); the newborn was immediately dried and wrapped in a towel (0.92); the woman was encouraged to ambulate or move around during labor (0.89); the woman delivered in a hospital versus health center (0.86); the newborn was immediately dried after birth (0.85); and the woman had fundal massage following delivery of the placenta (0.79).

Four indicators met the criteria for quality reporting based on both AUC and the IF. These were: the newborn was placed skin to skin against the mother, the woman identified her place of birth as a hospital versus a health center, the woman was encouraged to have a companion during labor or birth, and a support person was present during labor or birth.

Of the high-demand, evidence-based indicators tested, blood pressure measurement and two individual components of thermal care (baby was immediately dried and baby was placed skin to skin against the mother) met the criterion for inflation factor.

Two of the three indicators with an inflation factor of $>6.0$ had a true prevalence of 3% or less (labor augmentation, baby held upside down, fundal pressure applied before birth of baby). Such poor reporting is not surprising for low-prevalence indicators for which even small deviations from 100% in specificity can lead to extreme over-estimation in a survey. Indicators in Table 5 which could not be assessed due to cell size include: women was asked for a urine sample upon arrival at the health facility, cesarean section, instrumental birth, episiotomy, women physically mistreated, women verbally mistreated, and newborn slapped.

## Discussion

This study was able to test the validity of 27 key MNCH coverage indicators. We could not identify another published validation study of mothers' self-report of facility-based interventions delivered around the time of birth that compared women's reports against direct observation of the birth. Given the poor quality of clinical records in low-income settings, use of direct observation of care as the reference standard is a major strength of this study. Additional strengths include: the validity and reliability of clinical observers' observations were confirmed during training, observers were assigned to facilities other than their own places of work, and the home-based follow-up study closely mimicked the conditions of data collection in the DHS and MICS surveys.

Although the eight- to 10-month recall period for the follow-up study is shorter than that for the DHS (up to five years) and MICS

**Table 2.** Background characteristics: Percent distribution of respondents in the follow-up study; Women whose deliveries were observed in the Quality of Care Study [31]; and Women who delivered in a health facility in the 2008 MICS (a nationally representative sample of women of reproductive age).

| Socio-Demographic Characteristics | Follow-Up Survey (N = 304) | Mozambique Quality of Care Study (N = 525) | MICS 2008 (N = 3,011 Live Births in Two Years Preceding Survey) |
|---|---|---|---|
| **Age** | | | |
| 13–19 | 17.8 | 26.1 | (15–19) 17.6 |
| 20–24 | 33.9 | 29.9 | 29.3 |
| 25–29 | 22.4 | 21.0 | 23.0 |
| 30–34 | 15.1 | 14.1 | 16.3 |
| 35–39 | 8.6 | 7.0 | 10.5 |
| 40–44 | 2.0 | 0.8 | 2.5 |
| 45–49 | 0.3 | 0.0 | 0.8 |
| Don't know/ missing | 0.0 | 1.1 | 0.0 |
| **Level of Education** | | | |
| None | 12.8 | Not available | 23.8 |
| Primary | 55.6 | Not available | 62.2 |
| Secondary or higher | 31.6 | Not available | 13.4 |
| Don't know/ missing | 0.0 | Not available | 0.7 |
| **Marital Status** | | | |
| Married/ Cohabitating | 80.9 | Not available | Not available |
| Single | 12.5 | Not available | Not available |
| Divorced/ Separated | 6.6 | Not available | Not available |
| **Area of Residence** | | | |
| Urban | 54.0 | 55.0 [a] | 39.9 |
| Rural | 43.1 | 44.0 [a] | 60.1 |
| Missing | 3.0 | 1.0 [a] | 0.0 |
| **Region/Province** | | | |
| *North Region* | | | |
| Niassa | 9.9 | 8.4 | 7.9 |
| Cabo Delgado | 6.3 | 3.8 | 7.9 |
| Nampula | 15.5 | 14.1 | 18.3 |
| *Central Region* | | | 12.0 |
| Zambezia | 14.1 | 15.6 | 8.7 |
| Tete | 9.2 | 13.9 | 4.9 |
| Manica | 9.2 | 11.8 | 13.6 |
| Sofala | 8.9 | 8.8 | |
| *South Region* | | | |
| Inhambane | 2.0 | 2.1 | 6.4 |
| Gaza | 5.9 | 6.7 | 7.4 |
| Maputo Province | 4.6 | 4.8 | 6.9 |
| Maputo City | 12.2 | 10.1 | 5.9 |
| *Missing* | 0.3 | 0.0 | 0.0 |
| **Obstetric History** | | | |
| *Gravidity* | | | |
| 1 | 20.1 | 24.8 | Not available |
| 2–4 | 49.3 | 54.1 | Not available |
| 5+ | 30.5 | 21.1 | Not available |

[a]Based on 487 women observed in the Quality of Care Study [31] who gave consent to the follow-up interview.

doi:10.1371/journal.pone.0060694.t002

**Table 3.** Summary of validation results: Recommended indicators.

| Variable | Sensitivity (95% CI) of Follow Up Responses | Specificity (95% CI) of Follow Up Responses | N from Quality of Care Study | True Prevalence (%) among Those Observed in Quality Care Study | Population-Based Survey Estimate (%) Based on Sensitivity, Specificity | N for AUC Analysis | AUC | Inflation Factor (IF) | Recommend? Y/N (Selection Criteria: AUC>0.6 or 0.75<IF<1.25) |
|---|---|---|---|---|---|---|---|---|---|
| Woman was encouraged to have a companion during labor OR delivery | 0.49 (0.37–0.61) | 0.74 (0.67–0.81) | 525 | 33.6 | 33.8 | 212 | 0.62 (0.55–0.69) | 1.00 | Y (Both criteria) |
| Woman had a companion present during the labor OR delivery | 0.44 (0.33–0.54) | 0.78 (0.71–0.84) | 440 | 31.4 | 29.1 | 263 | 0.61 (0.55–0.66) | 0.93 | Y (Both criteria) |
| Woman delivered in a hospital (versus a health center) | 0.81 (0.75–0.87) | 0.94 (0.90–0.98) | 525 | 54.1 | 46.6 | 304 | 0.88 (0.84–0.91) | 0.86 | Y(Both criteria) |
| Newborn was placed skin to skin on mother's chest | 0.60 (0.52–0.69) | 0.69 (0.62–0.76) | 508 | 42.5 | 43.5 | 297 | 0.65 (0.59–0.70) | 1.02 | Y (Both criteria) |
| Woman had her labor augmented with injection | 0.67 (0.12–1.00) | 0.77 (0.72–0.83) | 525 | 2.4 | 23.6 | 264 | 0.72 (0.51–0.93) | 9.82 | Y (AUC only) |
| Woman received fundal pressure before birth of baby | 0.63 (0.39–0.87) | 0.70 (0.65–0.76) | 525 | 5.1 | 31.4 | 285 | 0.67 (0.55–0.78) | 6.16 | Y (AUC only) |
| Woman had her blood pressure taken | 0.63 (0.54–0.71) | 0.48 (0.38–0.59) | 378 | 59.3 | 58.2 | 212 | 0.57 (0.50–0.63) | 0.98 | Y (IF only) |
| Woman was encouraged to ambulate or move around during labor | 0.54 (0.46–0.63) | 0.63 (0.55–0.72) | 455 | 51.4 | 45.8 | 265 | 0.599 (0.53–0.65) | 0.89 | Y (IF only) |
| Woman delivered on her back | 0.92 (0.88–0.95) | 0 (0–0) | 507 | 96.4 | 91.9 | 297 | 0.46 (0.44–0.47) | 0.95 | Y (IF only) |
| Woman received fundal massage after delivery of the placenta | 0.56 (0.49–0.62) | 0.43 (0.31–0.54) | 507 | 71.0 | 56.1 | 289 | 0.49 (0.43–0.56) | 0.79 | Y (IF only) |
| Newborn dried and wrapped in a towel/cloth (among those not placed skin-to-skin with mother – composite indicator) | 0.59 (0.48–0.69) | 0.43 (0.29–0.56) | 312 | 63.1 | 58.3 | 144 | 0.51 (0.42–0.59) | 0.92 | Y (IF only) |
| Newborn immediately dried | 0.77 (0.72–0.82) | 0.31 (0.12–0.50) | 508 | 89.6 | 76.2 | 295 | 0.54 (0.44–0.63) | 0.85 | Y (IF only) |
| Birth with cephalic presentation | 0.95 (0.93–0.98) | 0 (0–0) | 507 | 97.4 | 95.5 | 288 | 0.56 (0.40–0.72) | 0.98 | Y (IF only) |

doi:10.1371/journal.pone.0060694.t003

**Table 4.** Summary of validation results: Indicators not recommended.

| Variable | Sensitivity (95% CI) of Follow Up Responses | Specificity (95% CI) of Follow Up Responses | N from Quality of Care Study | True Prevalence (%) among Those Observed in Quality of Care Study | Population-Based Survey Estimate (%) Based on Sensitivity, Specificity | N for AUC Analysis | AUC | Inflation Factor |
|---|---|---|---|---|---|---|---|---|
| Woman asked about her HIV status | 0.32 (0.24–0.40) | 0.64 (0.52–0.77) | 378 | 75.7 | 32.9 | 212 | 0.48 (0.41–0.55) | 0.43 |
| Woman offered HIV test (among women with unknown HIV status) | 0.40 (0.00–1.00) | 0.58 (0.47–0.70) | 378 | 1.9 | 13.8 | 82 | 0.50 (0.22–0.74) | 7.27 |
| Woman draped for privacy during labor | 0.60 (0.48–0.72) | 0.40 (0.33–0.47) | 455 | 23.3 | 59.8 | 261 | 0.52 (0.45–0.58) | 2.56 |
| Woman had more than one health care provider assisting during birth | 0.10 (0.02–0.17) | 0.87 (0.83–0.91) | 507 | 21.3 | 12.4 | 299 | 0.48 (0.44–0.53) | 0.58 |
| Woman received a uterotonic within 3 (a few) minutes after birth of baby | 0.38 (0.31–0.45) | 0.66 (0.58–0.75) | 507 | 56.8 | 36.3 | 289 | 0.52 (0.46–0.58) | 0.64 |
| Woman received controlled cord traction | 0.83 (0.77–0.89) | 0.25 (0.17–0.32) | 507 | 54.2 | 79.4 | 286 | 0.54 (0.49–0.59) | 1.47 |
| Active management of third stage of labor (composite indicator)[a] | 0.24 (0.15–0.32) | 0.81 (0.76–0.87) | 507 | 31.4 | 20.2 | 282 | 0.52 (0.47–0.58) | 0.64 |
| Newborn is placed skin to skin on mother covered with a cloth (composite) | 0.56 (0.42–0.69) | 0.21 (0.07–0.35) | 508 | 41.7 | 69.2 | 92 | 0.38 (0.29–0.48) | 1.66 |
| Newborn is wrapped in a towel/cloth | 0.82 (0.74–0.90) | 0.14 (0.02–0.26) | 508 | 41.5 | 84.2 | 130 | 0.54 (0.46–0.62) | 2.03 |
| Newborn immediately dried, placed skin to skin and covered with a towel/cloth (composite) | 0.40 (0.269–0.55) | 0.40 (0.25–0.55) | 508 | 38.2 | 52.5 | 92 | 0.40 (0.30–0.50) | 1.37 |
| Breastfeeding of newborn initiated within one hour of birth | 0.82 (0.72–0.92) | 0.25 (0.20–0.31) | 508 | 19.3 | 75.9 | 296 | 0.54 (0.48–0.59) | 3.94 |
| Newborn held upside down | 0.50 (0.05–0.95) | 0.40 (0.34–0.46) | 525 | 2.3 | 59.7 | 300 | 0.49 (0.31–0.68) | 25.94 |
| Newborn bathed within one hour (i.e., bathing not delayed) | 0.00 (0.00–0.00) | 0.99 (0.97–1.00) | 507 | 2.8 | 1.4 | 300 | 0.49 (0.49–0.50) | 0.49 |
| Woman for whom birth attendant manually explored uterus after birth of baby | 0.62 (0.46–0.79) | 0.42 (0.36–0.48) | 525 | 10.1 | 58.6 | 288 | 0.55 (0.46–0.63) | 5.80 |

[a]Uterotonic (injection) within a few minutes after birth, controlled cord traction AND fundal massage after delivery of placenta.

doi:10.1371/journal.pone.0060694.t004

**Table 5.** Summary of validation results: Indicators that could not be assessed due to small cell size*.

| Variable | N from Quality of Care Study | True Prevalence (%) among Those Observed in Quality of Care Study |
|---|---|---|
| Woman asked for urine sample upon arrival at the health facility | 525 | 1.6 |
| Woman delivered by cesarean section | 525 | 2.9 |
| Woman with an instrumental birth (forceps, vacuum extraction) | 525 | 2.9 |
| Woman received an episiotomy | 507 | 3.0 |
| Woman slapped, physically mistreated | 378 | 1.0 |
| Woman shouted at or otherwise verbally mistreated | 525 | 1.1 |
| Newborn slapped | 525 | 0.2 |

Sensitivity and specificity were not analyzed in cases where the n of any cell was <5.
doi:10.1371/journal.pone.0060694.t005

(up to two years) surveys, it represents a lengthy recall period not far from the average recall period of a MICS survey and is an improvement over validation studies interviewing women at facility discharge. Furthermore, lengthening the follow-up period would undoubtedly result in greater loss to follow up.

With some caveats, study data are representative of the population of women seeking facility-based care at birth in Mozambique. In the Quality of Care Study [31] all provinces in Mozambique were represented, though the sample somewhat over-represented urban and larger facilities. The response rate of the current study was 63% (lower than the 72% response rate of one published validation study using a lengthy recall period [27]), contributing to wider than anticipated confidence intervals for sensitivity and specificity of the indicators. Respondents to the follow-up study were somewhat more highly educated, younger, and more likely to be urban as compared to nationally representative MICS data. Some of this difference is likely due to the lower than expected response rate (as rural women, women without cell phone access and lower-educated women may be more mobile and difficult to locate). Thus, these results may overstate women's ability to self-report peripartum care if education is positively related to the accuracy of reporting, which was not explored in this paper. Among health facilities with surgical capacity, the sample of births observed missed women directly admitted through the emergency ward and prevented validation of delivery by cesarean section. It is unknown whether the observer's presence influenced women's ability to report on the care they received. Finally, our acceptability criterion of 0.60 for AUC may be considered low, and therefore a study limitation. However, given that so few key survey-based variables relied upon for international and national monitoring have been validated to date and that all of the variables assessed here occurred while women were in labor or shortly thereafter, these authors judged a minimum AUC of 0.60 to be acceptable.

Results from this study vary in comparison with other studies. In a study by Liu et al. in China [29], two indicators similar to those measured in this study included: blood pressure check and providing an HIV test. However, in the Liu study, questioning was about antenatal, not intrapartum, care. Liu et al. found higher sensitivity and lower specificity than the current study for both indicators. The differences may be explained by the fact that the prevalence for each of these screening tests was much higher in the Chinese study, possibly leading women to assume that they had always been performed, or that events during antenatal care were easier to be aware of and to recall than events when in labor.

Pacque et al. [38], whose study included home births, measured two indicators of immediate newborn care that are similar to indicators in the current study: "percent of mothers who breastfeed their infant within one hour of birth" and "percent of mothers whose newborn was immediately warmed (dried) and wrapped after birth." The sensitivity and specificity of the first indicator was relatively similar to results of the current study—0.88 compared with 0.82 for sensitivity in the current study with very low specificity in both studies. For the second indicator, Pacque et al. found lower sensitivity (0.64 versus 0.81) and higher specificity (0.52 versus 0.14) than the current study. It is unclear why the specificity in the current study is so low. The formulation of the question relating to "newborn immediately dried" in the two studies was slightly different.

Yoder et al. [39], who conducted a qualitative investigation of newborn and postnatal care among mothers in Malawi and Bangladesh detected problems with their understanding of (1) terminology related to newborn thermal care, (2) questions about the timing of events following birth, and (3) questions related to postnatal care health checks. The first two issues may have affected our study results as well as those of Pacque.

One of the indicators of greatest interest tested in this study was "woman received an injection within the first few minutes after birth," (i.e., received a prophylactic uterotonic against postpartum hemorrhage). Data on this important indicator are sparse and, disappointingly, this indicator cannot be recommended based on our results. It should be noted, however, that this and several other questions in the follow-up study were long, complex, and referred to specific time periods (e.g., before and after delivery of the baby; after delivery of the baby and before delivery of the placenta), all elements of questionnaire design best avoided. Qualitative research may assist in improving the formulation for these questions. We also hypothesize that educating women about key preventive interventions that they should expect to receive at birth may improve reporting by raising awareness of these interventions.

## Recommendations

We recommend that validation studies rely on two methods to assess the validity of selected indicators: sensitivity, specificity, and AUC, and the inflation factor or its mathematical equivalent. Two methods were selected because they are complementary and neither is sufficient alone if the goal is coverage monitoring via population-based surveys. Although high sensitivity and specificity are preferred for all indicators, knowing the estimated survey-based prevalence is also helpful, particularly for indicators of very low prevalence which are likely to be over-estimated without near-

perfect specificity. Likewise, in some cases, low sensitivity and specificity cancel out at the population level and may generate acceptable estimates for coverage monitoring purposes, even if not accurate for analysis at the individual level. An example from this study includes the indicator for newborns dried and wrapped in a towel (among those not placed skin-to-skin against the mother's chest), with sensitivity at 0.59, specificity at 0.43, and an inflation factor of 0.92.

We recommend that the 13 indicators which met acceptability criteria should be included in more detailed studies of maternal and newborn care, while noting that researchers should be cautious about measurement of low-prevalence indicators. For example, the prevalence of labor augmentation was only 2.4% in the Quality of Care Study [31]. Although its inflation factor was large, it scored well by the AUC criterion. It is important to monitor this intervention as it is frequently misused in other contexts, particularly in South Asia [40]. In such areas where labor augmentation is more prevalent, this indicator may be more accurately estimated in surveys. Among the high-demand, evidence-based indicators assessed, presence of a support person during labor/delivery and placement of the newborn skin to skin against the mother met both acceptability criteria and are therefore recommended for international survey programs.

The results of this study suggest that there are some aspects of peripartum care that women can report with adequate accuracy. Although Mozambique was selected as representative of high maternal and newborn mortality settings, additional studies with a modified design in other low-income settings are needed to confirm these results. Suggested modifications include: planning for a higher loss to follow up rate (e.g. at least 35%); where possible, improving question formulation for complex questions that relate to very specific time periods; and considering a design in which face-to-face interviews could be divided into two or three arms, with interviews at hospital discharge, at the woman's house after a lengthy interval and at the woman's house at the mid-point between interviews in the first and second arms of the study. Experience from this study suggests that a follow-up period of greater than approximately 12 months may not be feasible, given the effects of an extended period on the loss to follow-up rate. Cell phone penetration and population mobility will ultimately determine what is possible within a given context.

Such a design would allow one to determine if inaccuracy in reporting is due to recall or to the fact that women were never aware that certain procedures were performed. The three-arm design would also allow one to determine whether recall worsens over time. Depending on the results of such studies, data quality on peripartum care may be improved by restricting such detailed survey questions to births in the last two years.

## Supporting Information

**Checklist S1** Portuguese version of the questionnaire.
(DOCX)

**Checklist S2** English version of the questionnaire.
(DOCX)

## Acknowledgments

The authors would first of all like to thank the Mozambique Ministry of Health and the women and health care providers who participated in this study and made it possible. We would like to acknowledge the valuable contributions of Mercino Ombe and Michele Congolo in managing the data collection process and the initial data management. Gayane Yenokyan provided expert input on the sample size calculations. We further wish to acknowledge the important contributions of Jennifer Bryce in the initial study conceptualization. In addition, we would like to recognize the data collection teams for both the facility survey and household interviews.

## Author Contributions

Designed and managed the programs used on the tablets and mobile data system used in this study: DC. ICMJE criteria for authorship read and met: CKS BR MD MA DC LCho LCha MLV JR. Agree with manuscript results and conclusions: CKS BR MD MA DC LCho LCha MLV JR. Conceived and designed the experiments: CKS BR MD LCha MLV JR. Performed the experiments: BR MD MA JR DC. Analyzed the data: CKS BR JR. Wrote the paper: CKS BR MD JR DC LCho MLV LCha MA.

## References

1. World Health Organization (2012) Accountability for maternal, newborn and child survival: An update on progress in priority countries. WHO.
2. Bryce J, Requejo J (2012) Countdown to 2015 for maternal, newborn and child health: 2012 report. World Health Organization: Geneva.
3. Bennett A (1985) The birth of a first child: do women's reports change over time? Birth 12:153–158.
4. Bryant HE, Visser N, Love EJ (1989) Records, recall loss, and recall bias in pregnancy: a comparison of interview and medical records data of pregnant and postnatal women. Am J Public Health 79: 78–80.
5. Buka SL, Goldstein JM, Spartos E, Tsuang MT (2004) The retrospective measurement of prenatal and perinatal events: accuracy of maternal recall. Schizophr Res 71: 417–426.
6. Casey R, Rieckhoff M, Beebe SA, Pinto-Martin J (1992) Obstetric and perinatal events: the accuracy of maternal report. Clin Pediatr Phila 31: 200–204.
7. Delgado-Rodriguez M, Gomez-Olmedo M, Bueno-Cavanillas A, Garcia-Martin M, Galvez-Vargas R (1995) Recall bias in a case-control study of low birth weight. J Clin Epidemiol 48: 1133–1140.
8. Elkadry E, Kenton K, White P, Creech S, Brubaker L (2003) Do mothers remember key events during labor? Am J Obstet Gynecol 189: 195–200.
9. Elliott JP, Desch MS, Istwan NB, Rhea D, Collins AM et al (2010) The reliability of patient-reported pregnancy outcome data. Pop Health Manag 12: 27–32.
10. Githens PB, Glass CA, Sloan FA, Entman SS (1993) Maternal recall and medical records: an examination of events during pregnancy, childbirth, and early infancy. Birth 20: 136–141.
11. Hakim RB, Tielsch JM, See LC (1992) Agreement between maternal interview and medical record-based gestational age. Am J Epidem 136: 566–573.
12. Hewson D, Bennett A (1987) Childbirth research data: Medical records or women's reports? Am J Epidemiol 125: 484–491.
13. Lederman SA, Paxton A (1998) Maternal reporting of prepregnancy weight and birth outcome: Consistency and reliability of patient-reported pregnancy outcomes. Matern Child Health 2: 123–126.
14. Olson JE, Shu XO, Ross JA, Pendergrass T, Robison LL (1997) Medical record validation of maternally reported birth characteristics and pregnancy-related events: a report from the Children's Cancer Group. Am J Epidemiol 145: 58–67.
15. Quigley MA, Hockley C, Davidson LL (2007) Agreement between hospital records and maternal recall of mode of delivery: evidence from 12,391 deliveries in the UK Millennium Cohort Study. BJOG 114: 195–200.
16. Rice F, Lewis A, Harold G, van den Bree M, Boivin J et al (2006) Agrement between maternal report and antenatal records for a range of pre and perinatal factors: the influence of maternal and child characteristics. Early Hum Dev 83: 497–504.
17. Tate AR, Dezateux C, Cole TJ, Davidson L (2005) Factors affecting a mother's recall of her baby's birth weight. Int J Epidemiol 34: 688–695.
18. Tilley BC, Barnes AB, Bergstralh E, Labarthe D, Noller KL et al (1985) A comparison of pregnancy history recall and medical records. Am J Epidemiol 121: 269–281.
19. Troude P, L'Hélias LF, Raison-Boulley AM, Castel C, Pichon C et al (2008) Perinatal factors reported by mothers: do they agree with medical records? Eur J Epidemiol 23: 557–564.
20. Yawn BP, Suman VJ, Jacobsen SJ (1998) Maternal recall of distant pregnancy events. J Clin Epidemiol 51: 399–405.
21. Seoane G, Castrillo M, O'Rourke K (1998) A validation study of maternal self reports of obstetrical complications: implications for health surveys. Int J Gynecol Obstet 62: 229–236.
22. Filippi V, Rosnmans C, Ganhado T, Graham W, Alihonou E et al. (2000) Women's reports of severe (near-miss) obstetric complications in Benin. Stud Fam Plan 31: 309–324.

23. Stewart MK, Festin M (1995) Validation study of women's reporting and recall of major obstetric complications treated at the Philippine General Hospital. Int J Gyn Obs 48: 53–66.
24. Stewart MK, Stanton CK, Festin M, Jacobson N (1996) Issues in measuring maternal morbidity: lessons from the Philippines Safe Motherhood Survey Project. Stud Fam Plann 27: 29–35.
25. Ronsmans C, Achadi E, Cohen S, Zazri A (1997) Women's recall of obstetric complications in South Kalimantan, Indonesia. Stud Fam Plan 28: 203–214.
26. Sloan NL, Amoaful E, Arthur P, Winikoff B, Adjei S (2001) Validity of women's self-reported obstetric complications in rural Ghana. J Health Pop Nutr 19: 45–51.
27. Guo Y, Liu L, Li M, Yang L, Ju L, et al. (2013) Measuring coverage in MNCH: A validation study linking population survey derived coverage to maternal, newborn, and child health care records in rural China. PLoS ONE 8: e60762. doi:10.1371/journal.pone.0060762
28. Moran A, Kerber K, Sitrin D, Guenther T, Morrissey CS, et al. (2013) Measuring coverage in MNCH: Indicators for global tracking of newborn care. PLoS Med 10: e1001415. doi:10.1371/journal.pmed.1001415
29. United Nations (2013) Country profile: Mozambique. Available: http://data.un.org/CountryProfile.aspx?crName = Mozambique#Social.
30. Instituto Nacional de Estatística Moçambique, MEASURE DHS.ICF International (2012) MOÇAMBIQUE Inquérito Demográfico e de Saúde 2011. Available: http://www.measuredhs.com/pubs/pdf/PR14/PR14.pdf.
31. MCHIP (2012) Assessing the quality and humanization of maternity and ANC care in Mozambique. Available: http://www.mchip.net/node/1280. Accessed 6 December 2012.
32. World Health Organization (2000) Managing complications in pregnancy and childbirth: a guide for midwives and doctors. WHO.
33. World Health Organization (2003) Pregnancy, childbirth, postpartum and newborn care: a guide for essential practice. WHO.
34. Pepe M (2003) The statistical evaluation of medical tests for classification and prediction. Oxford University Press.
35. Zhou X, Obuchowski N, McClish D (2002) Statistical methods in diagnostic medicine. Wiley & Sons: New York, NY..
36. Vecchio TJ (1966) Predictive value of a single diagnostic test in unselected populations. N Engl J Med 274(21):1171–1173.
37. Campbell H, Biloglav Z, Rudan I (2008) Reducing bias from test misclassification in burden of disease studies: use of test to actual positive ratio—New test parameter. Croat Med J 49: 402–414.
38. Pacque M, Alayon S, Stallings R, Johnson K (2005) Saving Newborn Lives Initiative Validation of Newborn Health Indicators: Final Report. Macro International.
39. Yoder P, Rosato M, Mahmud R, Fort A, Rahman F, et al. (2010) Women's recall of delivery and neonatal care: A study of terms, concepts and survey questions. Calverton, Maryland, USA: ICF Macro.
40. Brhlikova P, Jeffrey P, Bhatia GP, Khurana S (2009) Intrapartum oxytocin (mis)use in South Asia. J Health Stud 2: 33–50.

PLOS ONE

# Measuring Coverage in MNCH: Validating Women's Self-Report of Emergency Cesarean Sections in Ghana and the Dominican Republic

**Özge Tunçalp[1]*, Cynthia Stanton[1], Arachu Castro[2], Richard Adanu[3], Marilyn Heymann[2], Kwame Adu-Bonsaffoh[4], Samantha R. Lattof[5], Ann Blanc[6], Ana Langer[5]**

**1** Johns Hopkins Bloomberg School of Public Health Baltimore, Maryland, United States of America, **2** Harvard Medical School, Harvard University, Boston, Massachusetts, United States of America, **3** School of Public Health, University of Ghana, Accra, Ghana, **4** Korle Bu Teaching Hospital, University of Ghana, Accra, Ghana, **5** Harvard School of Public Health, Harvard University, Boston, Massachusetts, United States of America, **6** Population Council, New York, New York, United States of America

## Abstract

***Background:*** Cesarean section is the only surgery for which we have nearly global population-based data. However, few surveys provide additional data related to cesarean sections. Given weaknesses in many health information systems, health planners in developing countries will likely rely on nationally representative surveys for the foreseeable future. The objective is to validate self-reported data on the emergency status of cesarean sections among women delivering in teaching hospitals in the capitals of two contrasting countries: Accra, Ghana and Santo Domingo, Dominican Republic (DR).

***Methods and Findings:*** This study compares hospital-based data, considered the reference standard, against women's self-report for two definitions of emergency cesarean section based on the timing of the decision to operate and the timing of the cesarean section relative to onset of labor. Hospital data were abstracted from individual medical records, and hospital discharge interviews were conducted with women who had undergone cesarean section in two hospitals. The study assessed sensitivity, specificity, and positive predictive value of responses to questions regarding emergency versus non-emergency cesarean section and estimated the percent of emergency cesarean sections that would be obtained from a survey, given the observed prevalence, sensitivity, and specificity from this study. Hospital data were matched with exit interviews for 659 women delivered via cesarean section for Ghana and 1,531 for the Dominican Republic. In Ghana and the Dominican Republic, sensitivity and specificity for emergency cesarean section defined by decision time were 79% and 82%, and 50% and 80%, respectively. The validity of emergency cesarean defined by operation time showed less favorable results than decision time in Ghana and slightly more favorable results in the Dominican Republic.

***Conclusions:*** Questions used in this study to identify emergency cesarean section are promising but insufficient to promote for inclusion in international survey questionnaires. Additional studies which confirm the accuracy of key facility-based indicators in advance of data collection and which use a longer recall period are warranted.

**Citation:** Tunçalp Ö, Stanton C, Castro A, Adanu R, Heymann M, et al. (2013) Measuring Coverage in MNCH: Validating Women's Self-Report of Emergency Cesarean Sections in Ghana and the Dominican Republic. PLoS ONE 8(5): e60761. doi:10.1371/journal.pone.0060761

**Editor:** Dr. Lucy Chappell, Kings College London, United Kingdom, in consultation with Carla AbouZahr, independent consultant, health statistics and policy

**Received** August 10, 2012; **Accepted** February 26, 2013; **Published** May 7, 2013



**Funding:** This study was funded by the Maternal Health Task Force at EngenderHealth (PSA-007-01 made to Stanton-Hill Research LLC and GMH-012-02 made to Harvard Medical School), via their funding from The Bill & Melinda Gates Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** ÖT and CS conducted this work under Stanton-Hill Research LLC, funded by Maternal Health Task Force at EngenderHealth. AC is a member of the Editorial Board of *PLOS Medicine*. All of the other authors declare no competing interests associated with the work supporting this manuscript.

**Abbreviations:** AUC, area under the receiver operating characteristic curve; DHS, Demographic and Health Survey/s; IF, inflation factor; IQR, interquartile range; MICS, Multiple Indicator Cluster Survey/s; OR, odds ratio.

* E-mail: otuncalp@jhsph.edu

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection.*

## Introduction

Cesarean section rates are rising in many low- and middle-income countries. For the first time, the World Health Organization's (WHO) *World Health Statistics 2012* reports a global cesarean section rate (16%) that exceeds the frequently used upper recommended limit of 15% [1,2]. Even in a low-income country like Bangladesh, recent data show the cesarean section rate increased from 3% to 12% between 2001 and 2010 [3]. Some middle-income Latin American and Asian countries report rates between 30% and 46%, and the cesarean section rate for upper-middle-income countries has surpassed that of high-income countries (31% and 28% respectively) [2]. Extreme socio-economic disparities in access to cesarean section exist within low-income countries as well. Women in the wealthiest households often have rates above 20%, whereas among the poorest households in many countries, cesarean section rates are less than one percent [4].

High and rising national rates indicate cause for concern, but provide no information on why or how these rates are changing or whether the increase is associated with any health gains. Likewise, very low rates, as seen in much of sub-Saharan Africa, provide no assurance that cesarean sections are serving women in greatest need.

Currently, cesarean section is the only surgery for which we have nearly global population-based data [2,5], as a result of the Demographic and Health Surveys (DHS) and UNICEF's Multiple Indicator Cluster Surveys (MICS). However, few surveys in low-income countries have incorporated questions that go beyond mode of delivery [6].

Although large-scale surveys provide the majority of global data on cesarean section, the question on cesarean section has not been validated. One study assessing the reliability of self-reported cesarean section rates in the DHS in six low-income countries showed that self-reported cesarean section rates were consistently higher than hospital-based cesarean section data applied to population-based births. However, in three quarters of the 31 sub-national observations assessed, hospital-based rates fell within 95% confidence intervals of the survey-based estimates. The differences between the two were often less than one percentage point [7]. It is not surprising that reliability of self-reported cesarean section is high since women are unlikely to forget or fabricate having undergone cesarean section.

In response to the need for more in-depth information related to cesarean section, the Maternal Health Task Force and the Child Health Epidemiology Reference Group at Johns Hopkins University sponsored a meeting in February 2010 in Baltimore, Maryland for maternal health researchers and program managers to propose an expanded list of indicators related to cesarean section [8]. Their top recommendation and the impetus for this study was the need to validate an indicator of emergency cesarean section which could be obtained from surveys of women of reproductive age.

Numerous definitions of emergency cesarean section exist, each of which identify a somewhat different group of women. A Medline search on emergency cesarean section from 1982 through 2007 by Schauberger and Chauhan [9] reported 28 studies which used at least 12 definitions based on varying criteria including: decision for cesarean section was made in labor, not scheduled, severe maternal/fetal complications (complications were specified in some but not all studies), immediate threat to mother/fetal life, timeliness from decision to incision or delivery, and various combinations of the above-mentioned criteria. In eight studies, no definition was provided. In a recent systematic review of cesarean section classification systems, Torloni and colleagues [10] identified nine classification systems (four based on indications and five based on various definitions of "urgency") that do not always use the term "emergency" but are similar in concept; for example: absolute maternal indication, obligatory, extreme emergency, and crash. In almost half of these studies, the classification system was designed for use in high-income countries with sophisticated record keeping.

This study, part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, has three objectives. The first is to validate self-reported data on emergency cesarean section among a sample of women who delivered by cesarean section. Two definitions of emergency cesarean section are tested. Cesarean section *by decision time* refers to a cesarean section for which the decision to perform the operation is made after the onset of labor. Cesarean section *by operation time* refers to a cesarean section performed after the onset of labor. Both indicators are dichotomous. We test two definitions because (1) in low-income settings, emergency cesarean section based on decision time may more accurately reflect the chronology of events than operation time given inadequate staffing and resources which often lead to delayed care; and (2) the timing of the operation relative to labor may be easier for women to report. To increase generalizability, large hospitals in two contrasting countries were selected for this study: Ghana and the Dominican Republic.

The second objective of the study is to estimate the percentage of emergency cesarean sections that would be obtained from a population-based survey, given the assessment of sensitivity and specificity from this study. The third objective is to identify characteristics of women who accurately report the status of their delivery by cesarean section.

### Contrasting Countries: Ghana Versus the Dominican Republic

In Ghana, maternal mortality is high at 378 per 100,000 births in 2007 [11]. Skilled attendance at birth in Ghana has increased over the past 20 years from 41% to 60% [12], most of which has occurred since 2003 when the Ghana Health Service began fee exemption for delivery services [13]. According to the Ghana DHS survey, the cesarean section rate increased from 4.5% to 6.4% between 1990 and 2005, with greater than 10-fold differentials in the rate by wealth quintile. As of 2005, the cesarean section rate for 40% of the population was under the WHO recommended minimum of 5%, and under 1% for the poorest quintile [12]. In contrast, the Dominican Republic is a country with nearly universal coverage of antenatal care and institutional delivery (>95%) [14], high maternal mortality compared to countries of similar income (179 per 100,000 live births between 2004 and 2008) [8], and rapidly increasing cesarean section rates. Between 1990 and 2006, the cesarean section rate in the Dominican Republic doubled from 22% to 44% [12].

## Methods

The study was conducted in two hospitals. Korle-Bu Hospital is one of the largest teaching hospitals in Ghana, situated in the capital, Accra. It is a tertiary referral center with 10,000 annual deliveries and a cesarean section rate of 30%. In the Dominican Republic, the study was conducted at the Maternity Hospital Nuestra Señora de la Altagracia, the national referral maternity hospital and a teaching hospital, in the capital, Santo Domingo. It is a tertiary level hospital with approximately 18,000 deliveries annually and a cesarean section rate of 33% [15]. Both of the facilities used partographs as routine practice during labor and delivery, although their use might not be consistent at times.

For the first objective, sensitivity, specificity, and positive predictive value of indicators related to caesarean section were calculated from women's responses to questions in the exit interview compared against hospital-based data (considered the reference standard). Area under the receiver operating characteristic curve (AUC) was estimated for each variable to compare overall validity for each indicator. Thus, research assistants undertook two data collection activities: (1) they abstracted data from the surgical and delivery room registers, individual case notes, and, occasionally, inquiries to the physician; and (2) they conducted face-to-face interviews just prior to hospital discharge of all women who had undergone cesarean section in each hospital. In Ghana, interviews were conducted in Twi and English. In the Dominican Republic, interviews were conducted in Spanish and Haitian Creole.

All women undergoing cesarean section were eligible for the study. Written informed consent was obtained upon admission to the hospital. Data were collected in Accra from June to August 2011, and from August to November 2011 in Santo Domingo. The following descriptive information was also collected: characteristics of the woman and the provider/patient communication

she experienced during her hospital stay (from the exit interview), and hospital characteristics such as the patient/provider ratio, volume of births, and deliveries by cesarean section (from hospital administrative data). The formulation of the questions assessed in this study is summarized in Box 1 (with Spanish version in Text S1), along with the two definitions of emergency cesarean section.

The method used for the second objective replicates methods used by Ronsmans and colleagues when assessing obstetric complications in Indonesia [16]. Using the equation below from Vecchio [17], sensitivity and specificity estimates from the validation study were used to calculate the prevalence of emergency cesarean section and other indicators of interest that would be obtained from a population-based survey, using the following equation:

$$\mathbf{Pr} = \mathbf{P} \times (\mathbf{SE} + \mathbf{SP} - 1) + (1 - \mathbf{SP}),$$

## Box 1. Questions Used in the Exit Interview

**GENERAL BACKGROUND QUESTIONS:**

- Previous to this pregnancy, have you ever had a cesarean section?
  - If YES, previous to this pregnancy, how many cesarean deliveries have you had?
- Other than that, have you ever had any surgery/operation in your pelvic area?
  - If YES, what was the surgery/operation?

**CURRENT DELIVERY:**

- Were you planning to deliver at Korle-Bu Teaching Hospital/Maternidad Altagracia?
  - If NO, where were you planning on delivering?
- Were you transferred from another facility?
  - IF YES, from where?
  - What was the reason for your transfer?
- What kind of delivery have you had here at Korle-Bu/ Maternidad Altagracia?
- What was the reason for your operation during your delivery? Choose the reason that best applies to your situation (includes a write-in option for other reasons)
- When was the decision made for you to have a cesarean/ operation?
- Whose idea was it for you to have a cesarean/operation? Please select the choice that best describes whose idea it was (includes a write-in option for other).
- Why did you request the cesarean?
- Who told you that you were having an operation/ −cesarean section?
- Did you go into labor by yourself/spontaneously?
- Did a health care provider give you a medication or drip to START your labor?
- Did you get a cesarean section BEFORE your labor pains began?
- How many weeks were you when you delivered?
- Was the baby born early? Was the baby born on time (at term)?

**EMERGENCY CESAREAN SECTION QUESTIONS**
**Decision Time:**

- When was the decision made for you to have a cesarean/ operation?
  - During antenatal clinic visits
  - Before the labor pains began
  - After labor pains began
  - Don't know

**Operation Time:**

- Did you go into labor by yourself/spontaneously?
  - Yes
  - No
  - Don't Know
- Did a health care provider give you a medication or drip to START your labor?
  - Yes
  - No
  - Don't Know
- Did you get a cesarean section BEFORE your labor pains began?
  - Yes
  - No
  - Don't Know

**EMERGENCY CESAREAN SECTION DEFINITIONS**

- Emergency Cesarean Section defined by Decision Time:
  - When was the decision made for you to have a cesarean?

    Answer: After labor pains began
- Emergency Cesarean Section defined by Operation Time:
  - Did you go into labor by yourself/spontaneously?

    Answer: Yes
  - Did a health care provider give you a medication or drip to START your labor?

    Answer: Yes/No (depending on the answer to the first question)
  - Did you get a cesarean section BEFORE your labor pains began?

    Answer: No

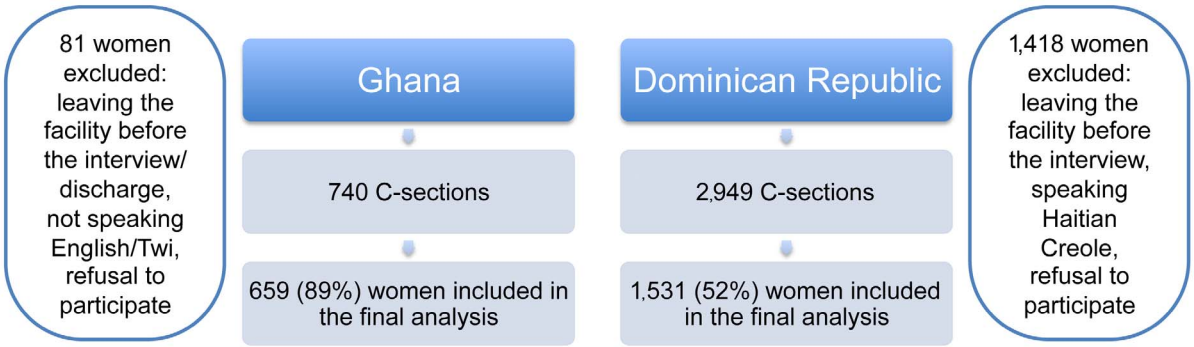


**Figure 1. Flowchart of participation in Ghana and the Dominican Republic.**
doi:10.1371/journal.pone.0060761.g001

where Pr is the estimate of survey-based prevalence, P is the hypothetical "true" prevalence in the population, SE is sensitivity, and SP is specificity. Results regarding the estimated population-based emergency cesarean section rate were expressed as an inflation factor (IF), that is, as an over- or under-estimation factor relative to the "true" rate. This equation is the mathematical equivalent of the ratio of Test to Actual Positives (TAP ratio) [18], which has been utilized in a number of papers in this Collection. Of note, two assumptions underlie this calculation:

1. Self-report of cesarean section is valid. Thus, the sample is restricted to women who had undergone cesarean. This sample is appropriate for a validation study of emergency cesarean section because in a survey questionnaire, only women who had delivered by cesarean would be asked questions regarding the characteristics of the procedure.
2. Results from interviews at hospital discharge are generalizable to survey-based responses about events up to three years prior to the survey; that is, we assume that poor recall of an event as major as pelvic surgery is low.

Unadjusted logistic regression was used to assess the third objective, with accurate self-report of emergency cesarean section as the dependent variable and women's characteristics as the independent variables.

Sample size for the study was calculated before the data collection and was based on an assumption of 80% sensitivity, a Type 1 error at 5% for a two-tailed test, ±5% precision and the true proportion of cesarean sections that are emergency cesarean sections at 30% in Ghana and 5% in Dominican Republic. Based on these assumptions, the target sample size was 450 women who had been delivered by cesarean section in the Ghanaian site and 1,460 in the Dominican Republic.

Ethical approval for the study in Ghana was provided by the Institutional Review Board of Korle-Bu Teaching Hospital, University of Ghana Medical School, College of Health Sciences, Accra, Ghana. The Harvard School of Public Health and the National Council of Bioethics of the Dominican Republic approved the study in the Dominican Republic.

## Results

### Study Population

In Ghana, 740 women were delivered by cesarean section during the study period, of which 89% (659 women) were interviewed prior to hospital discharge (Figure 1). Of 81 exit interviews that were missed, 64 were women who left the hospital before the interview, 15 did not speak English or Twi, one left the facility before her discharge, and one refused participation. The median number of days between the operation and the interview was 3 days (interquartile range [IQR] 2–3). In the Dominican Republic, 2,949 women were delivered by cesarean section during the study, of which 52% (1531 women) were interviewed before hospital discharge and included in the analysis. Twelve women refused participation, 92 women interviewed in Haitian Creole were excluded, and the rest (1,314 women) left the hospital before they could be interviewed. Factors that limited the Dominican Republic team's ability to invite the women to participate in the study included the lack of availability of medical files for review, the movement of patients within the hospital, and the early discharge practices of the hospital. The median number of days between the operation and the interview was one day (IQR 1–2).

### Characteristics of the Women in Ghana and the Dominican Republic

Table 1 presents the characteristics of the two study populations (as reported in exit interviews), which differ substantially. In Ghana as compared to the Dominican Republic, mean age and parity were higher and education was lower. The population in Ghana was more rural than in the Dominican Republic, as expected, and the distribution by religion varied. Proportions of women with previous cesarean sections were similar across the two populations (35% in Ghana and 38% in the Dominican Republic). Given the large difference in cesarean section rates in the two countries, a lower previous cesarean section rate in Ghana might have been expected. The rate of other pelvic surgery was low in both samples (4.7% in Ghana and 2.7% in the Dominican Republic).

The data on delivery plan and referral status best illustrate the difference in case mix between the two hospitals. In the Dominican Republic, nearly four fifths of women planned on delivering at Altagracia Hospital and one quarter of women report being referred to this hospital. In contrast, in Ghana 42% of women planned on delivering at Korle Bu Hospital and over three quarters of women were referred and transferred. Although both hospitals are large urban teaching hospitals, Korle Bu appears to be used more frequently as a referral hospital than Altagracia, suggesting that complicated deliveries likely represent a higher percentage of deliveries in Korle Bu than in Altagracia. This may partially explain the similar rates of previous cesarean sections in the two hospitals.

In both populations, 100% of women reported having undergone a cesarean section. In Ghana, 57% of women reported that the decision for delivery via cesarean was made before the onset of labor (nearly half of which during antenatal care visits); 42% reported that the decision was made after the onset of labor.

**Table 1**. Background characteristics of the study population based on women's exit interviews.

| Sociodemographic Characteristics | Ghana (N = 659) | Dominican Republic (N = 1,531) | *p*-Value[a] |
|---|---|---|---|
| **Age, years** | | | 0.0001 |
| 15–19 | 16 (2.4) | 416 (27.2) | |
| 20–24 | 89 (13.5) | 501 (32.7) | |
| 25–29 | 176 (26.7) | 332 (21.7) | |
| 30–34 | 219 (33.2) | 188 (12.3) | |
| 35–39 | 127 (19.3) | 76 (4.9) | |
| 40–49 | 32 (4.9) | 18 (1.2) | |
| **Education** | | | 0.0001 |
| None | 48 (7.3) | 26 (1.7) | |
| Primary | 87 (13.2) | 402 (26.3) | |
| Secondary | 433 (65.7) | 861 (56.2) | |
| Tertiary | 91 (13.8) | 242 (15.8) | |
| **Religion** | | | 0.0001 |
| Christian | 566 (85.9) | 1,065 (69.6) | |
| Muslim | 92 (13.9) | 0 (0.0) | |
| Other | 1 (0.2) | 22 (1.4) | |
| No religion | 0 (0.0) | 444 (29.0) | |
| **Marital Status** | | | 0.692 |
| Married/cohabitation | 556 (84.4) | 1,290 (84.3) | |
| Single | 100 (15.1) | 229 (15.0) | |
| Divorced/separated | 3 (0.5) | 12 (0.7) | |
| **Residence** | | | 0.0001 |
| Urban | 520 (78.9) | 1,396 (91.2) | |
| Rural | 136 (20.6) | 131 (8.6) | |
| Don't know | 3 (0.5) | 4 (0.2) | |
| **Obstetric history** | | | |
| *Number of pregnancies, mean (SD)* | 2.81 (1.59) | 3.01 (1.83) | 0.0158 |
| *Number of previous deliveries, mean (SD)* | 2.34 (1.39) | 2.12 (1.32) | 0.0003 |
| *Previous cesarean section* | | | 0.167 |
| No | 426 (64.6) | 942 (61.5) | |
| Yes | 233 (35.4) | 589 (38.5) | |
| *Previous pelvic surgery (other than cesarean section)* | | | 0.021 |
| No | 628 (95.3) | 1,490 (97.3) | |
| Yes | 31 (4.7) | 41 (2.7) | |
| **Current pregnancy** | | | |
| *Gestational age at delivery (weeks)* | | | 0.001 |
| <35 | 21 (3.2) | 141 (9.2) | |
| 35–37 | 52 (7.9) | 262 (17.1) | |
| 38–40 | 130 (19.7) | 797 (52.1) | |
| 41–43 | 38 (5.8) | 229 (14.9) | |
| Don't know | 418 (63.4) | 102 (6.7) | |
| *Gestational age in terms* | | | 0.001 |
| Preterm | 180 (27.3) | 263 (17.2) | |
| Term | 318 (48.2) | 1,230 (80.3) | |
| Post-term | 125 (19.1) | 13 (0.9) | |
| Don't know | 36 (5.4) | 25 (1.6) | |
| *Multiple pregnancy* | | | 0.24 |
| Single | 623 (94.5) | 1,465 (95.7) | |
| Multiple | 36 (5.5) | 66 (4.3) | |

**Table 1.** Cont.

| Sociodemographic Characteristics | Ghana (N = 659) | Dominican Republic (N = 1,531) | *p*-Value[a] |
|---|---|---|---|
| *Delivery plan* | | | 0.001 |
| Home | 14 (2.1) | 0 (0.0) | |
| Study hospital | 280 (42.5) | 1,201 (78.5) | |
| Other facility | 365 (55.4) | 330 (21.5) | |
| *Referral status* | | | 0.001 |
| No | 151 (22.9) | 1,139 (74.4) | |
| Yes | 508 (77.1) | 392 (25.6) | |
| **Cesarean-section indicators** | | | |
| *Reporting of cesarean section* | | | n/a |
| No | 0 (0) | 0 (0) | |
| Yes | 659 (100) | 1,531 (100) | |
| *Reporting of time of cesarean section decision* | | | 0.001 |
| During antenatal visits | 208 (31.6) | 751 (49.1) | |
| Before labor | 169 (25.6) | 165 (10.8) | |
| After onset of labor | 276 (41.9) | 597 (38.9) | |
| Don't know | 6 (0.9) | 18 (1.2) | |
| *Reporting of time of cesarean section* | | | 0.001 |
| Spontaneous labor | 328 (49.8) | 1,047 (68.4) | |
| Induced labor | 35 (5.3) | 5 (0.33) | |
| Cesarean section before labor | 278 (42.2) | 359 (23.4) | |
| Don't know | 18 (2.7) | 120 (7.8) | |
| **Communication** | | | |
| *Cesarean section decision maker* | | | 0.001 |
| The doctor | 591 (89.7) | 1,510 (98.6) | |
| The woman | 35 (5.3) | 10 (0.6) | |
| Other | 6 (0.9) | 0 (0) | |
| Don't know | 27 (4.1) | 11 (0.7) | |
| *Cesarean section information* | | | 0.001 |
| Doctor | 571 (86.6) | 1,393 (96.8) | |
| Nurse/midwife | 46 (7.0) | 10 (0.7) | |
| No one | 38 (5.8) | 33 (2.3) | |
| Other | 4 (0.6) | 3 (0.2) | |

[a]Pearson's Chi-square tests and/or Yates correction for continuity (when necessary) are used for bivariate and categorical variables. T-tests are used for continuous variables.

doi:10.1371/journal.pone.0060761.t001

In the Dominican Republic, women reported that the decision for a cesarean section was made before the onset of labor in 60% of cases, of which more than four fifths were made during antenatal care visits; in 39% of the cases, the decision was made after the onset of labor.

According to women's report, the onset of labor also varied across the two populations. Among women in Ghana, half of women had a spontaneous onset of labor, 5.3% of women had their labor induced, and 42% of women underwent cesarean section before the onset of labor. In the Dominican Republic, two-thirds of women had a spontaneous onset of labor, there were almost no inductions (0.3%), and 23% underwent cesarean section prior to the onset of labor.

The majority of the women in both of the study populations reported that the decision to perform a cesarean section was made by a doctor and that the doctor informed them about this decision.

## Validation of Cesarean Section Indicators

Table 2 presents the prevalence, sensitivity, specificity, positive predictive value, and AUC and IF for emergency cesarean defined by decision time and by operation time relative to the onset of labor. It should be noted that information on these indicators was mainly collected from the patient files in both of our study settings, represented as a percentage within the study population. For ten cases in Ghana (1.5%) and 36 cases in the Dominican Republic (2.3%), this was supplemented by information requested from the medical staff.

In Ghana, emergency cesarean section defined by decision time shows sensitivity and specificity of approximately 80% (79% and 82%, respectively) and an IF of 1.06. Emergency cesarean section defined by decision time in the Dominican Republic had similar specificity (80%), yet lower sensitivity (50%), leading to an IF suggesting almost 40% underestimation in a population-based survey (0.61). Given the higher prevalence of this indicator in

**Table 2.** Validation assessment of cesarean section indicators.

| Ghana (N = 659) | % (Within the Study Population)* | Sensitivity (95% CI) | Specificity (95% CI) | AUC Point Estimate (95% CI) | Positive Predictive Value (%) | Population-Based Survey Estimate (%) | IF |
|---|---|---|---|---|---|---|---|
| Previous cesarean section | 36.9 | 95 (91–97) | 98 (97–99) | 0.96 (0.95–0.98) | 97 | 37 | 0.98 |
| Emergency cesarean section by decision time | 39.8 | 79 (73–83) | 82 (78–85) | 0.80 (0.77–0.83) | 74 | 42 | 1.06 |
| Emergency cesarean section by the operation time (single question) | 48.8 | 84 (80–88) | 68 (63–73) | 0.79 (0.72–0.79) | 74 | 57 | 1.18 |
| Emergency cesarean section by the operation time (three–question algorithm) | 49.4 | 84 (80–88) | 70 (65–75) | 0.77 (0.74–0.80) | 72 | 57 | 1.15 |
| Spontaneous labor | 42.2 | 85 (80–89) | 73 (69–78) | 0.79 (0.76–0.82) | 70 | 51 | 1.21 |
| Induced labor | 7.2 | 37 (23–51) | 97 (96–98) | 0.67 (0.59–0.74) | 49 | 5 | 0.76 |
| Cesarean section before labor | 50.5 | 70 (65–75) | 84 (80–88) | 0.77 (0.74–0.80) | 82 | 43 | 0.86 |
| **Dominican Republic (N = 1,531)** | | | | | | | |
| Previous cesarean section | 38.2 | 96 (94–98) | 97 (96–98) | 0.96 (0.96–0.98) | 95 | 38.5 | 1.01 |
| Emergency cesarean Section by decision time | 64.7 | 50 (47–53) | 80 (77–83) | 0.65 (0.62–0.67) | 82 | 39 | 0.61 |
| Emergency cesarean section by the operation time (single question) | 66.0 | 83 (80–85) | 53 (48–57) | 0.67 (0.65–0.70) | 79 | 71 | 1.07 |
| Emergency cesarean section by the operation time (three question algorithm) | 67.0 | 88 (86–90) | 53 (48–57) | 0.70 (0.68–0.73) | 77 | 74 | 1.11 |
| Spontaneous labor | 62.0 | 89 (87–91) | 51 (46–55) | 0.70 (0.68–0.72) | 75 | 74 | 1.19 |
| Induced labor | 5.0 | 1.4 (−24–27) | 99.7 (0.99–1) | 0.50 (0.49–0.52) | 20 | 0.3 | 0.07 |
| Caesarean section before labor | 33.0 | 53 (48–57) | 88 (86–90) | 0.70 (0.68–0.73) | 68 | 26 | 0.77 |

*The percentages used in this column are based on the data collected from the reference standard (patient records).

doi:10.1371/journal.pone.0060761.t002

Dominican Republic, positive predictive value was higher in Dominican Republic than in Ghana (82% versus 74%).

Emergency cesarean section by operation time had sensitivity of 84%, specificity of 68%, and IF of 1.18 in Ghana; in the DR, sensitivity was 83%, specificity 53% and the IF was 1.07. Positive predictive value varied between 72% and 79% for both of the settings, slightly higher in the Dominican Republic. For exploratory purposes, the definition of operation time was refined by using the responses to two additional survey questions, which first specified that the woman did experience labor. Thus, women with emergency cesarean section were defined as: (1) those who reported a spontaneous onset of labor and that their cesarean section did not occur before the onset of labor, and (2) those who reported that their labor did not begin spontaneously, that the health care provider gave them some medication to start labor, and that their cesarean section did not occur before the onset of labor. In both countries, results for this more refined definition show slight improvements to validity, and small but opposing changes to the IF. In Ghana, the IF improved from 1.18 to 1.15 and, in the Dominican Republic, the IF increased from 1.07 to 1.11. The validity of the individual question on labor induction showed very low sensitivity and high specificity in both countries. Sensitivity of reporting on spontaneous onset of labor was 84% and 89% in Ghana and the Dominican Republic, respectively. Specificity was 70% in Ghana and 51% in the Dominican Republic.

Overall validity assessed by AUC estimates show that in Ghana the indicator on emergency cesarean section by decision time had the highest validity (0.80), followed by emergency cesarean section by the operation time (0.79). The indicators tested in the Dominican Republic had moderate validity, ranging between 0.65 and 0.70, with the exception of induction of labor, which was very low (0.50).

### Exploring Accurate Reporting of Emergency Cesarean Section Status

Unadjusted odds ratios (ORs) showing the association between accurate reporting of emergency cesarean and women's age, education, and gravidity are presented in Table 3. In Ghana, women who were referred were half as likely to report accurately on the emergency status of their cesarean section (defined by decision time) as compared to non-referrals (OR: 0.49, 95% CI 0.29–0.83, $p = 0.009$). Although there was a positive trend between emergency cesarean section (defined by decision time) and age and education, neither association was statistically significant. None of the associations with emergency cesarean section defined by operation time were statistically significant. In contrast, in the Dominican Republic there was a negative and statistically significant relationship between accurate reporting of emergency cesarean section defined by decision time and gravidity and age and between emergency cesarean section defined by operation time and age.

## Discussion

Given the demand for more in-depth information on cesarean section, this study validated women's self-report of emergency cesarean section using two definitions in two countries. Diverse populations were sought to increase generalizability and to identify survey questions, which could be recommended for use in surveys in large-scale survey programs. Although both of the study sites were referral facilities in capital cities, Ghana represents settings similar to others in much of sub-Saharan Africa and elsewhere with low skilled attendance at birth, very low population-based cesarean section rates, and high maternal mortality. In contrast, the Dominican Republic is a country with nearly universal skilled attendance at birth, and therefore high population and facility-based cesarean section rates, yet one of the highest maternal mortality ratios in Latin America.

Results from this study support the premise that self-reporting on cesarean section is valid. Although 100% of women reported that they had undergone a cesarean section, and self-report on previous cesarean section showed excellent results in both populations, validation for both of these questions would require that the question also be asked of women delivering vaginally. Nonetheless, these results, coupled with the high sensitivity and specificity for cesarean section indicator observed in the study from China in this Collection [19], increase our confidence in the widely available survey data on self-reported cesarean section.

Results from Ghana for validity and the IF for emergency cesarean section defined by decision time are promising. The poor sensitivity results for this indicator in the Dominican Republic compelled us to consider explanations with our local collaborators. On further exploration, it was discovered that this discrepancy was probably due to poor documentation of decisions during antenatal care and the practice in the delivery ward of not checking the antenatal clinical history even though most of the women who delivered at the facility also attended the antenatal clinic there. This suggests that it is likely that women's reports are more accurate than medical records for this specific question. Validity of responses for emergency cesarean section defined by operation time in Ghana was less favorable than by decision time. In the Dominican Republic, the IF for the definition based on operation time was better than that for decision time, though with a specificity of less than 60%. The three-question approach did not improve results in either country; therefore our results do not justify the more demanding data requirements for the three-question definition relative to the one question approach. The validation results for the individual question on induced labor, of interest to maternal health planners independent of their role in identifying emergency cesarean section, cannot be recommended based on these results. However, it could possibly be improved via experimentation with different formulations of the questions.

It is important to note that the IF in our analyses was used as a measure of indicator quality and not as an adjustment factor for population-based survey results. Furthermore, there were no strong or consistent associations between women's characteristics and accurate reporting on emergency cesarean section that could be used to adjust survey-based results.

The study has a number of limitations. First, the quality of the validation reference standard was not consistently high due to different registry systems at the hospitals, as can be observed in the Dominican Republic results. Second, this validation study does not fully replicate the conditions in the DHS and MICS surveys, because our recall period was a few days, compared to up to five years in some surveys. However, given that emergency cesarean section is a surgical intervention, we hypothesize that women are likely to remember the event and crucial circumstances surrounding it [20]. Third, even though we conducted the study in two contrasting countries, both study hospitals were tertiary care facilities in urban areas serving populations with greater access to care than in rural areas. Also, it should be noted that among the women who had cesarean sections in the Dominican Republic study, only 52% were included in the final analysis due to women who left the hospital before they could be interviewed. Given that the median duration of hospital stay across the entire Dominican Republic sample was one day, it is unlikely that this loss to follow-up biased the sample toward women with less complicated pregnancies.

Population-based cesarean section rates are essential but insufficient information for health care planners, particularly in

**Table 3.** Unadjusted odds of accurately reporting emergency cesarean section using two definitions in the Ghana and Dominican Republic samples.

| | Decision Time for Cesarean Section | | | Operation Time for Cesarean Section (Single Question) | | |
|---|---|---|---|---|---|---|
| | Odds Ratio | 95% Confidence Interval | *p*-Value>\|z\| | Odds Ratio | 95% Confidence Interval | *p*-Value>\|z\| |
| **Ghana (N = 659)** | | | | | | |
| **Age** | | | | | | |
| ≤24 | 1.00 | | | 1.00 | | |
| 25–34 | 1.29 | 0.77–2.16 | 0.33 | 0.85 | 0.51–1.43 | 0.55 |
| ≥35 | 1.81 | 0.97–3.39 | 0.06 | 1.01 | 0.56–1.83 | 0.97 |
| **Education** | | | | | | |
| None | 1.00 | | | 1.00 | | |
| Primary | 1.69 | 0.73–3.91 | 0.22 | 1.00 | 0.43–2.31 | 0.98 |
| Secondary | 1.65 | 0.88–3.09 | 0.12 | 0.94 | 0.49–1.83 | 0.87 |
| University | 2.14 | 0.95–4.83 | 0.07 | 0.92 | 0.42–2.02 | 0.84 |
| **Gravidity** | | | | | | |
| 1st | 1.00 | | | 1.00 | | |
| 2nd | 1.01 | 0.63–1.64 | 0.96 | 1.02 | 0.65–1.60 | 0.93 |
| 3rd | 1.11 | 0.64–1.95 | 0.71 | 0.82 | 0.49–1.36 | 0.44 |
| 4th | 1.12 | 0.48–2.95 | 0.69 | 1.71 | 0.67–4.34 | 0.26 |
| **Referral** | | | | | | |
| No | 1.00 | | | 1.00 | | |
| Yes | 0.49 | 0.29–0.83 | 0.01 | 1.39 | 0.92–2.10 | 0.115 |
| **Dominican Republic (N = 1,531)** | | | | | | |
| **Age** | | | | | | |
| ≤24 | 1.00 | | | 1.00 | | |
| 25–34 | 0.77 | 0.62–0.96 | 0.02 | 0.75 | 0.58–0.95 | 0.019 |
| ≥35 | 0.65 | 0.43–1.00 | 0.05 | 0.61 | 0.39–0.96 | 0.034 |
| **Education** | | | | | | |
| None | 1.00 | | | 1.00 | | |
| Primary | 1.03 | 0.45–2.35 | 0.95 | 0.64 | 0.23–1.76 | 0.38 |
| Secondary | 1.36 | 0.60–3.07 | 0.46 | 0.79 | 0.29–2.14 | 0.64 |
| University | 1.78 | 0.76–4.15 | 0.18 | 0.71 | 0.25–1.98 | 0.51 |
| **Gravidity** | | | | | | |
| 1st | 1.00 | | | 1.00 | | |
| 2nd | 0.42 | 0.32–0.55 | 0.00 | 0.90 | 0.66–1.22 | 0.49 |
| 3rd | 0.40 | 0.31–0.53 | 0.00 | 0.78 | 0.58–1.03 | 0.08 |
| 4th | 0.34 | 0.24–0.51 | 0.00 | 0.70 | 0.46–1.06 | 0.09 |
| **Referral** | | | | | | |
| No | 1.00 | | | 1.00 | | |
| Yes | 0.91 | 0.72–1.15 | 0.41 | 0.94 | 0.73–1.23 | 0.67 |

doi:10.1371/journal.pone.0060761.t003

countries without adequate routine health information systems to provide in-depth health facility-based cesarean-related data. The inadequacy of the cesarean section rate alone (without the proportion of emergency operations) is particularly acute in countries where the rate falls between 5% and 10%. In these settings, as the cesarean section rate increases, the poorest women may still not have access to life-saving delivery by cesarean section. However, the emergency cesarean section trends should be interpreted cautiously in settings such as Brazil, where cesarean sections are almost universal among certain sub-populations. [21] Although low-income countries should strive to establish robust routine health information systems which permit national-level monitoring, given current challenges, health care planners will need to rely on national surveys for the foreseeable future. Given our reliance on survey-based indicators, the most important aspect of data quality will vary by the purpose and use of the indicator. Although highly valid data are preferred for all purposes, highly sensitive and specific data are required for individual level analyses, whereas an IF near equality is sufficient for monitoring trends.

The results presented here are promising but insufficient to promote inclusion of the questions supporting the two definitions of emergency cesarean section into international survey program questionnaires. Further research on this indicator is warranted. Such studies should (1) confirm the accuracy of facility-based data on time of decision to operate in advance of data collection, (2) extend the recall period to be comparable to that of population-based surveys, and (3) based on results from the Mozambique validation study in this collection [22], allow for 50% loss to follow-up in sample size estimation to account for the extended recall period. Furthermore, qualitative research could lead to refined formulation of certain questions such as induction of labor, and potentially improve the validity of these additional indicators.

## Supporting Information

**Text S1** Exit interview in Spanish.
(DOC)

## References


1. World Health Organization (2012) World Health Statistics. Geneva, Switzerland: World Health Organization (WHO).
2. Betran AP, Merialdi M, Lauer JA, Bing-Shun W, Thomas J, et al. (2007) Rates of caesarean section: analysis of global, regional and national estimates. Paediatr Perinat Epidemiol 21: 98–113.
3. ICDDRB NM-E (2011) Bangladesh, Maternal Mortality and Health Care Survey 2010; Preliminary Results. Chapel Hill, North Carolina USA: University of North Carolina.
4. Ronsmans C, Holtz S, Stanton C (2006) Socioeconomic differentials in caesarean rates in developing countries: a retrospective analysis. Lancet 368: 1516–1523.
5. Debas HT, Gosselin R, McCord C, Thind A (2006) Surgery. In: D.B. Jamison J, Measham A, Alleyne G, Claeson M, Evans D et al., editors. Disease Control Priorities in Developing Countries, 2nd ed: Oxford University Press and The World Bank. 1245–1259.
6. Holtz S (2007) Clustering of Cesarean Deliveries in Developing Countries. Baltimore, MD: Johns Hopkins University.
7. Stanton CK, Dubourg D, De Brouwere V, Pujades M, Ronsmans C (2005) Reliability of data on caesarean sections in developing countries. Bull World Health Organ 83: 449–455.
8. CHERG MA #13 (2010) Expanding the List of Indicators regarding Cesarean Birth: Report from the meeting at The Johns Hopkins Bloomberg School of Public Health. Baltimore, MD.
9. Schauberger CW, Chauhan SP (2009) Emergency cesarean section and the 30-minute rule: definitions. Am J Perinatol 26: 221–226.
10. Torloni MR, Betran AP, Souza JP, Widmer M, Allen T, et al. (2011) Classifications for cesarean section: a systematic review. PLoS One 6: e14566.
11. Macro International, Ghana Statistical Service, Services. GH (2009) Ghana Maternal Health Survey 2007. Accra, Ghana and Calverton, Maryland USA.
12. Macro International (2011) Measure-DHS StatCompiler.
13. Witter S, Arhinful DK, Kusi A, Zakariah-Akoto S (2007) The experience of Ghana in implementing a user fee exemption policy to provide free delivery care. Reprod Health Matters 15: 61–71.
14. Castro A (2010) Prevention of Mother-to-Child Transmission of HIV and Syphilis in Latin America and the Caribbean. In: Castro A, editor. Challenges Posed by the HIV Epidemic in Latin America and the Caribbean 2009. Washington, DC: Pan American Health Organization, UNICEF, and UNAIDS.
15. Dirección General de Información y Estadística de Salud (2008) Salud Reproductiva por Regiones, Provincias, y Establecimientos de Salud. Santo Domingo: SESPAS.
16. Ronsmans C, Achadi E, Cohen S, Zazri A (1997) Women's recall of obstetric complications in south Kalimantan, Indonesia. Stud Fam Plann 28: 203–214.
17. Vecchio TJ (1966) Predictive value of a single diagnostic test in unselected populations. N Engl J Med 274: 1171–1173.
18. Campbell H, Biloglav Z, Rudan I (2008) Reducing bias from test misclassification in burden of disease studies: use of test to actual positive ratio–new test parameter. Croat Med J 49: 402–414.
19. Guo Y, Liu L, Li M, Yang L, Ju L, et al. (2013) Measuring coverage in MNCH: A validation study linking population survey derived coverage to maternal, newborn, and child health care records in rural China. PLoS ONE 8: e60762. doi:10.1371/journal.pone.0060762.
20. Coulter A, McPherson K, Elliott S, Whiting B (1985) Accuracy of recall of surgical histories: a comparison of postal survey data and general practice records. Community Med 7: 186–189.
21. Barros AJ, Santos IS, Matijasevich A, Domingues MR, Silveira M, et al. (2011) Patterns of deliveries in a Brazilian birth cohort: almost universal cesarean sections for the better-off. Rev Saude Publica 45: 635–643.
22. Stanton CK, Rawlins B, Drake M, Anjos M, Cantor D, et al. (2013) Measuring coverage in MNCH: Testing the validity of women's self-report of key maternal and newborn health interventions during the peripartum period in Mozambique. PLoS ONE 8: e60694. doi:10.1371/journal.pone.0060694.



## Acknowledgments

The authors wish to acknowledge Rod Knight of Principia International for his contribution to the calculation of the sample size, Isaac Newton Hotorvi, Alfred Aikins, and Godwin Binlinla of Korle-Bu Teaching Hospital for their assistance in the data collection and management in Ghana, and Ilonka Agramonte, Arismendy Benítez, Jean-René Louis, Merary Mota, and Angel Peguero for their assistance in the data collection in the Dominican Republic.


## Author Contributions

Conceived and designed the experiments: OT CS RA KAB AC MH AB AL SL. Performed the experiments: OT AC MH KAB RA. Analyzed the data: OT CS. Wrote the paper: OT CS. Contributed to the interpretation of data from the DR for the manuscript: AC MH. Provided input to the manuscript: AC MH KAB RA AB AL.

PLOS ONE

# Measuring Coverage in MNCH: A Validation Study Linking Population Survey Derived Coverage to Maternal, Newborn, and Child Health Care Records in Rural China

**Li Liu[1], Mengying Li[1], Li Yang[2], Lirong Ju[3], Biqin Tan[4], Neff Walker[1], Jennifer Bryce[1], Harry Campbell[5], Robert E. Black[1], Yan Guo[6]***

**1** Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, **2** Guangxi Medical University, School of Public Health, Nanning, China, **3** Capital Medical University, School of Public Health, Beijing, China, **4** Gongcheng County Maternal and Child Health Hospital, Gongcheng County, China, **5** University of Edinburgh Medical School, Edinburgh, United Kingdom, **6** Peking University Health Science Center, School of Public Health, Beijing, China

## Abstract

***Background:*** Accurate data on coverage of key maternal, newborn, and child health (MNCH) interventions are crucial for monitoring progress toward the Millennium Development Goals 4 and 5. Coverage estimates are primarily obtained from routine population surveys through self-reporting, the validity of which is not well understood. We aimed to examine the validity of the coverage of selected MNCH interventions in Gongcheng County, China.

***Method and Findings:*** We conducted a validation study by comparing women's self-reported coverage of MNCH interventions relating to antenatal and postnatal care, mode of delivery, and child vaccinations in a community survey with their paper- and electronic-based health care records, treating the health care records as the reference standard. Of 936 women recruited, 914 (97.6%) completed the survey. Results show that self-reported coverage of these interventions had moderate to high sensitivity (0.57 [95% confidence interval (CI): 0.50–0.63] to 0.99 [95% CI: 0.98–1.00]) and low to high specificity (0 to 0.83 [95% CI: 0.80–0.86]). Despite varying overall validity, with the area under the receiver operating characteristic curve (AUC) ranging between 0.49 [95% CI: 0.39–0.57] and 0.90 [95% CI: 0.88–0.92], bias in the coverage estimates at the population level was small to moderate, with the test to actual positive (TAP) ratio ranging between 0.8 and 1.5 for 24 of the 28 indicators examined. Our ability to accurately estimate validity was affected by several caveats associated with the reference standard. Caution should be exercised when generalizing the results to other settings.

***Conclusions:*** The overall validity of self-reported coverage was moderate across selected MNCH indicators. However, at the population level, self-reported coverage appears to have small to moderate degree of bias. Accuracy of the coverage was particularly high for indicators with high recorded coverage or low recorded coverage but high specificity. The study provides insights into the accuracy of self-reports based on a population survey in low- and middle-income countries. Similar studies applying an improved reference standard are warranted in the future.

**Citation:** Liu L, Li M, Yang L, Ju L, Tan B, et al. (2013) Measuring Coverage in MNCH: A Validation Study Linking Population Survey Derived Coverage to Maternal, Newborn, and Child Health Care Records in Rural China. PLoS ONE 8(5): e60762. doi:10.1371/journal.pone.0060762

**Editor:** Dr. Lucy Chappell, Kings Colleage London, United Kingdom, in consultation with Carla AbouZahr, independent consultant, health statistics and policy

**Received** August 14, 2012; **Accepted** February 26, 2013; **Published** May 7, 2013



**Funding:** This work was conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF (No. 50140 and No. 43386), with financial support from The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** AUC, area under the receiver operating characteristic curve; CHERG, Child Health Epidemiology Reference Group; CI, confidence interval; DHS, Demographic and Health Survey/s; GMNCHIS, Gongcheng MNCH Information System; HBsAb, hepatitis B antibody; LMIC, low- and middle-income country; MICS, Multiple Indicator Cluster Surveys; MNCH, maternal, newborn, and child health; ROC, receiver operating characteristic; TAP, test to actual positive.

* E-mail: guoyan@bjmu.edu.cn

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection*

## Introduction

Accurate data on coverage of key maternal, newborn, and child health (MNCH) interventions are crucial for monitoring progress toward the Millennium Development Goals 4 and 5 and ending preventable child deaths in a generation [1,2]. Recognizing its significance, the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF has made it a priority to improve coverage measurement for proven MNCH interventions. This paper is part of the *PLOS Medicine* "Measuring Coverage in MNCH" collection organized by CHERG for this purpose. Coverage estimates are generally obtained from routine population-based household surveys, such as the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys

(MICS), primarily through self-reporting [3]. However, little is known about the validity of self-reported coverage derived from these population-based surveys.

Previous validation studies comparing health care records with respondents' self-reports were mostly conducted in facility-based settings in high-income countries and found varying and often moderate validity across MNCH indicators studied. For example, an Australian study comparing medical records with women's reports of delivery interventions found that women's self-reports and medical records were both subject to errors [4]. In another recent population-based survey conducted in the UK, self-reported delivery mode was found to be highly reliable [5]. Two studies done in American hospitals showed generally unsatisfactory validity of self-reported medical interventions in the pregnancy, delivery, and postnatal periods [6,7], while one study among mothers of children with cancer in Canada and the US reported moderate to high validity for similar indicators [8]. According to one other US study, parents' reports of children's immunizations were of unsatisfactory validity, mainly due to poor initial encoding of the events [9].

Results from high-income countries may have limited generalizability when applied to the low- and middle-income countries (LMICs) due to different levels of coverage, intensities of service provision counseling, and degrees of recall bias that may be associated with the education level of respondents [10,11]. However, similar studies are sparse in the LMIC setting; most of them have focused on obstetrical complications rather than routine interventions [12–14]. To our knowledge, the current study and other validation studies in this collection are the only ones that aim to evaluate the accuracy of coverage of MNCH interventions in LMICs [15–18].

Most of the validation studies reviewed are facility-based, and therefore subject to selection biases, because the study sample is often not representative of the general population. The fact that such a facility-based study design is widely adopted is perhaps because validation studies based on population surveys are more methodologically challenging. This is particularly true in LMICs, because health-care recordkeeping systems are rarely complete or of adequate quality to be used as the reference standard. In this study, we sought to examine the validity of self-reported coverage of selected MNCH interventions in a relatively less developed rural area in China, a setting selected to increase the extent to which the study results would be generalizable to other LMICs. In addition, we attempted to minimize selection bias associated with facility-based validation studies by collecting the study sample through a population-based survey.

## Methods

### Study Site

We conducted the validation study in Gongcheng County, which is located in Guangxi Province in southwestern China. The county contains nine townships and 125 villages [19], with a total population of 285,058 based on a 2010 census, of which 59% were of Yao and 39% were of Han ethnicity [20]. Among the population aged 15 and above, 1.2% were illiterate. The majority of the population were fruit farmers and engage in citrus and persimmon production. In 2006, the county GDP per capita was reported to be around $1,500 [19].

In 2006–2010, the under-five mortality rate in Gongcheng County decreased from 15.2 to 8.1 per 1,000 live births, and the infant mortality rate declined from 13.4 to 6.8 per 1,000 live births [21]. During the same period, the maternal mortality ratio was on average 31.8 per 100,000 live births [21]. Most MNCH services are provided in county- and township-level hospitals and village clinics [22]. Coverage of antenatal care and institutional delivery is close to universal in the past five years [21]. Information on MNCH services is routinely recorded by service providers in a number of booklets, including, for example, the antenatal care booklet, the maternal and child health booklet, the child care booklet, and the vaccination booklet. The antenatal and child care booklets have been in use since 2003. These booklets are usually kept by women. In January 2007, an electronic MNCH information system was launched as part of the Guangxi Province MNCH information system. The system digitized key information collected from the booklets.

### Data Collection

Women aged 18 to 49 years who lived in Gongcheng County during the fieldwork and had delivered at least one live birth in the county in the five years preceding the survey (i.e., between 01 November 2006 and 01 November 2011) were eligible to participate in the study. Participants were selected via multi-stage stratified sampling with a target of interviewing mothers of 1,000 live births. The target sample size was determined based on the following consideration. The study was originally designed to also evaluate whether the validity of women's self-reports was worse based on a five-year compared to a two-year recall period. The study sample size needed to be sufficient to distinguish ten percentage point differences in validity (e.g., sensitivity) when comparing the two recall periods. Ten percent is considered to be programmatically important. Since no prior information was available on the sensitivity of the measurement of any indicators studied, 50% sensitive was assumed for the five-year recall period to yield the most conservative sample size. Based on a 10% difference, 60% sensitivity was assumed for the two-year recall. Assuming constant fertility, the number of live births born in the past two years is two-fifths of those born in the past five years. To ensure that the sample size was conservative, continuation correction was applied to improve the approximation of binomial distribution to the normal distribution [23]. With a significance level of 0.05 and 80% of power, sample size calculation for two-sample comparison of proportions using Stata 10 produced a total sample size of 714 live births in the past five years and 286 live births in the past two years [24]. We assumed that coverage of 20% of the study sample cannot be validated due to the lack of the reference standard. Taking into account a 10% non-response rate, a total of 992 or approximately 1,000 live births were needed. Information on 900 live births was anticipated to be actually collected.

Study women were selected via a multi-stage stratified sampling design. In the first stage, the nine townships were divided into three strata based on the population size as a proxy of the level of economic development, and one township was selected in each stratum. In the second stage, villages were divided into four groups according to their geographic location (east, west, north, and south), and one village was sampled from each geographic group. In the third stage, participants were recruited for interview by the village doctors based on availability by going through the vaccination roster, which is considered to have enlisted all children under five years of age in the villages. Recruitment stopped once the desired sample size in each township was reached. During the recruitment process, women were also asked to bring their MNCH booklets to the community interview for abstraction.

The study period overlapped with the persimmon harvesting season, which made it difficult to recruit participants based on our original plan. However, the fieldwork happened to fall on child immunization days in two of the three sampled townships, during

which young children were brought to the township hospitals for immunization and well-baby checkup. A fraction of the study sample was recruited on the immunization days at the end. As a result, children younger than two years were over-represented in the study sample.

Community-based face-to-face interviews were conducted in village centers with reasonable privacy after obtaining written informed consent. The survey instrument was adapted from the DHS and MICS questionnaires to suit the local context [25]. It was used to solicit information on coverage of selected MNCH interventions, including those routinely collected in the DHS or MICS or of local relevance. Some additional modifications were made in the wording of a number of questions, including those on child vaccination, in an effort to reduce the length of the questionnaire. The exact wording of the questions used in the survey in comparison with those used in the DHS and MICS questionnaires is provided in Table S1. The questionnaires were designed in English, translated into Chinese, and verified through back-translation. One questionnaire was administered to each eligible woman to collect information on household characteristics and her socio-demographic background. A second questionnaire was administered for each eligible live birth to collect information on services received during the antenatal, delivery, newborn, postnatal, and child care periods.

We abstracted relevant records from available booklets after completion of the interview using a structured template. We also extracted relevant records from the electronic system for all women residing in Gongcheng who delivered locally between the initiation of the electronic system (01 January 2007) and 01 October 2011. Because study women may not have all booklets available for abstraction, the electronic system was not in operation for the first few months of the study reference period, and a small set of indicators were only available in the booklets, we combined records from booklet abstraction and the electronic systems. The study reference standard was created by giving preference to the electronic system and is referred to as Gongcheng MNCH Information System (GMNCHIS). Only indicators for which information was available from GMNCHIS were included in the validation. A complete list of validated indicators can be found in Table S1.

## Data Analyses

We cleaned and matched data collected through the community survey and those abstracted from the GMNCHIS. A number of databases were exported from the electronic system, including pregnant women's general information, early antenatal care, other antenatal care, high-risk pregnancy, antenatal screening tests, delivery, postnatal, and child care databases. When processing the exported data, record of results of a test or examination was treated as the evidence of the receipt of the test or examination, and the lack of such a record was treated as evidence of not receiving the service. Records in different databases were matched using the maternal and child health identification unique for each pregnancy. Record of receipt of a service in any of the databases was considered evidence of the receipt of the service.

To identify potentially duplicated records for the same pregnancy in the electronic system, we first identified records for the same woman, defined as records with either the same national identification or the same name and village, and less than four years of differences in reported age. We then identified the records for the same pregnancy, defined as records of the same woman who had first antenatal visits fewer than 30 days apart, or last menstrual periods fewer than 30 days apart, or delivery records fewer than 150 days apart so that a pregnancy loss before another pregnancy was not identified as the same pregnancy. Lastly, we collapsed all the duplicated records for the same pregnancy. A total of 15,189 unique women with 16,049 unique pregnancies for the whole county were identified in the electronic system.

To match women's recall in the community-based survey with those collected through the GMNCHIS, we first matched survey data with those exported from the electronic system using the following combinations of information: (1) maternal and child health identification and women's names, or (2) first 14 digits of the national identification which includes an area code, date of birth, women's names, and children's date of birth, or (3) women's names, village names, and children's date of birth. Then we used data abstracted from the booklets as the reference standard for those women and indicators which were not matched using data exported from the electronic system. Our unit of analysis was live birth.

We grouped the results into four categories for presentation, including antenatal care, delivery care, postnatal care, and child vaccination services. Antenatal care includes routine antenatal care, blood screening for sexually transmitted diseases, and blood screening for congenital abnormalities. During the routine antenatal care, the first ultrasound scan was done on the first antenatal visit, usually around 10–14 weeks. It provided information on gestational age and examined fetal conditions, including measuring nuchal translucency to screen for Down's syndrome. Normally, a few antenatal tests are done during each antenatal visit, including weight/height/blood pressure measurements, urine test, and fetal heart monitoring. For these repeated tests, we only measured whether they were received at least one time. Pregnant women are screened for thalassemia by a combination of mean cell volume count, erythrocyte osmotic fragility test, and hemoglobin electrophoresis [26,27].

We calculated sensitivity and specificity of the self-reported coverage. We graphed the overall validity in a receiver operating characteristic (ROC) plot with true positive rate, or sensitivity, plotted against false positive rate, or one minus specificity. We also quantified the overall validity by the area under the ROC curve (AUC) [28]. The uncertainty associated with validity, as represented by the 95% confidence interval (CI), was estimated assuming a binomial distribution.

Population-level accuracy of the coverage estimates was also examined, which is measured by the test to actual positive (TAP) ratio or the reported over-recorded coverage [29]. It can be demonstrated mathematically that the TAP ratio is determined by the validity of self-reported coverage in combination with the actual (or recorded) coverage [29]. If the recorded coverage is high, the TAP ratio approximately equals sensitivity and is independent of specificity [29]. A combination of low recorded coverage and low to moderate specificity results in a high TAP ratio [29,30]. We also investigated and discussed the complex mathematical relationship observed between the TAP ratio, validity, and recorded coverage.

For the purpose of describing the study results, we categorized coverage, sensitivity, and specificity as low, moderate, and high, applying two cut-off points at 0.33 and 0.66. We also considered the overall validity high if the AUC was at or above 0.67 and moderate and low otherwise. We qualitatively defined the population-level bias based on the TAP ratio as small ($0.8<$TAP ratio$<1.2$), moderate ($0.5<$TAP ratio$<1.5$), and large (TAP ratio$<0.5$ or TAP ratio$>1.5$). We also conducted two sensitivity analyses to treat information solely abstracted from the booklets as the reference standard and to limit the study sample to women who gave birth in the past two years.

### Ethical Review

The study protocol was approved by the Institutional Review Boards of the Johns Hopkins Bloomberg School of Public Health and Peking University.

## Results

### Characteristics of the Study Sample

Nine hundred and thirty-six women were recruited between 10 and 22 November 2011. Among them, 914 agreed to participate in the survey and delivered 994 eligible live births. Interviews on 961 eligible live births were completed, among whom mothers of 431, 115, 343, and 793 live births brought the antenatal, maternal and child health, child care, and vaccination booklets to the community survey, respectively. Seven hundred and twelve live births were matched using electronic information. Another 196 live births were matched using information from at least one booklet, yielding a total of 908 matched live births. The remaining 53 live births did not have any matched indicators and could not be validated.

The socio-demographic characteristics of the surveyed live births by matching status are presented in Table 1. Overall, mothers of almost 60% live births were aged between 25 and 34 years, and of more than half had secondary or higher education. Similar to Gongcheng's general population, mothers of 59% of live births were of Yao ethnicity. The majority (84%) of the live births sampled was the only one born in the past five years and 42% were under one year old. More than half of the live births lived in households with an annual income per capita ranging between 1,000 and 5,000 Yuan, or 158 and 791 dollars.

Mother's age, education, and household annual income per capita were not significantly different between the matched and unmatched live births. However, the matched live births were more likely to have been born to mothers of Yao ethnicity, to be the only live birth in the last five years, and to be younger than 24 months.

### Validity of the MNCH Indicators

Reported coverage derived from the community survey, recorded coverage derived from the GMNCHIS and TAP ratio are reported in Table 2. Sensitivity, specificity, AUC, and their corresponding 95% CI are presented in Table 3. The reported coverage of the routine antenatal care indicators was high (>81%). Their recorded coverage was also high, with the exception of the first antenatal visit before 12 weeks of gestational age. Self-reported coverage of routine antenatal interventions had sensitivity close to 0.90 and specificity below 0.25. Recorded coverage of the antenatal HIV and hepatitis B antibody (HBsAb) tests was similar, although their reported coverage differed greatly. Self-reported coverage of the HIV test had moderate sensitivity (0.59 [95% CI: 0.54–0.65]) and specificity (0.53 [95% CI: 0.46–0.60]), whereas that of the HBsAb test had high sensitivity (0.89 [95% CI: 0.86–0.92]) and low specificity (0.18 [95% CI: 0.13–0.22]). Among the screening tests for Down's syndrome, neural tube defects, and thalassemia, despite similar levels of reported coverage of 54%–68%, the recorded coverage varied, ranging between 18% and 52%. Self-reported coverage of each of these screening tests had sensitivities of 0.75–0.87 and specificities of 0.46–0.52.

Coverage of cesarean section was reported to be 36%, compared to the recorded 24%. It had high sensitivity (0.96 [95% CI: 0.93–0.99]) and specificity (0.83 [95% CI: 0.80–0.86]). Among the postnatal care indicators, coverage of occurrence of at least one postnatal visit was reported to be higher than the recorded value, with a moderate sensitivity (0.57 [95% CI: 0.50–0.63]) and a high specificity (0.72 [95% CI: 0.68–0.76]). The rest of the postnatal indicators had high reported and recorded coverage, with moderate to high sensitivity (0.66–0.93) and low to moderate specificity (0.21–0.35). Reported and recorded coverage of vaccination was consistently high, with the exception of measles vaccine. Self-reported coverage of vaccination also had high sensitivity (>0.86) and a wide range of specificity (0.02–0.70).

The AUC estimates reported in Table 3 and the ROC plot shown in Figure 1 demonstrate the overall validity of self-reported coverage by indicator. Self-reported coverage of cesarean section had the highest overall validity when compared to the reference standard, with the AUC being 0.90 [95% CI: 0.88–0.92]. Diphtheria-pertussis-tetanus (DPT) vaccine ranked the second, with the AUC being 0.80 [95% CI: 0.75–0.84]. Self-reported coverage of thalassemia screening and measles vaccine also had high overall validity (AUC>0.69). The remaining indicators had either moderate or low overall validity, with the AUC of a number of them not significantly different from 0.5, indicating validity equivalent to a random guess.

Despite varying overall validity, the TAP ratios ranged between 0.8 and 1.5 for self-reported coverage of 24 of the 28 indicators examined, suggesting mostly small to moderate degree of bias at the population level (Figure 2). However, it was particularly large for four indicators, including measles vaccine (TAP ratio = 2.0), first antenatal visit before 12 weeks of gestational age (TAP ratio = 2.7), screening for neural tube defect (TAP ratio = 2.7), and screening for Down's syndrome (TAP ratio = 3.2). Both sensitivity analyses for using the booklets as the reference standard and limiting the sample to women who gave birth in the last two years gave quantitatively similar results (not shown).

## Discussion

To our knowledge, this is the first study to validate self-reported coverage of a range of MNCH indicators by systematically comparing women's self-reports solicited from a population-based survey of MNCH care records in a LMIC. We found that across the indicators examined, self-reported coverage had moderate to high sensitivity and low to moderate specificity. The overall validity is high for the self-reported coverage of a few indicators including cesarean section, diphtheria-pertussis-tetanus vaccine, measles vaccine and screening for thalassemia, yet moderate to low for the remaining indicators. The finding of moderate levels of overall validity is not unexpected, as similar results have been reported in previous studies in high-income countries [6,9].

The variation in validity across indicators seems to suggest that the more distinctive the experiences women had while receiving certain interventions, the better was the validity of the self-reported coverage. The positive association between event distinctiveness and recall accuracy is supported by the psychology literature [31]. The variation could also be the result of the social desirability bias associated with self-reports. That is, when women perceived that it was socially desirable to receive a certain service, they were more likely to report receipt of the service regardless of whether they had actually received it or not. An example illustrating the potential social desirability bias can be drawn from the coverage and validity of the HIV and HBsAb tests. The two tests had similar levels of recorded coverage, yet widely different levels of reported coverage. We hypothesize that women may be less willing to report receipt of an HIV test than an HBsAb test, as the former is less socially desirable.

Despite varying validity, self-reported coverage of the majority of the examined indicators had only a small to moderate degree of population-level bias. At least two reasons can perhaps explain

**Table 1.** Socio-demographic characteristics of surveyed live births by matching status.

| Category | Total | | Matched | | Unmatched | | p-Value[a] |
|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | |
| **Mother's age, years** | | | | | | | 0.164 |
| 24 and below | 162 | 16.9% | 159 | 17.5% | 3 | 5.7% | |
| 25–29 | 317 | 33.1% | 299 | 33.0% | 18 | 34.0% | |
| 30–34 | 251 | 26.2% | 232 | 25.6% | 19 | 35.8% | |
| 35–39 | 161 | 16.8% | 151 | 16.7% | 10 | 18.9% | |
| 40 and above | 68 | 7.1% | 65 | 7.2% | 3 | 5.7% | |
| Total[b] | 959 | 100.0% | 906 | 100.0% | 53 | 100.0% | |
| **Mother's education** | | | | | | | 0.139 |
| Primary | 336 | 35.0% | 311 | 34.3% | 25 | 47.2% | |
| Secondary | 510 | 53.1% | 486 | 53.5% | 24 | 45.3% | |
| Tertiary | 115 | 12.0% | 111 | 12.2% | 4 | 7.5% | |
| Total | 961 | 100.0% | 908 | 100.0% | 53 | 100.0% | |
| **Mother's ethnic group** | | | | | | | 0.045 |
| Yao | 569 | 59.2% | 546 | 60.1% | 23 | 43.4% | |
| Han | 340 | 35.4% | 313 | 34.5% | 27 | 50.9% | |
| Other | 52 | 5.4% | 49 | 5.4% | 3 | 5.7% | |
| Total | 961 | 100.0% | 908 | 100.0% | 53 | 100.0% | |
| **Number of births in the last 5 years** | | | | | | | <0.001 |
| 1 | 806 | 84.2% | 780 | 86.3% | 26 | 49.1% | |
| 2 | 147 | 15.4% | 120 | 13.3% | 27 | 50.9% | |
| 3 | 4 | 0.4% | 4 | 0.4% | 0 | 0.0% | |
| Total[b] | 957 | 100.0% | 904 | 100.0% | 53 | 100.0% | |
| **Children's age, months** | | | | | | | <0.001 |
| Below 12 | 404 | 42.0% | 401 | 44.2% | 3 | 5.7% | |
| 12–23 | 192 | 20.0% | 188 | 20.7% | 4 | 7.5% | |
| 24–35 | 138 | 14.4% | 128 | 14.1% | 10 | 18.9% | |
| 36–47 | 112 | 11.7% | 105 | 11.6% | 7 | 13.2% | |
| 48–59 | 115 | 12.0% | 86 | 9.5% | 29 | 54.7% | |
| Total | 961 | 100.0% | 908 | 100.0% | 53 | 100.0% | |
| **Annual household income per capita, Yuan** | | | | | | | 0.373 |
| Below 1,000 | 104 | 14.4% | 99 | 14.6% | 5 | 10.6% | |
| 1,000–1,999 | 157 | 21.7% | 143 | 21.1% | 14 | 29.8% | |
| 2,000–4,999 | 253 | 34.9% | 235 | 34.7% | 18 | 38.3% | |
| 5,000–9,999 | 141 | 19.5% | 136 | 20.1% | 5 | 10.6% | |
| 10,000 and above | 69 | 9.5% | 64 | 9.5% | 5 | 10.6% | |
| Total[b] | 724 | 100.0% | 677 | 100.0% | 47 | 100.0% | |

[a]Based on the Chi-square test comparing matched and unmatched samples.
[b]Live births born to mothers who answered "don't know" to or refused to answer the corresponding questions are not presented.
doi:10.1371/journal.pone.0060762.t001

this. The first reason is based on the mathematical relationship between bias, validity, and the recorded coverage. It can be demonstrated that when recorded coverage is high, the TAP ratio approximately equals sensitivity and is independent of specificity [29]. Because a large proportion of the indicators had high recorded coverage and high sensitivity, their TAP ratio did not deviate greatly from 1. Of note, although high coverage may have limited our power to accurately estimate specificity, the accuracy of coverage is not much affected, as specificity is almost irrelevant to population-level bias when coverage is high.

The second reason is better recall and recognition of certain interventions due to better community knowledge associated with high coverage. We observed that the higher the recorded coverage is, the higher the sensitivity is, and the correlation is marginally significant ($p = 0.06$). This should not be the case under normal circumstances, as sensitivity and specificity are in theory intrinsic to the estimation of self-reported coverage, and are usually independent of the actual coverage [32]. As a result, high coverage is also likely associated with high sensitivity.

On the other hand, the self-reported coverage of a few indicators had large bias, including screening for Down's

**Table 2.** Coverage of selected MNCH indicators.

| Indicators | Number of Observations | | | Coverage | | |
|---|---|---|---|---|---|---|
| | Missing[a] | Indeter-minate[b] | Validated | GMNCHIS | Community Survey | TAP Ratio |
| **Routine antenatal care** | | | | | | |
| First antenatal care <12 weeks of gestational age | 194 | 86 | 675 | 0.31 | 0.81 | 2.66 |
| At least four antenatal visits | 176 | 246 | 539 | 0.70 | 0.91 | 1.30 |
| Weight measurement | 165 | 14 | 782 | 1.00 | 0.98 | 0.98 |
| Height measurement | 182 | 10 | 769 | 0.96 | 0.89 | 0.92 |
| Blood pressure measurement | 165 | 6 | 790 | 0.99 | 0.98 | 0.99 |
| Hemoglobin test | 132 | 53 | 776 | 0.85 | 0.92 | 1.09 |
| Urine test | 166 | 8 | 787 | 0.93 | 0.99 | 1.06 |
| Fetal heart rate monitoring | 175 | 6 | 780 | 0.99 | 0.99 | 1.00 |
| Ultrasound exam | 189 | 4 | 768 | 0.86 | 0.99 | 1.15 |
| **Screening for STD** | | | | | | |
| HIV test | 256 | 152 | 553 | 0.60 | 0.54 | 0.91 |
| HBsAb test | 175 | 86 | 700 | 0.59 | 0.86 | 1.46 |
| **Screening for congenital abnormalities** | | | | | | |
| Down's syndrome screening | 244 | 206 | 511 | 0.18 | 0.59 | 3.21 |
| Neural tube defect screening | 200 | 228 | 533 | 0.20 | 0.54 | 2.73 |
| Thalassemia screening | 769 | 41 | 151 | 0.52 | 0.68 | 1.32 |
| **Delivery care** | | | | | | |
| Cesarean section | 184 | 0 | 777 | 0.24 | 0.36 | 1.47 |
| **Postnatal care** | | | | | | |
| At least 1 postnatal visit | 231 | 3 | 727 | 0.29 | 0.37 | 1.25 |
| Blood pressure | 810 | 3 | 148 | 0.83 | 0.87 | 1.05 |
| Temperature | 854 | 3 | 104 | 0.78 | 0.87 | 1.11 |
| Breast exam | 810 | 2 | 149 | 0.83 | 0.66 | 0.80 |
| Uterus exam | 810 | 5 | 146 | 0.83 | 0.77 | 0.93 |
| Lochia exam | 810 | 2 | 149 | 0.83 | 0.77 | 0.93 |
| Perineum exam | 811 | 3 | 147 | 0.84 | 0.81 | 0.97 |
| Family planning advice | 821 | 1 | 139 | 0.68 | 0.75 | 1.11 |
| **Child vaccination** | | | | | | |
| BCG vaccine | 240 | 151 | 570 | 0.91 | 0.94 | 1.04 |
| Polio vaccine | 258 | 14 | 689 | 1.00 | 0.86 | 0.86 |
| HBV vaccine | 256 | 31 | 674 | 0.92 | 0.98 | 1.07 |
| DPT vaccine | 230 | 273 | 458 | 0.75 | 0.74 | 0.99 |
| Measles vaccine | 253 | 206 | 502 | 0.35 | 0.70 | 2.00 |

[a]Missing indicates live births who were not matched.
[b]Indeterminate indicates live births whose mothers answered "Don't know" or refused to give an answer to the corresponding survey questions.
BCG, Bacillus Calmette-Guérin; DPT, diphtheria-pertussis-tetanus; HBV, hepatitis B virus.
doi:10.1371/journal.pone.0060762.t002

syndrome and neural tube defects, first antenatal visit before 12 weeks of gestational age, and measles vaccine. This is likely due to a similar mathematical relationship–a combination of low recorded coverage and low to moderate specificity results in a high TAP ratio [29,30]. In fact, for the current recorded coverage and sensitivity of screening for Down's syndrome, for example, specificity needs to be as high as 0.96, compared to the current 0.46, to yield a TAP ratio of 1 [29]. For a combination of low prevalence and low specificity, coverage derived from self-reports in population-based surveys always overestimates the actual coverage. The degree of overestimation increases with the decrease of the actual coverage and specificity.

In summary, despite moderate and varying validity, the population-level bias in coverage estimates was mostly small to moderate in this study, particularly for indicators with high recorded coverage or low recorded coverage but high specificity. Of note, although the bias may not be large at the population level, the degree of misclassification at the individual level could still be large due to unsatisfactory validity of some indicators.

Our study is subject to a number of limitations. First, our reference standard has some caveats. The fact that the self-reported coverage of a number of indicators had low or lower than expected specificity, including that of cesarean section which would normally have closer to 100% specificity, suggests that the

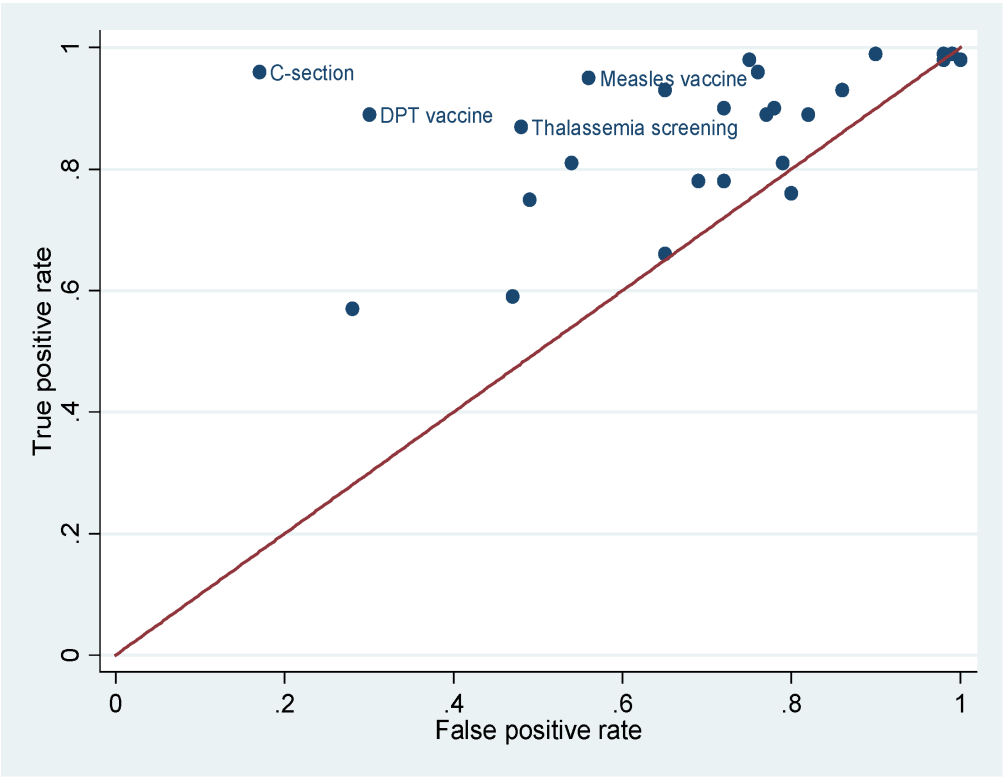


**Figure 1. ROC plot of validated indicators.** The red line indicates the diagonal.
doi:10.1371/journal.pone.0060762.g001

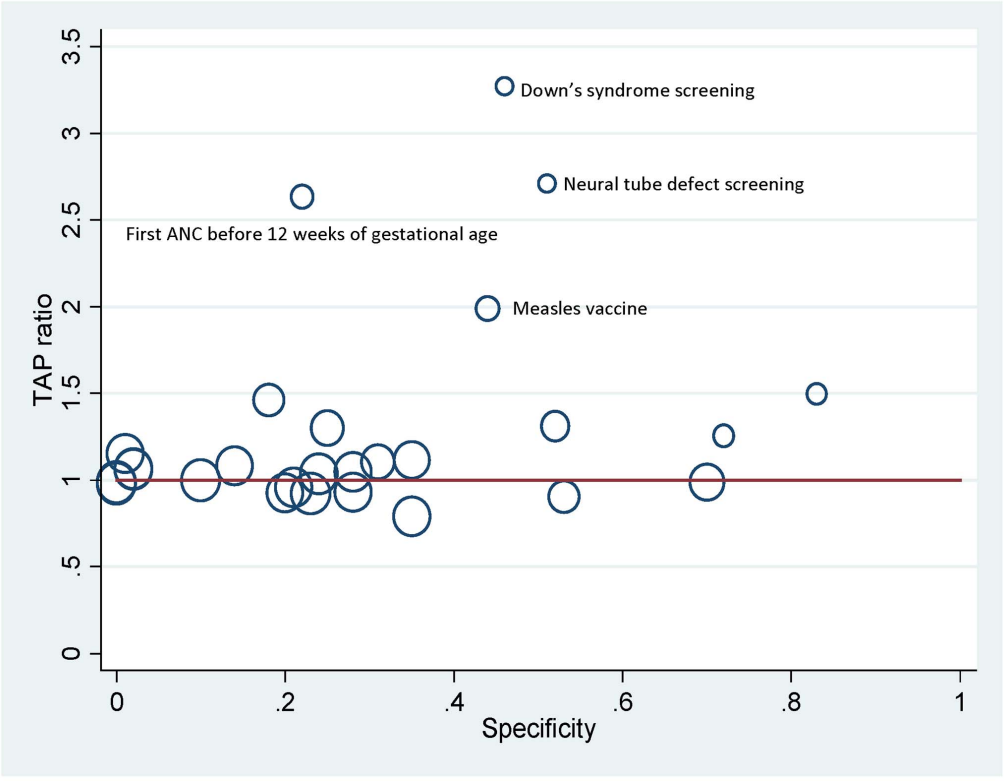


**Figure 2. TAP ratio by specificity of self-reported coverage of selected MNCH indicators.** The size of the circle represents the recorded coverage in the GMNCHIS.
doi:10.1371/journal.pone.0060762.g002

**Table 3.** Validity of selected MNCH indicators.

| Indicators | Sensitivity | | Specificity | | AUC | |
|---|---|---|---|---|---|---|
| | Est. | 95% CI | Est. | 95% CI | Est. | 95% CI |
| **Routine antenatal care** | | | | | | |
| First antenatal visit <12 weeks of gestational age | 0.90 | (0.86, 0.94) | 0.22 | (0.19, 0.26) | 0.56 | (0.54, 0.59) |
| At least 4 antenatal visits | 0.98 | (0.96, 0.99) | 0.25 | (0.19, 0.32) | 0.62 | (0.58, 0.65) |
| Weight measurement | 0.98 | (0.97, 0.99) | 0.00 | (0.00, 0.00) | 0.49 | – |
| Height measurement | 0.89 | (0.87, 0.91) | 0.23 | (0.08, 0.37) | 0.56 | (0.48, 0.63) |
| Blood pressure measurement | 0.98 | (0.97, 0.99) | 0.00 | (0.00, 0.00) | 0.49 | (0.49, 0.50) |
| Hemoglobin test | 0.93 | (0.91, 0.95) | 0.14 | (0.08, 0.21) | 0.54 | (0.50, 0.57) |
| Urine test | 0.99 | (0.98, 1.00) | 0.02 | (0.00, 0.05) | 0.50 | (0.48, 0.52) |
| Fetal heart rate monitoring | 0.99 | (0.98, 1.00) | 0.10 | (0.00, 0.29) | 0.54 | (0.45, 0.64) |
| Ultrasound exam | 0.99 | (0.98, 1.00) | 0.01 | (0.00, 0.03) | 0.50 | (0.49, 0.51) |
| **Screening for STD** | | | | | | |
| HIV test | 0.59 | (0.54, 0.65) | 0.53 | (0.46, 0.60) | 0.54 | (0.51, 0.56) |
| HBsAb test | 0.89 | (0.86, 0.92) | 0.18 | (0.13, 0.22) | 0.56 | (0.52, 0.60) |
| **Screening for congenital abnormalities** | | | | | | |
| Down's syndrome screening | 0.81 | (0.73, 0.88) | 0.46 | (0.41, 0.5.0) | 0.63 | (0.59, 0.68) |
| Neural tube defect screening | 0.75 | (0.67, 0.84) | 0.51 | (0.46, 0.56) | 0.63 | (0.58, 0.68) |
| Thalassemia screening | 0.87 | (0.80, 0.95) | 0.52 | (0.41, 0.64) | 0.70 | (0.63, 0.76) |
| **Delivery care** | | | | | | |
| Cesarean section | 0.96 | (0.93, 0.99) | 0.83 | (0.80, 0.86) | 0.90 | (0.88, 0.92) |
| **Postnatal care** | | | | | | |
| At least one postnatal visit | 0.57 | (0.50, 0.63) | 0.72 | (0.68, 0.76) | 0.64 | (0.60, 0.68) |
| Blood pressure | 0.90 | (0.85, 0.95) | 0.28 | (0.10, 0.46) | 0.59 | (0.50, 0.68) |
| Temperature | 0.93 | (0.87, 0.98) | 0.35 | (0.15, 0.54) | 0.64 | (0.53, 0.74) |
| Breast exam | 0.66 | (0.57, 0.74) | 0.35 | (0.16, 0.53) | 0.50 | (0.40, 0.60) |
| Uterus exam | 0.76 | (0.68, 0.84) | 0.20 | (0.04, 0.36) | 0.48 | (0.39, 0.57) |
| Lochia exam | 0.78 | (0.71, 0.85) | 0.28 | (0.10, 0.46) | 0.53 | (0.43, 0.63) |
| Perineum exam | 0.81 | (0.74, 0.88) | 0.21 | (0.05, 0.37) | 0.51 | (0.42, 0.60) |
| Family planning advice | 0.78 | (0.69, 0.86) | 0.31 | (0.18, 0.45) | 0.54 | (0.46, 0.62) |
| **Child vaccination** | | | | | | |
| BCG vaccine | 0.96 | (0.94, 0.97) | 0.24 | (0.13, 0.35) | 0.60 | (0.54, 0.66) |
| Polio vaccine | 0.86 | (0.84, 0.89) | – | – | – | – |
| HBV vaccine | 0.98 | (0.97, 0.99) | 0.02 | (0.00, 0.05) | 0.50 | (0.48, 0.52) |
| DPT vaccine | 0.89 | (0.86, 0.92) | 0.70 | (0.61, 0.78) | 0.80 | (0.75, 0.84) |
| Measles vaccine | 0.95 | (0.92, 0.98) | 0.44 | (0.38, 0.49) | 0.69 | (0.66, 0.72) |

BCG, Bacillus Calmette-Guérin; DPT, diphtheria-pertussis-tetanus; HBV, hepatitis B virus.
doi:10.1371/journal.pone.0060762.t003

quality of the GMNCHIS is perhaps not optimal. For a distinctive event like cesarean section, self-reporting might even be more reliable. MNCH services received outside the study county may not be recorded in the GMNCHIS, although one had to deliver in Gongcheng to be included in the study. Anecdotal evidence also suggests that the completeness of the electronic system has improved over time since its initiation in 2007. This was further supported by the finding that live births were significantly less likely to be matched using the electronic system in 2007–2009 compared to more recent years ($p<0.001$). However, our sensitivity analysis shows that the validity is the same between women who gave birth in the past five years and those in the past two years. This suggests that although the completeness may be lower in 2007–2009, among women recorded in the electronic system, the validity of their self-reports in the past five years is similar to that in the past two years. If however, self-reports of women that were not captured by the electronic system in its early stage had very different validity from those captured later, the limited completeness may introduce bias to our findings, although the direction is difficult to determine.

Second, our study sample may not necessarily be representative of the whole county. The study participants were drawn from village vaccination rosters, which may have missed children born outside of the national family planning policy. However, this bias is likely to be small as Gongcheng is a minority-concentrated area and such women could usually have up to two children. In addition, within the primary sampling units, study participants were recruited based on availability by going through the

vaccination rosters until the desired sample size was reached. This process may not be completely random, which could introduce selection bias, although we do not have reason to believe that the bias is systematic. The fact that our matched and unmatched study sample differed in mothers' ethnicity, number of live births in the past five years, and children's age may also introduce bias if these characteristics are associated with recall accuracy. However, the lack of representativeness does not affect the validation results, but may limit the study generalizability.

Third, there are other factors that may limit our generalizability. The study was conducted in a setting where coverage of selected MNCH indicators was in general higher than that in countries where DHS and MICS surveys are normally conducted. In addition, we conducted this study in an area that is relatively more developed than those where most DHS and MICS surveys are usually conducted. If socio-economic development factors, such as education, are associated with validity as found previously [10,11], the study results could not be directly applied to other settings with different development levels. Despite these limitations, our study results could be subject to fewer selection biases and be more generalizable than other facility-based studies in this Collection, although they would have a higher-quality reference standard based on direct observation [15–18].

Factors associated with the design and implementation of the survey may also affect the external validity of the study. We interviewed women in a central location in the community rather than in their households, which may affect validity of certain indicators. However, we speculate that this influence is likely to be small for the indicators studied, most of which are not sensitive at all in this context. Validity or reliability of questions included in the survey instruments could also affect the study's internal and external validity. For instance, we failed to include the age limit of the measles vaccine in the questionnaire, which is 8 months or older in China [33]. As a result, the coverage of measles vaccine had high false positive rate and large bias. It is illustrated by the fact that children older than 8 months only constituted 60% of the matched live births, whereas the measles vaccine coverage rate was reported to be 70%, which is unlikely to be true.

In conclusion, more population-based validation studies are warranted with an improved reference standard and survey instruments. Future research should further examine the generalizability of observed validity to other LMIC settings. Nevertheless, the current study contributes to our understanding of validity of self-reported coverage of a range of MNCH interventions. It provides insights into the population-level accuracy of self-report based on a population survey in the LMICs.

## Supporting Information

**Table S1** List of the 28 matched indicators and the corresponding questions used in the community questionnaire, in comparison with those used in the DHS and MICS.
(DOCX)

## Acknowledgments

We are thankful for the inputs from Melinda Munos and Ingrid Friberg during the design stage of the study. We also thank Haoying Wang, Ming Chen, Dasheng Luo, Jian Wang, Xiaoming Zhao, Kaiyong Huang, Meihong Yao, Aihu Dong, Rongjun Zhou, and other field researchers for their hard work in collecting valuable study information. We are indebted to Gongcheng Department of Health staff members and doctors of the sampled townships and villages for their support to organize and facilitate the community survey. We also acknowledge the contribution of our study participants, without whom the study would not have been possible.

## Author Contributions

Interpretation of results and subsequent revision of the manuscript: LL ML LY LJ BT NW JB HC REB YG. Conceived and designed the experiments: LL NW JB HC RB YG ML. Performed the experiments: LL ML LJ LY BT YG. Analyzed the data: ML LL. Wrote the paper: LL ML.

## References

1. Glass RI, Guttmacher AE, Black RE (2012) Ending preventable child death in a generation. JAMA 308: 141–142. doi:10.1001/jama.2012.7357.
2. The Countdown Group (2012) Building a future for women and children: The 2012 Countdown Report. Geneva: WHO.
3. Hancioglu A, Arnold F (2013) Measuring coverage in MNCH: Tracking progress in health for women and children using DHS and MICS household surveys. PLoS Med 10: e1001391. doi:10.1371/journal.pmed.1001391.
4. Hewson D, Bennett A (1987) Childbirth research data: medical records or women's reports? Am J Epidemiol 125: 484–491.
5. Quigley MA, Hockley C, Davidson LL (2007) Agreement between hospital records and maternal recall of mode of delivery: Evidence from 12,391 deliveries in the UK Millennium Cohort Study. BJOG 114: 195–200. doi:10.1111/j.1471-0528.2006.01203.x.
6. Casey R, Rieckhoff M, Beebe SA, Pinto-Martin J (1992) Obstetric and perinatal events: The accuracy of maternal report. Clin Pediatr (Phila) 31: 200–204.
7. Elkadry E, Kenton K, White P, Creech S, Brubaker L (2003) Do mothers remember key events during labor? Am J Obstet Gynecol 189: 195–200. doi:10.1067/mob.2003.371.
8. Olson JE, Shu XO, Ross JA, Pendergrass T, Robison LL (1997) Medical record validation of maternally reported birth characteristics and pregnancy-related events: A report from the Children's Cancer Group. Am J Epidemiol 145: 58–67.
9. Willis G, Brittingham A, Lee L, Tourangeau R, Ching P (1999) Response errors in surveys of children's immunizations. National Center for Health Statistics.
10. Hakim RB, Tielsch JM, See LC (1992) Agreement between maternal interview- and medical record-based gestational age. Am J Epidemiol 136: 566–573.
11. Buka SL, Goldstein JM, Spartos E, Tsuang MT (2004) The retrospective measurement of prenatal and perinatal events: Accuracy of maternal recall. Schizophr Res 71: 417–426. doi:10.1016/j.schres.2004.04.004 [pii];10.1016/j.schres.2004.04.004 [doi].
12. Filippi V, Ronsmans C, Gandaho T, Graham W, Alihonou E, et al. (2000) Women's reports of severe (near-miss) obstetric complications in Benin. Stud Fam Plann 31: 309–324.
13. Ronsmans C, Achadi E, Cohen S, Zazri A (1997) Women's recall of obstetric complications in south Kalimantan, Indonesia. Stud Fam Plann 28: 203–214.
14. Sloan NL, Amoaful E, Arthur P, Winikoff B, Adjei S (2001) Validity of women's self-reported obstetric complications in rural Ghana. J Health Popul Nutr 19: 45–51.
15. Eisele T, Silumbe K, Yukich J, Hamainza B, Keating J, et al. (2013) Measuring coverage in MNCH: Accuracy of measuring diagnosis and treatment of childhood malaria from household surveys in Zambia. PLoS Med 10: e1001417. doi:10.1371/journal.pmed.1001417.
16. Hazir T, Bequem K, el Arifeen S, Khan AM, Huque MH, et al. (2013) Measuring coverage in MNCH: A prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia. PLoS Med 10: e1001422. doi:10.1371/journal.pmed.1001422.
17. Stanton CK, Rawlins B, Drake M, dos Anjos M, Cantor D, et al. (2013) Measuring coverage in MNCH: Testing the validity of women's self-report of key maternal and newborn health interventions during the peripartum period in Mozambique. PLoS ONE 8: e60694. doi:10.1371/journal.pone.0060694.
18. Tunçalp Ö, Stanton C, Castro A, Adanu RK, Heymann M, et al. (2013) Measuring coverage in MNCH: Validating women's self-report of emergency cesarean sections in Ghana and the Dominican Republic. PLoS ONE 8: e60761. doi:10.1371/journal.pone.0060761.
19. Guilin Economic and Social Statistical Yearbook Editorial Board (2007) Guilin Economic and Social Statistical Yearbook 2007. Beijing: China Statistics Press.
20. Gongcheng Yao Autonomous County Statistics Bureau (2010) The sixth national census preliminary results, Gongcheng Yao Autonomous County.
21. Gongcheng County Department of Health (2011) Maternal and child health statistics in Gongcheng County: 2006–2010. Accessed through Xiuqin Tan.
22. Liu Y (2004) China's public health-care system: facing the challenges. Bull World Health Organ 82: 532–538.
23. Royston P, Babiker A (2002) A menu-driven facility for complex sample size calculation in randomized controlled trials with a survival or binary outcome. Stata J 2: 151–163.
24. StataCorp (2007) Stata Statistical Software: Release 10. College Station, TX: StataCorp LP.
25. Institut Superieur des Sciences de la Population (ISSP and Institute for International Programs Johns Hopkins University IIP-JHU) (2012) Evaluation Independante du Projet d'Acceleration de la Reduction de la Mortalite

Maternelle, Neonatale et Infanto-Juvenile dans les Regions Sanitaires du Nord et Centre-Nord au Burkina Faso: Rapport d'Analyse de l'Enquete de Couverture de Base. Ouagadougou, Burkina Faso: ISSP and IIP-JHU.
26. Huang H, Gao X, Guo L (2009) Assessing MCV, EOFT and Hb electrophoresis combination test in thalassaemia screening. Shangdong Medicine 49: 89–90.
27. Liao C, Mo QH, Li J, Li LY, Huang YN, et al. (2005) Carrier screening for alpha- and beta-thalassemia in pregnancy: The results of an 11-year prospective program in Guangzhou Maternal and Neonatal hospital. Prenat Diagn 25: 163–171. 10.1002/pd.1079 [doi].
28. Zhou X, Obuchowski N, McClish D (2002) Statistical methods in diagnostic medicine. New York, NY: Wiley & Sons.
29. Campbell H, Biloglav Z, Rudan I (2008) Reducing bias from test misclassification in burden of disease studies: Use of test to actual positive ratio–new test parameter. Croat Med J 49: 402–414.
30. Anker M (1997) The effect of misclassification error on reported cause-specific mortality fractions from verbal autopsy. Int J Epidemiol 26: 1090–1096.
31. Tourangeau R (1999) Remembering what happened: Memory errors and survey reports. In: Stone AA, Bachrach CA, Jobe JB, Kurtzman HS, Cain VS, editors. The Science of Self-report: Implications for Research and Practice. Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 29–47.
32. Vecchio TJ (1966) Predictive value of a single diagnostic test in unselected populations. N Engl J Med 274: 1171–1173. doi:10.1056/NEJM196605262742104.
33. The Central People's Government of the People's Republic of China (2005) Guideline for National Child Immunization Programme. In: State Council of the People's Republic of China Order Number 434.

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: Indicators for Global Tracking of Newborn Care

**Allisyn C. Moran[1]*, Kate Kerber[2,3], Deborah Sitrin[1], Tanya Guenther[1], Claudia S. Morrissey[1], Holly Newby[4], Joy Fishel[5], P. Stan Yoder[5], Zelee Hill[6], Joy E. Lawn[2]**

**1** Save the Children, Washington, District of Columbia, United States of America, **2** Save the Children, Cape Town, South Africa, **3** University of Western Cape, Belleville, South Africa, **4** Statistics and Monitoring Section, Division of Policy and Strategy, United Nations Children's Fund, New York, New York, United States of America, **5** ICF International, Calverton, Maryland, United States of America, **6** Institute of Child Health, London, United Kingdom

**Abstract:** Neonatal mortality accounts for 43% of under-five mortality. Consequently, improving newborn survival is a global priority. However, although there is increasing consensus on the packages and specific interventions that need to be scaled up to reduce neonatal mortality, there is a lack of clarity on the indicators needed to measure progress. In 2008, in an effort to improve newborn survival, the Newborn Indicators Technical Working Group (TWG) was convened by the Saving Newborn Lives program at Save the Children to provide a forum to develop the indicators and standard measurement tools that are needed to measure coverage of key newborn interventions. The TWG, which included evaluation and measurement experts, researchers, individuals from United Nations agencies and non-governmental organizations, and donors, prioritized improved consistency of measurement of postnatal care for women and newborns and of immediate care behaviors and practices for newborns. In addition, the TWG promoted increased data availability through inclusion of additional questions in nationally representative surveys, such as the United States Agency for International Development–supported Demographic and Health Surveys and the United Nations Children's Fund–supported Multiple Indicator Cluster Surveys. Several studies have been undertaken that have informed revisions of indicators and survey tools, and global postnatal care coverage indicators have been finalized. Consensus has been achieved on three additional indicators for care of the newborn after birth (drying, delayed bathing, and cutting the cord with a clean instrument), and on testing two further indicators (immediate skin-to-skin care and applications to the umbilical cord). Finally, important measurement gaps have been identified regarding coverage data for evidence-based interventions, such as Kangaroo Mother Care and care seeking for newborn infection.

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection.*

## Introduction

Neonatal mortality accounts for 43% of under-five mortality [1] and is becoming a political priority at both global and country levels [2,3]. Ministries of Health, partners, and donors are increasingly focused on implementing packages and specific interventions aimed at improving newborn health and survival [4]. To assess changes and improvements, accurate and accessible data on coverage of evidence-based interventions are essential. However, few of the highest impact interventions for newborn care are systematically measured [4].

Given the relatively recent focus on scale-up of newborn care, and the fact that many of the interventions are linked to packages of care for women or children, there has been limited agreement on indicators to monitor and evaluate newborn care. To address this gap, Save the Children's Saving Newborn Lives program convened the interagency Newborn Indicators Technical Working Group (TWG) in 2008 focused on improving the capture and measurement of newborn care. This group of about 20 individuals includes people from United Nations agencies, nongovernmental organizations and national institutes of health, researchers and academics, evaluation and measurement experts, donors, and other stakeholders. The group aims to reach consensus on key indicators, including definitions and standard measurement tools. It also promotes inclusion of agreed-upon indicators in nationally representative household surveys such as the United Nations Children's Fund–supported Multiple Indicator Cluster Surveys (MICS) and the United States Agency for International Development–supported Demographic and Health Surveys (DHS) [5]. In addition, the TWG disseminates standard indicators and tools for inclusion in specialized data collection efforts at subnational levels. The main TWG group meets twice per year, with subgroups working on particular areas meeting more frequently.

The objective of this article, which is part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, is to describe the process of developing and achieving consensus on indicators to

**Citation:** Moran AC, Kerber K, Sitrin D, Guenther T, Morrissey CS, et al. (2013) Measuring Coverage in MNCH: Indicators for Global Tracking of Newborn Care. PLoS Med 10(5): e1001415. doi:10.1371/journal.pmed.1001415

**Academic Editor:** Carla AbouZahr, Independent consultant: Health statistics and policy

**Published** May 7, 2013



**Funding:** Funding for the Newborn Indicators TWG was provided through Save the Children's Saving Newborn Lives Program, supported by The Bill & Melinda Gates Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** DHS, Demographic and Health Surveys; MICS, Multiple Indicator Cluster Surveys; TWG, Newborn Indicators Technical Working Group; UNICEF, United Nations Children's Fund; WHO, World Health Organization.

* E-mail: allisynmoran@gmail.com

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

monitor and evaluate coverage of evidence-based interventions for newborns. The TWG's initial focus was on three areas of measurement: postnatal care, immediate care behaviors and practices for newborns, and health facility assessments. We will focus on achievements in the first two areas, provide recommendations for indicators and tools, and identify areas for future research efforts.

## Measurement of Coverage of Postnatal Care

Around three-quarters of neonatal and maternal deaths occur in the first week of life, with up to half in the first 24 hours [6]. Postnatal care to provide counseling and to identify and treat complications after birth for both the mother and her baby was identified by the World Health Organization (WHO) as a critical care package at an expert review meeting in 1997 and has been included in its guidelines for pregnancy, childbirth, postnatal, and newborn care since 2003 [7]. Research has demonstrated that a postnatal care home visit from a trained provider within two days of delivery can result in 30% to 60% reduction in neonatal mortality [8–10]. The evidence is less clear for maternal outcomes, mostly due to challenges with measuring maternal mortality. Nevertheless, the 2003 WHO guidelines integrate care for the woman and for her baby from shortly after childbirth until six weeks after birth, as such care is often provided at the same time by the same provider. In 2007, postnatal care was redefined as being for both woman and newborn in the context of the continuum of care [11]. Further, the Countdown to 2015: Maternal, Newborn and Child Survival initiative defines postnatal care as integrated care for women and their newborns [12,13].

Given the importance of the postnatal period and the clear evidence of the effectiveness of early postnatal care for newborns, an indicator that measures a contact with a health worker to provide counseling and to assess health within the first two days after birth is critical for global monitoring. Countdown to 2015, which was established in 2005, includes early postnatal care for women and newborns in its initial list of priority indicators [13]. More recently the Commission on Information and Accountability for Women's and Children's Health prioritized early postnatal care as one of seven selected coverage indicators to measure progress for maternal, newborn, and child health [14]. Although there is consensus on the importance of care provided during this time period, there are many challenges associated with measuring coverage in large-scale household surveys (Table 1). Each of these challenges was discussed during TWG meetings, research and secondary analyses were conducted, and consensus on an indicator for postnatal care within two days of birth regardless of delivery location was achieved (Table 2). Following TWG deliberations, this indicator has been tested by MICS, and data collection has been harmonized across DHS and MICS surveys [15].

### Measurement Challenges and Recommendations

Similar to other long-standing, global indicators such as antenatal care and skilled attendant at delivery that focus on contact with the health system, the measurement focus for postnatal care is a contact with a health care provider. This contact provides the platform for delivering lifesaving interventions, but measurement of contact does not provide any indication about the content or quality of care. Measurement of the first postnatal care contact provides a complementary indicator to antenatal and delivery care, and extends the ability to monitor contacts across the continuum of care from pregnancy through childbirth and the postnatal period [16].

Postnatal care includes a specific contact for both the woman and her newborn. It is distinct from intrapartum care, which is provided by the delivery attendant to the woman and her child around the time of birth. Postnatal care focuses on prevention of complications through counseling on important health messages, assessment of the woman and newborn, and referral for complications, if identified. Although a postnatal contact should contain elements of care for both women and newborns, this does not always happen in programs. As a result, for the time being it is necessary to measure postnatal care for the woman and newborn separately to ensure that an inclusive contact has taken place [16].

DHS surveys collect data on postnatal care for women who gave birth either at home or in a health facility, but, historically, most surveys have collected postnatal care data for newborns only for those born at home. This distinction is based on the untested assumption that women who give birth in facilities will have difficulty knowing whether their newborn received postnatal care. Based on recent qualitative work in Bangladesh and Malawi [17] and in Ghana (Hill et al., unpublished report, 2010, Institute for Child Health, University of London), in which women with facility deliveries were able to answer questions about postnatal care contact for their newborns (Box 1), the standard indicator has now been modified to include all births regardless of place of delivery (Tables 1 and 2). The reference period has also been revised to include births occurring within two years preceding the survey (as opposed to five years), which is long enough to provide enough cases for analysis, but not so long that it may hamper recall. To date, 25 countries have collected data on postnatal care coverage for women, and four have collected data on postnatal care coverage for newborns using this definition [13].

All health care providers are included in the current definition of the global postnatal care indicator (including traditional birth attendants and community health workers), regardless of the skill level of the provider. This differs from the current global standards for at least one antenatal care visit and skilled attendance at delivery, which are based on contact with a skilled provider. Although information on where postnatal care took place (e.g., at home or at a health facility) is of programmatic interest, it is not reported on separately in the global indicator. Postnatal care can be effectively provided by a variety of health workers, ranging from community health workers to skilled providers, and may differ for the woman and newborn, as well as the place of care [10]. The global indicator therefore includes postnatal care for mother and newborn by any provider, but standard tables in both DHS and MICS reports disaggregate coverage by type of provider and place of birth.

Another critical challenge facing the measurement of postnatal care contact is the element of timing. The postnatal period refers to the first 42 days following delivery for both the woman and the newborn. Recent WHO and UNICEF guidelines recommend a postnatal care visit for mother and newborn on day 1, day 3, and day 7 after birth, with continuing contacts throughout the first six weeks of life [10]. Evidence from Bangladesh has demonstrated that a postnatal visit to the newborn in the first 48 hours can significantly reduce mortality, whereas first postnatal contacts after that time were not associated with reduced mortality [8]. Postnatal care within two days of delivery aligns well with measurement in household surveys, given that there is potential misclassification in recall on the day of delivery and the day after delivery.

The global postnatal care indicators recommended by the TWG are useful for standard measurement across large-scale household surveys to monitor progress toward achieving coverage along the continuum of care. The global indicators alone, however, are not sufficient for programmatic needs. Programs that focus on improving maternal and newborn survival and

**Table 1.** Postnatal care indicator: measurement issues and advances.

| Topic | Issue(s) | What Has Been Accomplished | What Needs to Be Done |
|---|---|---|---|
| Recall and validity | • Uncertainty about mother's knowledge about what happened to baby after birth, especially for facility births<br>• Lack of recall of past births up to five years prior to survey<br>• Potential misunderstanding of survey questions on postnatal care | • Formative research indicates that women have a good idea what happens to their baby regardless of where they deliver<br>• DHS and MICS questionnaires revised to include postnatal care for all newborns, regardless of place of birth<br>• Formative research indicates that women have difficulty understanding the term "postnatal care"<br>• DHS and MICS questionnaires revised to include an introductory statement for postnatal care questions<br>• Standard tables in DHS and MICS updated to include postnatal care coverage for newborns | • Review data from new DHS and MICS questionnaires and revise tools as needed |
| Timing | • Lack of criteria to distinguish between intrapartum and postnatal care, e.g., should all contacts from birth count, or is postnatal care valid only if it takes place after the intrapartum period<br>• Potential overestimation of true postnatal contacts | • Detailed postnatal care module developed and tested for MICS<br>• Distinct measurement of the first pre-discharge and first post-discharge contact in MICS | • WHO recommendation to define postnatal contact (e.g., a cutoff of one hour after birth)<br>• Formative research to differentiate intrapartum and postnatal care<br>• Implementation research on facility pre-discharge checklist, etc. |
| Number of visits | • Only the first postnatal contact captured in DHS and MICS surveys, and may be an intrapartum contact but no further question asked<br>• Lack of ability to capture pre-discharge postnatal care more accurately<br>• Lack of data on home or clinic postnatal contacts after facility births | • Detailed postnatal care module developed and tested for MICS<br>• Distinct measurement of the first pre-discharge and first post-discharge contact in MICS<br>• No information captured on subsequent postnatal contacts | • Capture additional visits in optional module or specialized surveys |
| Content | • Data not currently collected in national surveys | • Formative research indicates women could recall specific actions for newborns during postnatal care (such as use of equipment, undressing baby, giving advice)<br>• Consensus on five measureable signal functions for postnatal care for newborns | • Work with maternal health community on signal functions for postnatal care for women<br>• WHO meeting to define postnatal care interventions for women and newborns<br>• Test household survey module for content of postnatal care for newborns and women |

doi:10.1371/journal.pmed.1001415.t001

health should include additional questions on other key elements of postnatal care in the surveys used to monitor progress, including timing of each postnatal care contact, location and provider of postnatal care, and, ideally, content of care. The TWG has agreed on five signal functions to assess the content of postnatal care for the newborn. These functions are part of the recommended postnatal care package and are considered feasible for reporting by women in surveys based on qualitative work in Ghana (Box 1) and on analysis of Saving Newborn Lives household surveys (Box 2). The signal functions are (1) checking the newborn's umbilical cord, (2) assessing the newborn's temperature, (3) observing/counseling on breastfeeding, (4) counseling on newborn danger signs, and (5) weighing the baby, if appropriate for a given country. The TWG suggests including these signal functions in an optional module in nationally representative surveys, and in smaller subnational surveys where there is interest in exploring these areas, although consideration must be given to the length of questionnaires as well as data quality. Finally, there have also been preliminary discussions with the TWG and external experts on signal functions for postnatal care for the woman, but more discussion is needed with the wider community.

## Remaining Research Gaps

The first important area where additional research is needed is the timing of postnatal care. The end of the postnatal period is relatively well defined, although there is global discussion about extension of the postnatal period for the woman beyond the currently accepted 42 days after delivery. However, there is much less consensus around when the intrapartum period ends and the postnatal period begins. For example, maternal health experts

**Table 2.** Global indicators for postnatal care coverage, 2010.

| Category | Indicator | Numerator | Denominator |
|---|---|---|---|
| Postnatal care for women | Percent of women who received postnatal care within two days after last delivery | Number of women who received postnatal care within two days after last delivery | Number of women with a live birth in the last two years |
| Postnatal care for newborns | Percent of newborns who received postnatal care within two days after delivery | Number of newborns who received postnatal care within two days after delivery | Number of women with a live birth in the last two years |

doi:10.1371/journal.pmed.1001415.t002

## Box 1. Qualitative Research Studies to Assess Postnatal Care and Immediate Care Behaviors and Practices for Newborns

The TWG supported two qualitative studies to assess postnatal care and immediate care behaviors and practices for newborns—one in rural Ghana in 2010 (Hill et al., unpublished report, 2010, Institute for Child Health, University of London) and one in Bangladesh and Malawi in 2009 [17]. Both studies used birth narratives to describe the birth process and immediate newborn care. The narrative approach is used to understand what is salient by probing specific areas, an approach that is more appropriate for low-income developing country settings than cognitive testing. In Ghana, the narratives were supplemented with focus groups to assess the terminology used to describe postnatal care and interviews with health workers. In Bangladesh and Malawi, the narratives were followed by a structured questionnaire on specific care behaviors and practices, including the timing of events. Women with a birth within the past year were interviewed in both studies, stratified by place of birth and time since delivery.

In Ghana (40 birth narratives, four focus groups, ten health worker interviews)

- Questions on postnatal care were not easily understood by women.
- Few women gave spontaneous answers about the postnatal care that they or their newborn had received.
- Probing phrases about specific activities that could take place during postnatal care were essential.
- Questions had to be carefully phrased to allow for local terminologies and perceptions.
- Mothers were most likely to report an observed specific action during postnatal care such as the use of equipment or undressing the baby.
- Being tired or sick after delivery, time since delivery, number of visits, outcome of the visits, and where the newborn was cared for after facility-based birth influenced recall of postnatal care.

In Bangladesh and Malawi (84 and 80 narratives/questionnaires, respectively)

- The narratives indicated that women can recall the main birth events (their labor pains, the delivery of the placenta, cord care, and newborn care) regardless of place of delivery.
- There was limited difference in recall between home and facility births.
- The majority of women gave nonnumeric answers to questions about timing of events.
- When asked how long after delivery their checkup occurred, answers ranged from minutes to days.
- Many women did not understand what was meant by a postnatal care contact defined as a "health checkup," or "a check on your health" after the baby was born unless study teams defined the term. In Malawi, interviewers provided examples of postnatal care content; in Bangladesh, interviewers described what was meant by health in general.

## Box 2. Survey Questions on Newborn Care Practices: Summary Results and Recommendations

Save the Children's Saving Newborn Lives program conducted household surveys in five countries in Asia, Africa, and Latin America, maintaining standardized measurement of key indicators (drying, delayed bathing, and cutting the cord with a clean instrument) across countries as much as possible. Multiple questions were often used to measure the same indicator, which provided the opportunity to test different question formulations. In each country, cross-sectional household surveys with pre/post design using community cluster sampling were conducted among women with a live birth in the 12 months prior to the survey (nine months for Bangladesh). Countries surveyed included Malawi (2011, $n = 900$), Bangladesh (2010, $n = 794$), Nepal (2011, $n = 630$), Viet Nam (2011, $n = 1{,}050$), and Indonesia (2011, $n = 400$). Target sample sizes were calculated based on expected changes in key indicators, ranging from around 600 women in Nepal to 900 women in Malawi.

- For drying and delayed bathing there were low rates of "don't know" or missing data responses (<10%).
- For cord care, there were differences between women with home and with facility births.
- For home births, women were more able to report the type of instrument used to cut the cord, compared with facility births.
- Rates of "don't know" and missing data among women with facility births were around 30% in Malawi, compared with less than 5% among women with home births.

Based on these findings, the TWG recommends the questions shown in Figure 3 for use in household surveys.

typically define the intrapartum period as ending after the delivery of the placenta, whereas newborn health experts typically set one hour after birth as the end of this period. As there are few countries with postnatal care coverage data for newborns among all births, to get an idea of how much reported postnatal care might actually be intrapartum care, we used recent DHS data from 35 countries to assess postnatal care coverage for women [18]. Our analysis indicates that 35% of postnatal care contacts occur within six hours of birth (Figure 1). Thus, many of these contacts, though important, could be part of routine intrapartum care and not what most programs would consider distinct postnatal care contacts.

WHO conducted a meeting in September 2012 to review postnatal care guidelines, including recommendations for timing of discharge after facility delivery, the number and timing of postnatal care contacts, and the content of those contacts. Part of the discussion was focused on defining the "postnatal period" and differentiating postnatal care from intrapartum care for both women and newborns (results are forthcoming), which may assist in resolving some of the measurement challenges referred to above. However, it is important to consider that even if consensus on a particular cutoff were reached, it would be difficult to operationalize this definition in large-scale household surveys. First, women's recall of a contact occurring before, at, or after four, five, or six hours after birth may not be reliable, especially two years after the birth. Second, standard data collection currently includes only the first postnatal care contact, so if

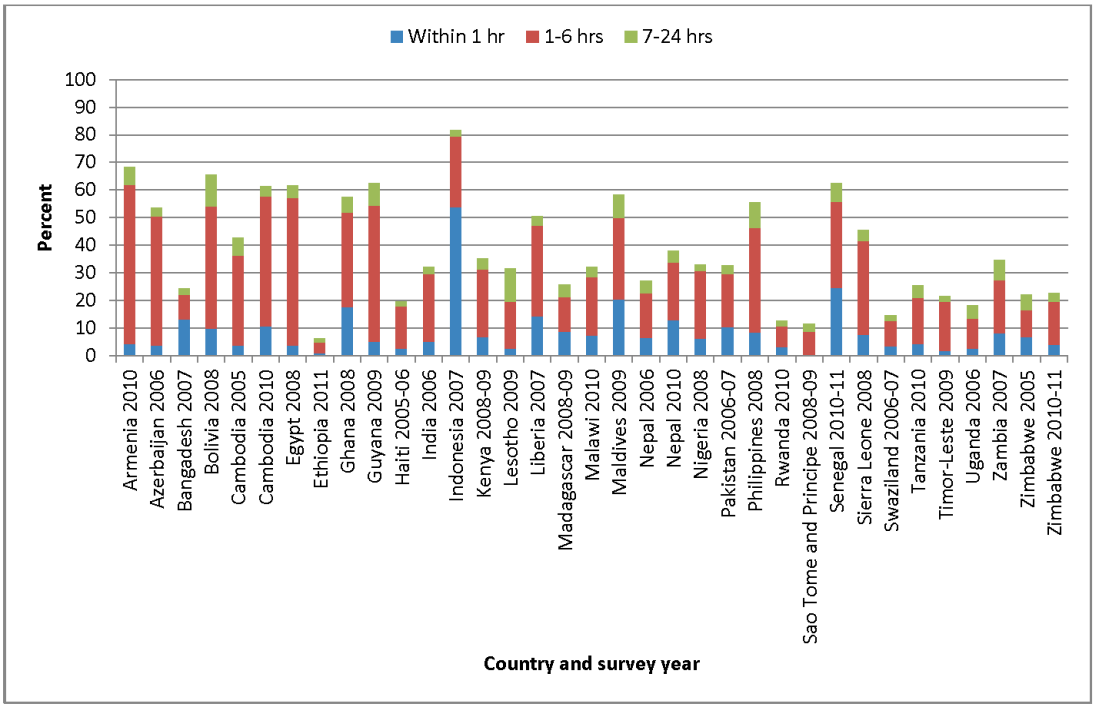


**Figure 1. Proportion of women who received postnatal care within two days of delivery by time of first visit, DHS survey data 2005–2011 [18].**
doi:10.1371/journal.pmed.1001415.g001

additional contacts take place, women do not have the opportunity to report them in the survey, which may result in an underestimation of the overall coverage.

Another important research gap concerns the providers of postnatal care. Currently, as discussed earlier, the global indicator for postnatal care includes all health care providers regardless of

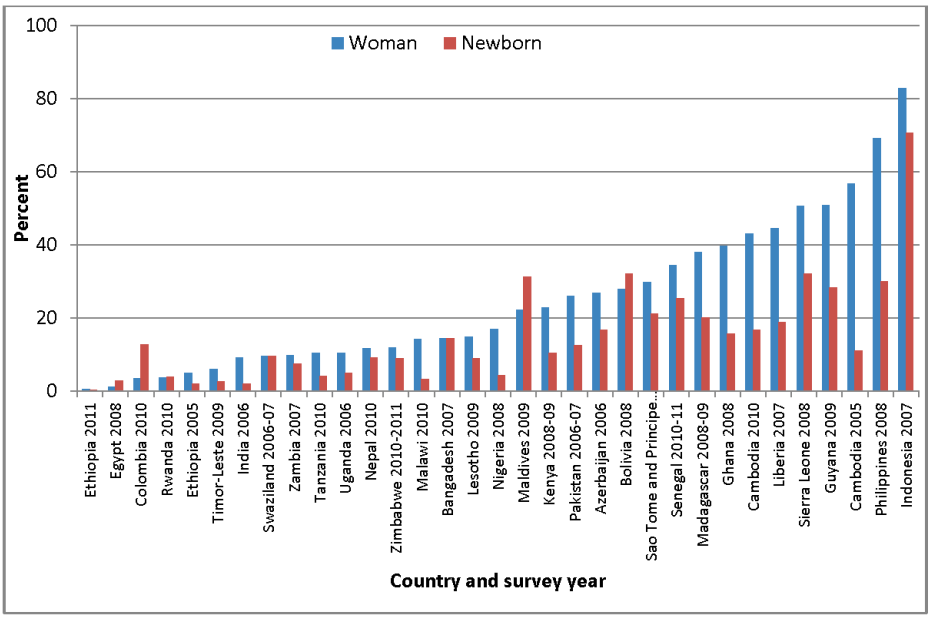


**Figure 2. Proportion of home births for which women and babies received postnatal care within two days of delivery, DHS survey data 2005–2011 [18].**
doi:10.1371/journal.pmed.1001415.g002

**Table 3.** Recommended indicators for care behaviors and practices for newborns.

| Program Element | Indicator | Numerator | Denominator | Comments | Recommended Question(s) |
|---|---|---|---|---|---|
| **Recommended** | | | | | |
| Thermal care: drying | Percent of newborns dried after birth | Number of newborns dried after birth | Number of live last births in the *X* years prior to the survey | All births; timing assessment optional | Was (NAME) dried (wiped) after delivery? |
| Thermal care: delayed bath | Percent of newborns with first bath delayed at least six hours after birth | Number of newborns with first bath delayed at least six hours after birth | Number of live last births in the *X* years prior to the survey | All births; different timing categories can be calculated | How long after delivery was (NAME) bathed for the first time? |
| Cord care: clean cord cutting | Percent of newborns with cord cut with clean instrument | Number of newborns with cord cut using new blade or boiled instrument | Number of live last births (at home) in the *X* years prior to the survey | Home births only; questions on use of clean delivery kits can be included | What was used to cut the cord? Was the instrument used to cut the cord boiled prior to use? |
| **Additional testing needed** | | | | | |
| Thermal care: skin-to-skin | Percent of newborns placed on the mother's bare chest after delivery | Number of newborns placed on the mother's bare chest after delivery | Number of live last births in the *X* years prior to the survey | All births; timing assessment optional | After the birth, was (NAME) put directly on the bare skin of your chest? (Show mother example of skin-to-skin position) |
| Cord care: applications to the umbilical cord | Percent of newborns with nothing (harmful) applied to cord | Number of newborns with nothing (harmful) applied to cord | Number of live last births in the *X* years prior to the survey | All births; "harmful" to be defined locally | Was anything applied to the cord after the cord was cut and tied, until the cord fell off? (If "yes") What was applied to the cord? |

Surveys will vary in period of recall. Typically, DHS surveys use a recall period of five years, while MICS surveys use a two-year period. Interviewer records all substances put on the cord from cutting until it falls off. Harmful substances are determined locally and split out during analysis.

doi:10.1371/journal.pmed.1001415.t003

their skill level. More research is needed to determine whether traditional birth attendants should be removed from the indicator definition since they are rarely trained in postnatal care programs. Although it can be difficult to identify skilled health personnel reliably, other indicators such as antenatal care and skilled birth attendance also rely on this identification [16].

Another issue for future consideration is the feasibility of combining the measurement of postnatal care contacts for women and newborns. Ideally, these contacts should take place at the same time and include elements for both the woman and her newborn. However, based on our analysis of DHS data from 33 countries between 2006 and 2011 [18], there is great variability in terms of postnatal care coverage for women and newborns (Figure 2). In Bangladesh and Zambia, postnatal care coverage for home births is similar for women and their newborns (14% for both women and newborns in Bangladesh; 10% for women and 8% for newborns in Zambia). However, in Ghana, 40% of women who gave birth at home reported receiving postnatal care for themselves, but only 16% reported that their newborns received care. Similarly, in Cambodia, 43% of women who gave birth at home reported receiving postnatal care for themselves, but only 17% reported postnatal care for their newborns. This discrepancy, which may be due to measurement error, to misunderstanding the questions, or to actual differences in coverage, requires further exploration, especially as more comparable data (for all births regardless of place of delivery) become available in the next several years.

## Measurement of Coverage of Immediate Care Behaviors and Practices for Newborns

The second area of measurement that the TWG specifically concentrated on during its deliberations is care of the newborn immediately after birth, which includes a number of functions. Three of these functions can be provided at both community and facility levels: (1) thermal care (drying, skin-to-skin care, and delayed bathing), (2) hygienic and clean skin and cord care (clean cord cutting and dry cord care; hand washing prior to delivery), and (3) breastfeeding (immediate initiation; not discarding colostrum; no foods other than breast milk) [19]. If practised routinely, these practices and care behaviors could reduce newborn deaths by up to 30% [19]. It is therefore vital to measure their coverage in valid and comparable ways across countries. Unfortunately, with the exception of breastfeeding, national surveys do not routinely include information on these care behaviors. Moreover, although some South Asian countries (e.g., Nepal, Bangladesh, and India) have incorporated some of these indicators into their DHS surveys, the questions and response categories vary and are not comparable across countries. For example, in India, women are asked using a prompted question if the baby was immediately wiped dry and then wrapped, while in Nepal, women are asked separate questions about how long after birth the baby was dried and then how long after birth the baby was wrapped. The TWG has not addressed indicators for immediate behaviors and practices relevant to the mother and her care, and would encourage other groups to address this important issue.

Similar to measurement concerns around postnatal care contacts, there are methodological issues with the measurement of immediate care behaviors and practices for newborns that relate to the recall and timing of specific actions. To tackle these methodological issues, the TWG developed a list of recommended indicators, including drying/wrapping the baby, delayed bathing, and cord care. Qualitative work to assess these indicators and address measurement issues was conducted in Bangladesh and Malawi (Box 1), and different questions were tested within surveys undertaken by Save the Children in 2010 and 2011.

Now I'm going to ask you some questions about what happened soon after (NAME OF BABY) was born.

| | | | |
|---|---|---|---|
| 1 | Was (NAME) dried (wiped) after delivery? | Yes.....1<br>No.....2<br>Don't know . . . .....98 | <br>→ 2<br>→ 2 |
| 1a | How soon after birth was (NAME) dried (wiped)?<br><br>IF LESS THAN ONE HOUR, RECORD 00.<br><br>*Optional question and may be removed.* | Immediately: 00<br>HOURS: __\| \|<br><br>Don't know.....998 | |
| 2 | How long after delivery was (NAME) bathed for the first time?<br><br>IF LESS THAN ONE HOUR, RECORD 00.<br><br>*May also record Days in areas where it is recommended to delay bathing >1 day. Otherwise, days can be recorded in hours (1day=24hrs, 2days=48hrs, 3days=36hrs, etc)* | Immediately: 00<br>HOURS: __\| \|<br><br>Don't know.....998 | |
| 3 | ASK FOR HOME BIRTHS ONLY:<br>What was used to cut the cord?<br><br>*List of answer options may be modified.* | Blade from delivery bag or New blade.....1<br>Blade used for other purposes.....2<br>Sickle .....3<br>Scissor .....4<br>Other, Specify: _____<br>Don't Know/Can't Remember.....98 | →<br>END |
| 4 | Was the instrument used to cut the cord boiled prior to use? | Yes.....1<br>No.....2<br>Don't know . . . .....98 | GO<br>TO<br>END |
| 5 | ASK FOR FACILITY BIRTHS ONLY:<br>Did you bring a Clean Delivery Kit or a new blade to the facility the time you gave birth to (NAME)?<br><br>*Question may not be appropriate in all settings and may be removed.* | Yes.....1<br>No.....2<br>Don't know . . . .....98 | |

**Figure 3. Standard questionnaire for measuring coverage of immediate newborn care.**
doi:10.1371/journal.pmed.1001415.g003

## Measurement Challenges and Recommendations

The qualitative research (Box 1) indicates that women can recall the event sequence for delivery and immediate care practices for newborns, that there is no difference in recall between women with facility-based births and home births, and that the timing of the birth relative to the survey has no effect on recall. Although women did have difficulty recalling the exact timing of events as measured in hours and minutes, women with facility births could estimate the amount of time elapsed between delivery and immediate newborn care. This finding is replicated in other studies [20,21]. We recommend, therefore, that questions about care behaviors and practices for newborns should be asked for all live births, but that questions on timing should be simplified by limiting response categories to hours after birth (as opposed to minutes).

Importantly, there was overlap between the measurement of coverage based on the indicators of drying and of wrapping the baby. Based on Save the Children survey data (Box 2), more than 90% of babies who were dried were also wrapped. Thus, it is not necessary to include indicators for both of these actions (data not shown). Survey data were also examined to assess rates of "don't know" and missing responses. Questions around delayed bathing, immediate drying, and cutting the cord for home births had low levels of "don't know" and missing responses (<10%) (Box 2). Women were less able to report on cord cutting for facility births, with "don't know" responses at higher levels. Based on these findings, we recommend three indicators for inclusion in household surveys: (1) percent of newborns dried after birth, (2) percent of newborns with bath delayed at least six hours after birth, and (3) percent of newborns with cord cut with clean instrument (for home births only) (Table 3; Figure 3). These indicators can be collected via specialized surveys, and the TWG is working with DHS and MICS on an optional newborn module. This module could also include questions on postnatal care beyond the first contact, as well as the content of postnatal care. As more data become available, these indicators may need additional revisions.

## Remaining Research Gaps

Skin-to-skin care is an important intervention to maintain thermal regulation and encourage breastfeeding and bonding, but more work is needed to assess the validity of this indicator in terms of women's comprehension of survey questions concerning the intervention, as well as recall. This indicator has been recently tested in a validation study in Mozambique that is described elsewhere in this Collection [22] and is being investigated in

## Key Points

- Neonatal mortality accounts for 43% of under-five mortality; countries are increasingly implementing programs to improve newborn survival, but there is a dearth of data to monitor and evaluate newborn care programs.
- The interagency Newborn Indicators Technical Working Group (TWG) convened by Save the Children's Saving Newborn Lives program has prioritized clarity and consistency of postnatal care metrics within two days of delivery for all births.
- Indicators and survey questions to measure care behaviors and practices in the immediate newborn period have been agreed upon, pertaining to drying of the newborn, delayed bathing, and cutting the cord with a clean instrument
- Two additional indicators related to skin-to-skin care and applications to the umbilical cord are recommended for future testing.
- In the future, the TWG will address metrics for other evidence-based interventions such as Kangaroo Mother Care and care seeking for newborn sepsis, and will work to strengthen national health information systems, including vital registration.

several Saving Newborn Lives surveys. Saving Newborn Lives has also assessed a variety of indicators to look at applications to the umbilical cord after birth. The current WHO recommendation is dry cord care except in settings where the risk of bacterial infection is high [23]. However, recent evidence has demonstrated significant reductions in mortality (up to 23%) among babies who have 4% chlorhexidine applied to the cord after birth [24–26]. Once a global recommendation is finalized, this indicator will require further testing.

## The Way Forward for Newborn Care Indicators

Over the past several years, consistency in the measurement of newborn care interventions has improved. National data on first postnatal care contact for all newborns regardless of place of delivery are now available in four countries, and more should be forthcoming through new surveys. Development of a supplemental module is underway for national surveys to measure immediate care behaviors and practices for newborns and to provide more detailed information on postnatal care content and quality, and for visits beyond the first contact. The TWG provides a valuable forum for discussion around critical indicators using data, experience, and expert opinions.

In the future, the TWG will investigate the measurement of other evidence-based interventions that address the three main killers of newborns—complications of preterm birth, infection, and intrapartum-related deaths due to asphyxia—for which we lack coverage data. For example, facility-based Kangaroo Mother Care (care for preterm or low birth weight infants in which the baby is carried, usually by the mother, with skin-to-skin contact, and in which breastfeeding support is provided) can prevent up to half of neonatal deaths in stable newborns weighing less than 2,000 g [27]. Coverage with this intervention is not currently captured through nationally representative surveys or routine information systems but should be measurable through maternal recall. Indeed, a list of facility-based indicators to capture training of providers and coverage of Kangaroo Mother Care has been agreed upon [28], but these indicators need to be tested, refined, and incorporated into national surveys or project-based surveys. Over 700,000 newborns die of severe infection (mainly sepsis and pneumonia) each year. DHS and MICS surveys capture care seeking and treatment for fever and symptoms of acute respiratory infection for all children under five years. However, the available sample is often too small for disaggregation, and the data are not specifically presented for newborns. In addition, the symptoms assessed include only cough and difficulty breathing, so newborns with additional danger signs that indicate possible severe infection may be missed. Notably, as discussed elsewhere in this Collection, the reliability of the household-survey-based indicator that measures the proportion of children treated for pneumonia is questionable [29], which raises concerns about measuring management of other newborn problems through household surveys. Thus, innovative methodologies to complement survey research need to be developed and reviewed to improve coverage for interventions designed to manage severe infection. It is also essential that additional care provided to preterm babies and resuscitation of asphyxiated babies is tracked. The validity of asking questions about resuscitation in household surveys needs to be assessed. Facility-level data on this intervention from programs such as Helping Babies Breathe should be available soon [30].

Other indicators for maternal and newborn care require additional evaluation and development, especially for quality of care. As the percentage of facility births increases, it will become more and more critical to assess effectiveness and efficiency. Twelve million more women gave birth with a skilled attendant in 2010 compared to 2000; in South Asia, 49% of births are now assisted by skilled personnel, compared to just 30% a decade ago [4]. In this context, understanding what happens before women leave the health facility after giving birth and then at home in the postnatal period is crucial. The MICS4 module addresses these different areas of care, and different methodological approaches should be considered and differences validated [15,16].

The TWG also advocates for improving vital registration (official registration of all births and deaths in a population), which can be linked to facility care at birth and to postnatal care [31]. Measurement of stillbirths, especially intrapartum stillbirths, and of preterm birth and low birth weight should be closely linked to improved measurement of neonatal deaths and to tracking pregnancy outcomes more comprehensively [32]. Notably, the recently published Global Burden of Disease Study 2010 did not count stillbirths, an omission that makes comprehensive tracking of global pregnancy outcomes more difficult [33]. In the future, the TWG plans to improve existing methods to capture stillbirths, by comparing data from pregnancy history modules with that from demographic surveillance sites, and by investigating the validity of shorter pregnancy history modules. The TWG will also work with partners to explore innovative methods to register births, such as using mobile phone technology. Finally, although to date the TWG has focused on measurement of newborn care and coverage through large-scale, nationally representative household surveys, in the future, it will focus more on routine data collection systems, such as health management information systems, to improve the quality of data and its use for decision-making.

## Conclusions

The world's estimated 287,000 maternal deaths, 3 million newborn deaths and 2.6 million third trimester stillbirths each year represent a huge burden that affects both families and communities. Recent increases in global attention and resource allocation for postnatal

maternal and newborn care require concomitant increases in the availability of programmatic indicators and data to track change over time [34]. Progress has been made over the last five years in measurement of coverage for postnatal care and for immediate care behaviors and practices for newborns, and the TWG has provided a forum for standardizing indicators and developing common tools. However, considerable work is still needed to develop and use metrics to track progress toward improving newborn survival and care.

## Acknowledgments

The authors acknowledge the participation and input of all members of the TWG. The authors would also like to acknowledge Ladys Ortiz, who conducted the data analyses of the DHS data.

## Author Contributions

Conducted qualitative study in Bangladesh and Malawi: PSY. Conducted qualitative study in Ghana: ZH. Presented DHS analyses at TWG meetings: JF. Key member of discussions at TWG meetings and presentation of UNICEF data and recommendations: HN. Initiator of TWG and key technical input into newborn indicators: JL. Analyzed the data: KK DS TG . Wrote the first draft of the manuscript: AM KK DS TG. Contributed to the writing of the manuscript: CM HN JF JL PSY ZH. ICMJE criteria for authorship read and met: AM KK DS TG CM HN JF PSY ZH JL. Agree with manuscript results and conclusions: AM KK DS TG CM HN JF PSY ZH JL.

## References

1. United Nations Children's Fund, World Health Organization, The World Bank, United Nations (2012) Levels and trends in child mortality: report 2012—Estimates developed by the UN Inter-agency Group for Child Mortality Estimation. New York: United Nations Children's Fund. Available: http://www.unicef.org.uk/Documents/Publications/UNICEF_2012_IGME_child_mortality_report.pdf. Accessed 21 January 2013.
2. Smith SL, Neupane S (2010) Factors in health initiative success: Learning from Nepal's newborn survival initiative. Soc Sci Med 72: 568–575.
3. Shiffman J (2010) Issue attention in global health: The case of newborn survival. Lancet 375: 2045–2049.
4. Lawn JE, Kinney MV, Black RE, Pitt C, Cousens C, et al. (2012) Newborn survival: A multi-country analysis of a decade of change. Health Policy Plan (Suppl 3): 1–23.
5. Hancioglu A, Arnold F (2013) Measuring coverage in MNCH: Tracking progress in health for women and children using DHS and MICS household surveys. PLoS Med 10: e1001391. doi:10.1371/journal.pmed.1001391
6. Lawn JE, Lee AC, Kinney M, Sibley L, Carlo WA, et al. (2009) Two million intrapartum-related stillbirths and neonatal deaths: Where, why, and what can be done? Int J Gynaecol Obstet 107 (Suppl 1): S5–S19.
7. World Health Organization. Pregnancy, Childbirth, Postpartum and Newborn Care: A guide for essential practice. Geneva: World Health Organization; 2003. Available: http://www.who.int/maternal_child_adolescent/documents/924159084x/en/index.html. Accessed 21 January 2013.
8. Baqui AH, Ahmed S, El Arifeen S, Darmstadt GL, Rosecrans AM, et al. (2009) Effect of timing of first postnatal care home visit on neonatal mortality in Bangladesh: A observational cohort study. BMJ 339: b2826. doi: 10.1136/bmj.b2826
9. Baqui AH, Arifeen SE, Williams EK, Ahmed S, Mannan I, et al. (2009) Effectiveness of home-based management of newborn infections by community health workers in rural Bangladesh. Pediatr Infect Dis J 28: 304–10.
10. World Health Organization, United Nations Children's Fund (2009) Home visits for the newborn child: A strategy to improve survival. Geneva: World Health Organization. Available: http://www.who.int/maternal_child_adolescent/documents/who_fch_cah_09_02/en/index.html. Accessed 21 January 2013.
11. Kerber KJ, de Graft-Johnson JE, Bhutta ZA, Okong P, Starrs A, et al. (2007) Continuum of care for maternal, newborn, and child health: From slogan to service delivery. Lancet 370: 1358–1369.
12. Countdown to 2015 (2005) Tracking progress in child survival: the 2005 report. New York: United Nations Children's Fund. Available: http://www.countdown2015mnch.org/reports-and-articles/previous-reports/2005. Accessed 21 January 2013.
13. Bryce J, Requejo J, Countdown to 2015 (2012) Building a future for women and children: The 2012 report. Geneva: World Health Organization. Available: http://www.countdown2015mnch.org/reports-and-articles/2012-report. Accessed 21 January 2013.
14. World Health Organization (2012) Accountability commission for health of women and children. Available: http://www.who.int/topics/millennium_development_goals/accountability_commission/en/. Accessed 21 January 2013.
15. UNICEF. Statistics and monitoring: Multiple Indicator Cluster Surveys. Available: http://www.unicef.org/statistics/index_24302.html. Accessed 1 February 2013.
16. Requejo JH, Newby H, Bryce J (2013) Measuring coverage in MCNH: Challenges and opportunities in the selection of coverage indicators for global monitoring. PLoS Med 10: e1001416. doi:10.1371/journal.pmed.1001416
17. Yoder P, Rosato M, Mahmud R, Fort A, Rahman F, et al. (2010) Women's recall of delivery and neonatal care in Bangladesh and Malawi: a study of terms, concepts, and survey questions. Calverton (Maryland): ICF Macro. Available http://www.measuredhs.com/pubs/pdf/QRS17/QRS17.pdf. Accessed 22 January 2013.
18. MEASURE DHS Demographic and Health Surveys. Available: http://www.measuredhs.com/. Accessed 21 January 2013.
19. Darmstadt GL, Bhutta ZA, Cousens S, Adam T, Walker N, et al. (2005) Evidence-based, cost-effective interventions: how many newborn babies can we save? Lancet 365: 977–988.
20. Hill Z, Tawiah-Agyemang C, Okeyere E, Manu A, Fenty J, et al. (2010) Improving hygiene in home deliveries in rural Ghana: how to build on current attitudes and practices. Pediatr Infect Dis J 29: 1004–1008.
21. Hill Z, Tawiah-Agyemang C, Manu A, Okyere E, Kirkwood BR (2010) Keeping newborns warm: beliefs, practices and potential for behaviour change in rural Ghana. Trop Med Int Health 15: 1118–1124.
22. Stanton CK, Rawlins B, Drake M, Anjos M, Cantor D, et al. (2013) Measuring coverage in MNCH: Testing the validity of women's self-report of key maternal and newborn health Interventions during the peripartum period in Mozambique. PLoS ONE 8: e60694. doi:10.1371/journal.pone.0060694
23. Capurro H (2004) Topical umbilical cord care at birth. The WHO Reproductive Health Library. Geneva: World Health Organization. Available: http://apps.who.int/rhl/newborn/cd001057_capurroh_com/en/index.html. Accessed 22 January 2013.
24. Mullany LC, Darmstadt GL, Khatry SK, Katz J, LeClerq SC, et al. (2006) Topical applications of chlorhexidine to the umbilical cord for prevention of omphalitis and neonatal mortality in southern Nepal: a community-based, cluster-randomised trial. Lancet 376: 910–918.
25. Soofi S, Cousens S, Imdad A, Bhutto N, Ali N, et al. (2012) Topical application of chlorhexidine to neonatal umbilical cords for prevention of omphalitis and neonatal mortality in a rural district of Pakistan: a community-based, cluster-randomised trial. Lancet 397: 1029–1036.
26. El Ariefeen S, Mullany LC, Shah R, Mannan I, Rahman SM, et al. (2012) The effect of cord cleansing with chlorhexidine on neonatal mortality in rural Bangladesh: a community-based, cluster-randomised trial. Lancet 379: 1022–1028.
27. Lawn JE, Mwansa-Kambafwile J, Horta BL, Barros FE, Cousens S (2010) 'Kangaroo mother care' to prevent neonatal deaths due to preterm birth complications. Int J Epidemiol 39: i144–i154.
28. Standley J (2011) Kangaroo mother care implementation guide. Washington (District of Columbia): Maternal and Child Health Integrated Program. Available: http://www.mchip.net/sites/default/files/MCHIP%20KMC%20Guide.pdf. Accessed 22 January 2013.
29. Campbell H, el Arifeen S, Hazir H, O'Kelly J, Bryce J, et al. (2013) Measuring coverage in MNCH: Challenges in monitoring the proportion of young children with pneumonia who receive antibiotic treatment. PLoS Med 10: e1001421. doi:10.1371/journal.pmed.1001421
30. Helping Babies Breathe (2013) Welcome to the website for Helping Babies Breathe. Available: http://www.helpingbabiesbreathe.org/. Accessed 22 January 2013.
31. Lawn JE, Gravett MG, Nunes TM, Rubens CE, Stanton C (2010) Global report on preterm birth and stillbirth (1 of 7): definitions, description of the burden and opportunities to improve data. BMC Pregnancy Childbirth 10 (Suppl 1): S1.
32. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller A, et al. (2012) National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet 379: 2162–2172.
33. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, et al. (2012) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380: 2224–2260.
34. Pitt C, Lawn JE, Ranganathan M, Mills A, Hanson K (2012) Donor funding for newborn survival: An analysis of donor-reported data, 2002–2010. PLoS Med 9: e1001332. doi:10.1371/journal.pmed.1001332

PLOS | MEDICINE

# Measuring Coverage in MNCH: A Prospective Validation Study in Pakistan and Bangladesh on Measuring Correct Treatment of Childhood Pneumonia

**Tabish Hazir[1]*[9], Khadija Begum[2][9], Shams el Arifeen[2], Amira M. Khan[1], M. Hamidul Huque[2], Narjis Kazmi[1], Sushmita Roy[2], Saleem Abbasi[1], Qazi Sadeq-ur Rahman[2], Evropi Theodoratou[3], Mahmuda Shayema Khorshed[2], Kazi Mizanur Rahman[2], Sanwarul Bari[2], M. Mahfuzul Islam Kaiser[2], Samir K. Saha[4], A. S. M. Nawshad Uddin Ahmed[4], Igor Rudan[3], Jennifer Bryce[5], Shamim Ahmad Qazi[6], Harry Campbell[3]**

**1** Children's Hospital, Pakistan Institute of Medical Sciences, Islamabad, Pakistan, **2** International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh, **3** Centre for Population Health Sciences, University of Edinburgh, Edinburgh, United Kingdom, **4** Dhaka Shishu (Children) Hospital, Dhaka, Bangladesh, **5** Department of International Health, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States of America, **6** Department of Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland

## Abstract

***Background:*** Antibiotic treatment for pneumonia as measured by Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) is a key indicator for tracking progress in achieving Millennium Development Goal 4. Concerns about the validity of this indicator led us to perform an evaluation in urban and rural settings in Pakistan and Bangladesh.

***Methods and Findings:*** Caregivers of 950 children under 5 y with pneumonia and 980 with "no pneumonia" were identified in urban and rural settings and allocated for DHS/MICS questions 2 or 4 wk later. Study physicians assigned a diagnosis of pneumonia as reference standard; the predictive ability of DHS/MICS questions and additional measurement tools to identify pneumonia versus non-pneumonia cases was evaluated. Results at both sites showed suboptimal discriminative power, with no difference between 2- or 4-wk recall. Individual patterns of sensitivity and specificity varied substantially across study sites (sensitivity 66.9% and 45.5%, and specificity 68.8% and 69.5%, for DHS in Pakistan and Bangladesh, respectively). Prescribed antibiotics for pneumonia were correctly recalled by about two-thirds of caregivers using DHS questions, increasing to 72% and 82% in Pakistan and Bangladesh, respectively, using a drug chart and detailed enquiry.

***Conclusions:*** Monitoring antibiotic treatment of pneumonia is essential for national and global programs. Current (DHS/MICS questions) and proposed new (video and pneumonia score) methods of identifying pneumonia based on maternal recall discriminate poorly between pneumonia and children with cough. Furthermore, these methods have a low yield to identify children who have true pneumonia. Reported antibiotic treatment rates among these children are therefore not a valid proxy indicator of pneumonia treatment rates. These results have important implications for program monitoring and suggest that data in its current format from DHS/MICS surveys should not be used for the purpose of monitoring antibiotic treatment rates in children with pneumonia at the present time.

*Please see later in the article for the Editors' Summary*.

**Citation:** Hazir T, Begum K, el Arifeen S, Khan AM, Huque MH, et al. (2013) Measuring Coverage in MNCH: A Prospective Validation Study in Pakistan and Bangladesh on Measuring Correct Treatment of Childhood Pneumonia. PLoS Med 10(5): e1001422. doi:10.1371/journal.pmed.1001422

**Academic Editor:** David Osrin, Wellcome Trust Senior Research Fellow in Clinical Science, UCL Reader in International Child Health, Honorary Consultant, Great Ormond Street Hospital for Children, United Kingdom

**Received** September 18, 2012; **Accepted** March 4, 2013; **Published** May 7, 2013



**Funding:** This work was supported by a sub-grant "MA13 project: careseeking and coverage of child health interventions" from the Child Health Epidemiology Reference Group (CHERG; www.cherg.org) grant from The Bill & Melinda Gates Foundation grant no. 50140 (www.gatesfoundation.org). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** DHS, Demographic and Health Survey/s; MICS, Multiple Indicator Cluster Survey/s; WHO, World Health Organization.

* E-mail: tabishhazir@hotmail.com

These authors contributed equally to this work.

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection.*

## Introduction

Globally, of the estimated 6.9 million annual deaths in children younger than 5 y, 1.2 million (18%) die from pneumonia [1]. Prompt treatment with appropriate antibiotics in children with pneumonia is an effective intervention for reducing mortality [2,3]. The proportion of children with pneumonia in a population who receive antibiotic treatment (antibiotic treatment rate) is a key indicator for tracking progress in achieving Millennium Development Goal 4 targets [4]. The validity of this indicator depends on both the correct identification of pneumonia and the use of an antibiotic to treat the condition. Current measures of antibiotic use in pneumonia rely on household-based surveys such as the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS). It is, therefore, crucial that the DHS and MICS measures are reliable and accurate.

Pneumonia indicators in current surveys are based on interviews with mothers (DHS) or primary caregivers (MICS) using structured questions on cough and short, rapid breathing or difficulty in breathing in the previous 2 wk and whether these were chest-related. The intent is to identify the best possible proxy for pneumonia in order to assess treatment coverage based on mother/caregiver recall. Although the DHS survey labels this condition as "symptoms of acute respiratory infection," the MICS survey uses the term "suspected pneumonia." Irrespective of the terminology used, for program purposes these cases are classified as pneumonia. Therefore, the accuracy and reliability of these questions and the algorithm as a valid proxy of pneumonia in children has been questioned [5,6], and there is a need to assess the validity of this approach.

Our primary study objective was to assess the validity of the caregiver's responses to the standard DHS/MICS questions about whether the child had pneumonia in the recent past, and if so, how it was treated. The reference standard was physician-diagnosed pneumonia (as per World Health Organization [WHO] definitions) [7]. A further objective was to determine if these measures can be improved by adding additional questions to the DHS/MICS surveys or using alternative measurement tools. We also aimed to assess if there was any difference in a caregiver's recall at 2 wk (recall period in current DHS/MICS surveys) and 4 wk.

## Methods

### Ethics Statement

Ethical approval was obtained from the Hospital Ethics Committee (Pakistan Institute of Medical Sciences), the Ethical Review Committee of the International Centre for Diarrhoeal Disease Research, Bangladesh, and the WHO Ethics Review Committee.

We employed a two-stage written consent procedure: caregivers were informed about the study and permission was obtained first at the time of diagnosis and enrollment, and then again at the start of the follow-up home interviews.

### Setting

DHS and MICS surveys are conducted very widely throughout low- and middle-income countries and in very diverse settings, which range from rural settings served by community services to urban and peri-urban settings, where the first level health services are provided by local hospital outpatient departments. We thus conducted the study in three diverse urban and rural settings.

The study was conducted in two countries: Pakistan and Bangladesh. In Pakistan, study participants were recruited in an urban setting, from the outpatient department of Children's Hospital, Pakistan Institute of Medical Sciences, Islamabad, Pakistan. In Bangladesh, the participants were recruited from both urban (Dhaka Shishu Hospital, a tertiary pediatric hospital) and rural (community-based recruitment in four unions [Mirzapur, Gorai, Bhatgram, and Jamurki] of Mirzapur, 63 km north of Dhaka) settings. The latter setting involved initial identification of pneumonia and "no pneumonia" cases during weekly home visits by Village Health Workers, later reassessed by study physicians according to WHO guidelines and enrolled.

### Study Design

The study identified and recruited two groups of children with acute respiratory infections: those who were confirmed to have pneumonia and those who did not have pneumonia. Two to four weeks after recruitment caregivers were surveyed to assess the accuracy of their recall of the diagnosis and treatment given. This enabled an assessment to be made of the degree to which DHS and MICS measures of antibiotic treatment in those with reported symptoms of pneumonia were valid measures of the antibiotic treatment of true pneumonia in a study population.

The study had two phases. Phase 1 involved recruiting large numbers of children with symptoms of acute respiratory infection and having trained study physicians establish whether pneumonia (the reference standard for the test) was present or absent. We purposefully selected hospital outpatient departments as two of the three study sites to increase the probability that we would find sufficient numbers of children presenting with symptoms of acute respiratory infection, and to allow careful training and monitoring of the study physicians to ensure that the reference standard was robust. At the third, rural site in Bangladesh, patients were also identified in the community. Phase 2 involved having trained field workers interview the caregiver of each child in their home using DHS and MICS algorithms and alternative tools, either 2 or 4 wk after their recruitment.

### Participants and Selection Criteria

The study participants were children 0–59 mo with physician-diagnosed clinical pneumonia or "no pneumonia" and their caregivers. A single WHO technical expert (S. A. Q.) oversaw staff training and study monitoring to ensure comparability of case definitions across sites.

### Enrollment and Follow-Up

**Enrollment procedure.** In the two urban settings (Islamabad and Dhaka) all children aged 0 to 5 y who were assessed and managed by hospital outpatient department physicians were referred to study physicians, who then screened, reassessed, and diagnosed the children using WHO acute respiratory infection guidelines and determined their eligibility for inclusion in the study. Children were enrolled in one of two groups: those with pneumonia and those with "no pneumonia" [7]. Treatment provided by outpatient department physicians was recorded. Children with the following conditions were excluded: recurrent wheezing (airway disorder/asthma), severe pneumonia requiring hospitalization (since DHS and MICS surveys typically gather data on very few of these cases due to their very low period prevalence in community surveys and since hospitalization, being

a dramatic event, could induce recall bias), symptoms of chronic cough of more than 4 wk duration, history of pneumonia within past 10 d, history of congenital heart disease, and nonresident of study catchment area. Children with "cough or cold/no pneumonia" were frequency matched to pneumonia cases so that in any one week the age (and sex, in the case of the Bangladesh sites) distribution of cases of pneumonia and "cough or cold" was similar.

For each pneumonia case enrolled, a corresponding age-matched (and sex-matched, in case of Bangladesh sites) "no pneumonia" case was selected using a computer-generated randomization list. Details of the recruitment procedures for each site are given in Box 2.

In the rural setting (Mirzapur), the Village Health Workers identified all children with pneumonia or "no pneumonia" through their weekly household visits and informed the study physicians by mobile phone. The study physicians visited at home all children with "possible pneumonia" and a sample of children with "no pneumonia" who lived nearest to the case of "possible pneumonia." They reassessed using Integrated Management of Childhood Illness guidelines, and screened and enrolled eligible children.

**DHS and MICS questionnaires and alternative measurement tools.** DHS and MICS surveys both have an algorithm of questions about the presence or absence of specific signs and symptoms of suspected pneumonia (denoted as "symptoms of acute respiratory infection" in the DHS survey and "suspected pneumonia" in the MICS questionnaire; copies of questionnaires are given in Texts S1, S2, S3). Additional study tools were developed by ARI Research Cell, Children's Hospital, Pakistan Institute of Medical Sciences, Islamabad, and field tested in both Pakistan and Bangladesh. These tools included a pneumonia score questionnaire, which consisted of questions on 20 commonly reported signs and symptoms of pneumonia (Box 3), and a video depicting children in three scenarios: with signs and symptoms of pneumonia, with severe pneumonia, and with cough and cold but without pneumonia. The video had nine different clips of no pneumonia, pneumonia (fast breathing), and severe pneumonia (lower chest indrawing), one set for each of three age groups, i.e., up to 2 mo, 2–11 mo, and 12–59 mo. Country-specific drug charts (flip chart and computer-based) were developed showing medicines (especially antibiotics) commonly used in pneumonia and other febrile illnesses in children 0–59 mo old. All questionnaires, including standard DHS and MICS questionnaires, were translated into local languages (Urdu and Bangla—see Texts S1, S2, S3 for copies of questionnaires).

**Home follow-up.** Using computer-generated randomization lists, two-thirds of enrolled children were randomly allocated to follow-up at 2 wk, and one-third to follow-up at 4 wk. The mothers/caretakers were (i) interviewed using DHS/MICS questions on cough and chest-related short, rapid breathing and difficult breathing, (ii) asked about specific signs of pneumonia using a pneumonia score questionnaire, (iii) shown video clips to find which clip best represented their child's respiratory illness episode, and (iv) shown a "computer-based drug chart" and "a drug flip chart" to identify the drugs the child was treated with. The order of administering the DHS and MICS questionnaires was alternated with each child, so that each questionnaire was administered first on 50% of occasions.

### Box 1. Key DHS and MICS Acute Respiratory Infection Questions Used for the Identification of Children with a Caregiver Report of Symptoms and Signs of Pneumonia

Q.535 (PDHS)/Q.B3 (BDHS)/Q.CA8 (MICS-P)/Q.C2 (MICS-B): When (name) had an illness with a cough, did s/he breathe faster than usual with short, rapid breaths or have difficulty in breathing?
Q.536 (PDHS)/Q.B4 (BDHS)/Q.CA9 (MICS-P)/Q.C3 (MICS-B): Was the fast or difficult breathing due to a problem in the chest or to a blocked or runny nose?
BDHS, Bangladesh DHS; MICS-B, MICS Bangladesh; MICS-P, MICS Pakistan; PDHS, Pakistan DHS.

### Box 2. Details of the Matching Procedures at the Study Sites

**Pakistan** For each pneumonia case enrolled a corresponding age-matched "no pneumonia" case was randomly selected. Randomization was done with a computer-generated randomization list with blocks of four (each having a randomly selected number). For every enrolled pneumonia case, a "no pneumonia" case was picked as per the sequence number of the corresponding block. There were two separate randomization lists: one for children 0–12 mo and one for those >12–59 mo (to match for age). Time matching was also ensured, whereby in any one week a similar number of age-matched "no pneumonia" and pneumonia cases were enrolled.

**Bangladesh** The enrollment and randomization procedure followed a hierarchical procedure where an attempt was made to match each pneumonia case in a given week one-on-one with a non-pneumonia case based on age, sex, and study physician. The procedure for matching, in order of preference, was as follows:

1. Matched on case's sex and age (±2 mo), and assessed by the same study physician
2. Matched on case's sex and age (±2 mo), and assessed by different study physician
3. Matched on case's sex and same age category (≤12 mo or >12 mo), and assessed by the same study physician
4. If criteria 1–3 could not be met, we matched on case's sex and same age category under different study physician

If more than one child with no pneumonia was found for a case (pneumonia), we randomly selected only one control using computer-generated random numbers. A child was selected as a control only once. This procedure was followed at the end of each week throughout the enrollment period. An additional criterion was added after 2 mo of study enrollment to ensure adequate enrollment of controls: if matched controls were not found from the children with "no pneumonia" assessed in the same week as the cases, controls were selected from children with "no pneumonia" assessed in the week prior to or following the week of the case. Eligible and consented pneumonia patients were excluded from the study if no age- and sex-matched "no pneumonia" patient could be found.

## Data Management and Analysis

After double data entry and data cleaning, descriptive statistics were used to assess the socio-demographic and clinical characteristics of the study children. Sensitivity and specificity were calculated to assess the discriminative power of each measurement

**Box 3. Pneumonia Score Features**

1. Cough
2. Fever
3. Chills/sweats
4. Restlessness
5. Irritability
6. Loss of appetite
7. Abnormally sleepy
8. Wheezing
9. Shortness of breath
10. Fast breathing
11. Flaring of nostrils
12. Refusal to drink
13. Lower chest indrawing
14. Chest pain
15. Difficulty in breathing
16. Vomiting
17. Grunting
18. Blue coloration of skin
19. Coughing up blood
20. Convulsions

tool (DHS, MICS, pneumonia score questionnaire, pneumonia video) and for various combinations of tools.

For the pneumonia score, a composite score was developed by adding one point for each symptom or sign present. Sensitivity and specificity were calculated for each of the composite scores (ranging from lowest to highest). The intent was to identify a cutoff that can be used to reach a decision on the presence or absence of pneumonia. SPSS 11.0 (SPSS Inc.) and STATA 10 (StataCorp) were used in Pakistan and Bangladesh, respectively.

### Sample Size Calculation

Based on baseline estimates of sensitivity of 60%–70% and specificity of 70%–90% of mother/caregiver recall of symptoms of acute respiratory infection or suspected pneumonia in predicting true pneumonia, it was estimated that 300 children under 5 y with physician-diagnosed pneumonia and 300 with "no pneumonia" should be enrolled at each site in order to estimate sensitivity and specificity with a precision of ±5%.

The study is reported in line with the STARD statement (checklist in Text S4). The protocol is provided in Text S5.

## Results

### Enrollment and Follow-Up Status

At the Pakistan site, 752 children were enrolled in the study—361 with pneumonia and 391 with "no pneumonia"—between October 2010 and February 2011 (Figure 1). Follow-up of 329 pneumonia and 343 "no pneumonia" cases was successfully carried out; 456 at 2 wk and 216 at 4 wk.

In Bangladesh, between March and August 2011, 1,178 children—589 each in the pneumonia and "no pneumonia" groups—were enrolled, of which 700 were from Dhaka Shishu Hospital (Figure 2) and 478 from rural Mirzapur (Figure 3). "No pneumonia" cases were enrolled using recruitment procedures similar to those used in Pakistan.

### Baseline Characteristics of the Study Children

The socio-demographic characteristics and the clinical features of the study children are presented in Table 1.

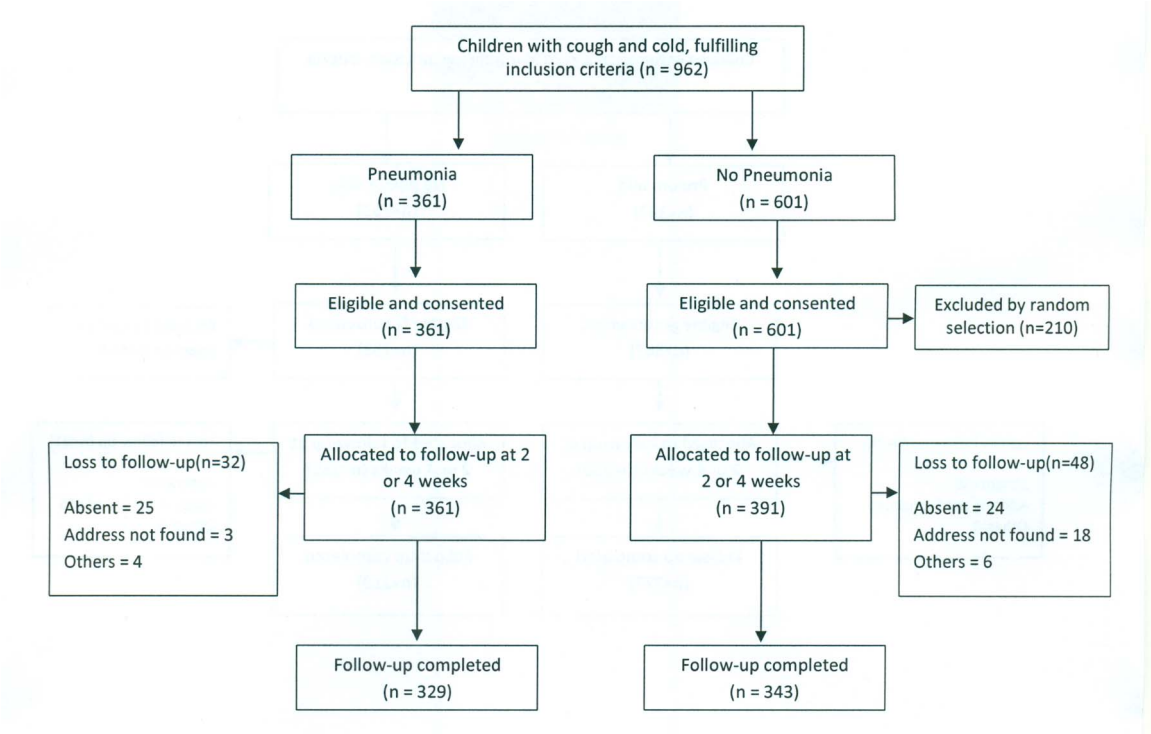


**Figure 1. Pakistan urban site—flowchart of selection of pneumonia and no-pneumonia cases.**
doi:10.1371/journal.pmed.1001422.g001

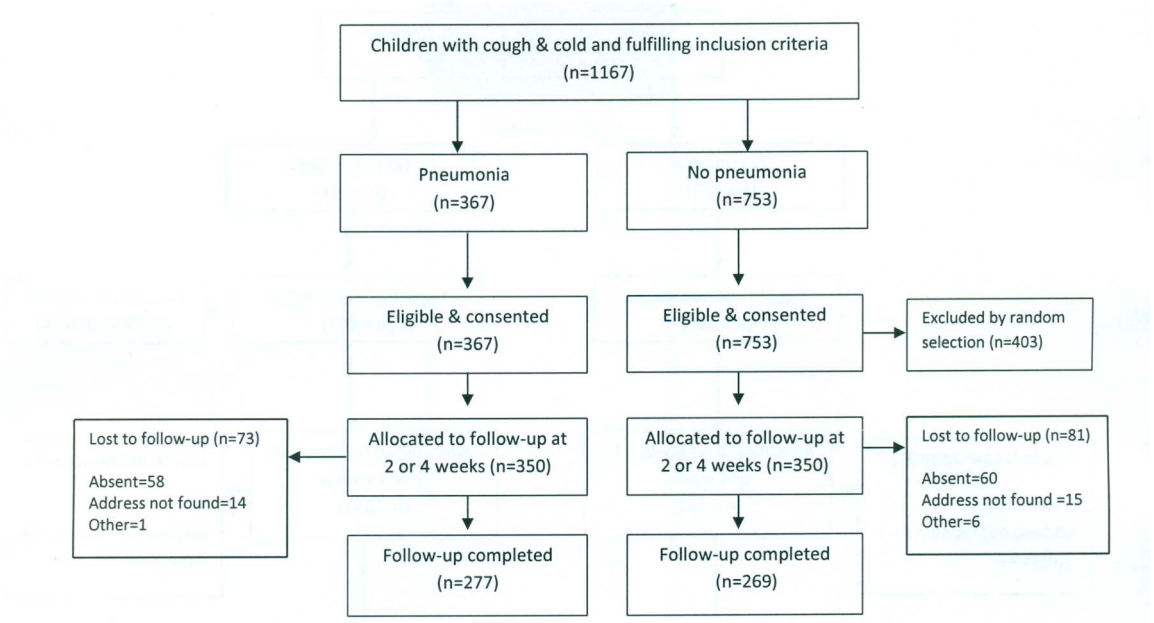


**Figure 2. Bangladesh urban site—flowchart of selection of pneumonia and no-pneumonia cases in Dhaka Shishu Hospital, Bangladesh.**
doi:10.1371/journal.pmed.1001422.g002

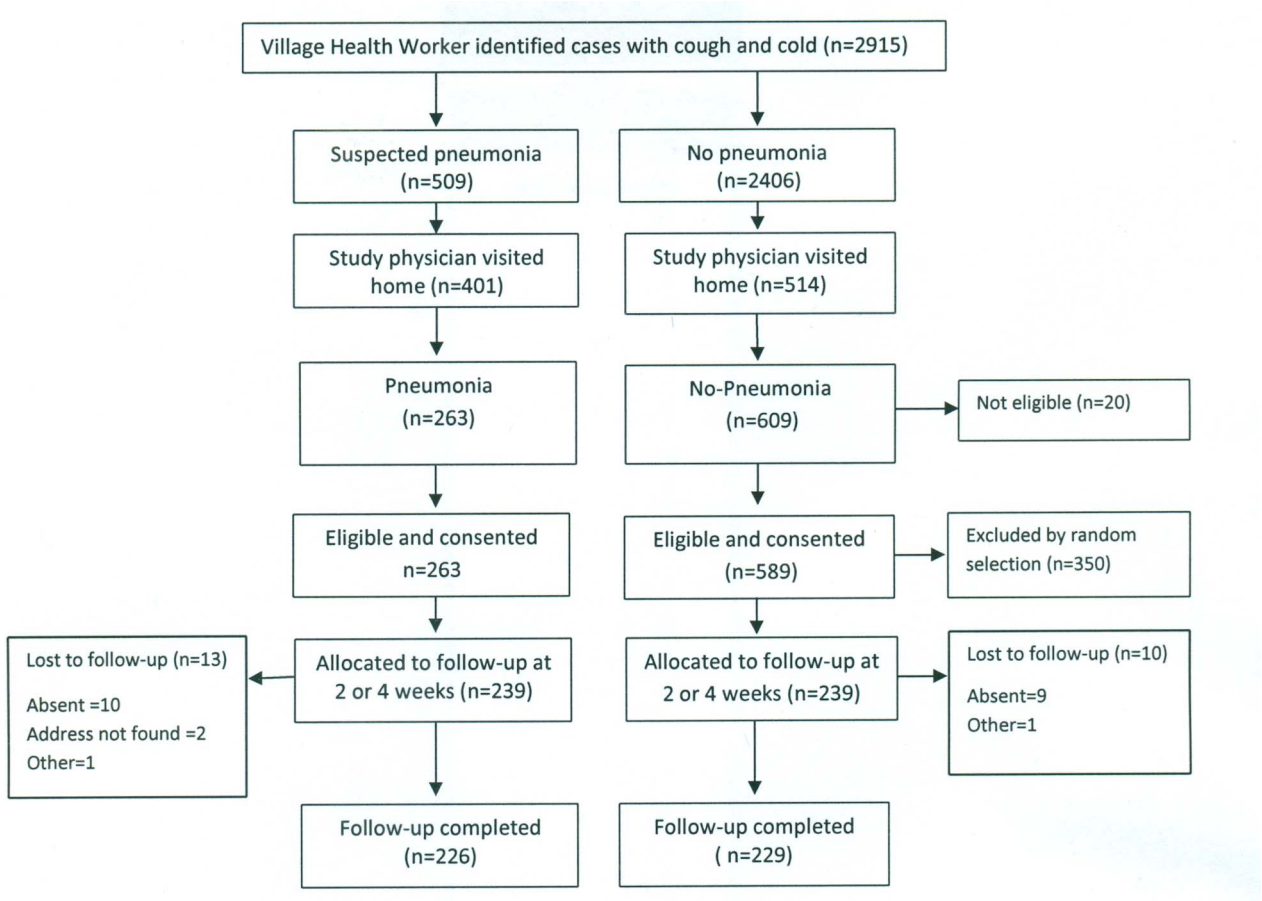


**Figure 3. Bangladesh rural site—flowchart of selection of pneumonia and no-pneumonia cases in the rural setting.**
doi:10.1371/journal.pmed.1001422.g003

### Discriminative Power of Survey Tools

Sensitivity, specificity, and 95% CIs for the various instruments that were studied are presented in Tables 2 and 3.

Data are given as percent (95% CI).

**DHS/MICS questions.** Results at both sites showed poor discriminative power for DHS and MICS questions. Individual patterns of sensitivity and specificity varied substantially by study site and between urban and rural sites (Table 2). Results were similar for DHS and MICS questions, although not identical, because of some differences in questionnaire skip patterns. There was no significant difference in the validity of DHS and MICS questions at 2- versus 4-wk follow-up intervals (Table 2).

**Pneumonia score.** As expected, specificity increased and sensitivity declined with increasing pneumonia score, i.e., the number of positive symptoms from among the pneumonia score questions. An area-under-the-curve analysis was performed for the pneumonia score (see below) and is presented in Figure 4. Thresholds could be chosen so that the performance was comparable to DHS/MICS questions (e.g., >9 in Bangladesh) or to selectively increase test specificity.

**Video tool.** In Pakistan (where the video was developed and child images recorded), the video tool had a much higher discriminative power than in Bangladesh. There was no significant difference in recall with the video tool at 2 versus 4 wk in any of the three sites (Table 2).

### Recall of Antibiotic Use

In children with pneumonia, antibiotic treatment was recalled correctly by 66.0% and 66.8% of caregivers using DHS or MICS questions in Pakistan and Bangladesh, respectively, with no significant difference between 2- and 4-wk recall. Correct recall increased to 72.0% using a drug flip chart/photo album and to 71.3% using a computer-based drug chart in Pakistan, and to 78.8% and 81.2%, respectively, in Bangladesh.

### Association of Caregiver's Recall with Socio-Demographic Characteristics

In both Bangladesh study sites, but not in Pakistan, test specificities were higher for mothers who had higher levels of education. At the Bangladesh rural site the test specificity was 59% for mothers who had secondary or higher education, compared to a specificity of 37% for mothers whose education level was below primary. However, at the Bangladesh urban site these figures were 86% and 73%, respectively.

## Discussion

### Importance of Monitoring Antibiotic Treatment of Pneumonia

Despite recent falls in pneumonia mortality over the past ten years, it remains the largest single cause of death in children [8]. It is therefore essential that global and national programs track the coverage of effective interventions against pneumonia if Millennium Development Goal 4 is to be achieved. The antibiotic treatment of pneumonia is highly effective and, together with immunization, is one of the two main control strategies [9]. DHS and MICS household surveys are the primary tools used to measure intervention coverage in low- and middle-income countries where health information systems are weak, and are the primary source of information on common childhood illnesses and treatment coverage. We report here on the first study evaluating the validity of the estimates of pneumonia treatment coverage produced by the DHS and MICS survey instruments and discuss the implications of these findings for program monitoring.

### Validity of DHS/MICS Data for Estimation of Pneumonia Prevalence

Community-based epidemiological studies have estimated pneumonia incidence to be about 0.3 episodes per child per year in children under 5 y in low- and middle-income countries, which is some 12–18 times lower than the reported incidence of upper respiratory infections in this age group [10]. This low prevalence of pneumonia among all children with cough who are surveyed by DHS and MICS requires that the survey instruments must have very high specificity to identify pneumonia, otherwise the great majority of cases identified in the survey will not represent true cases of pneumonia (as reviewed by Campbell and colleagues in this *PLOS Medicine* Collection [11]). An example of this effect can be seen in a recent Pakistan DHS survey, which reported in 2006–2007 that 2,508/8,367 (29.9%) children under 5 y had cough and 1,178/8,367 (14.1%) had symptoms consistent with pneumonia [12]. This ratio of children with reported symptoms and signs of "cough and cold only" versus "suspected pneumonia" of 2:1 is in marked contrast to that reported in community studies, where this ratio is typically very much higher, as noted above [13].

DHS and MICS reports clearly caution that these data (denoted as "suspected pneumonia" in MICS surveys) should not be used as a proxy measure of the prevalence of pneumonia in the community. However, many pneumonia control programs in the developing world, faced with a lack of data on this important parameter for planning purposes, use the information in this way. The results of our study show that the specificity of these survey tools is well below the very high levels required for this proxy to give an accurate estimation of pneumonia prevalence, and reinforce that these data should not be used for this purpose as they will lead to large overestimations of pneumonia prevalence [11].

### Validity of DHS/MICS Data for Estimation of Proportion of Children with Pneumonia Who Receive Antibiotic Treatment

DHS and MICS surveys identify children whose caregivers report that they had symptoms and signs consistent with pneumonia, and then ask whether these children were treated with an antibiotic. The validity of this proxy indicator of the proportion of children with pneumonia who receive antibiotic treatment is therefore entirely dependent on the validity of DHS/MICS-reported symptoms and signs of pneumonia as a measure of true pneumonia. For this to represent a valid denominator for this important indicator, it is important that a high proportion of children with "symptoms of acute respiratory infection" (DHS)/"suspected pneumonia" (MICS) actually have "true" pneumonia. If this proportion is low, then results of monitoring antibiotic coverage among these children will not only be inaccurate but also misleading. The results of this study show that DHS and MICS surveys at both sites have poor discriminative power for the identification of episodes of pneumonia in young children. Previous studies have shown that fever is a significant predictor of pneumonia and that adding fever to WHO criteria increased the specificity of pneumonia diagnosis [14,15]. Our findings support this observation, as specificity increased when a question on the presence of fever was included (Table 3). Since DHS/MICS surveys currently collect information on the presence of fever in the last 2 wk, these data could be readily incorporated into the definition of "suspected pneumonia." However, the addition of

**Table 1.** Baseline characteristics of the study children.

| Characteristic | Pakistan (Urban) | | Bangladesh (Urban) | | Bangladesh (Rural) | |
|---|---|---|---|---|---|---|
| | Pneumonia (*n* = 329) | No Pneumonia (*n* = 343) | Pneumonia (*n* = 350) | No Pneumonia (*n* = 350) | Pneumonia (*n* = 239) | No Pneumonia (*n* = 239) |
| **Age of child (months) (mean ± SD)** | 10.6±10.1 | 12.9±13.2 | 12.2±12.0 | 12.8±12.8 | 16.8±12.6 | 18.9±14.6 |
| **Age category of child (months)** | | | | | | |
| 0 to 2 | 27 (8.2) | 34 (9.9) | 84 (24.0) | 88 (25.1) | 1 (0.4) | 10 (4.2) |
| 2 to 11 | 208 (63.2) | 210 (61.2) | 116 (33.1) | 118 (33.7) | 84 (35.2) | 83 (34.7) |
| 12 to 59 | 94 (28.6) | 99 (28.9) | 150 (42.9) | 144 (41.1) | 154 (64.4) | 146 (61.1) |
| **Gender** | | | | | | |
| Male | 199 (60.5) | 172 (50.1) | 204 (58.4) | 203 (58.1) | 123 (51.5) | 124 (51.9) |
| Female | 130 (39.5) | 171 (49.9) | 146 (41.6) | 147 (41.9) | 116 (48.5) | 115 (48.1) |
| **Siblings** | | | | | | |
| No siblings | 73 (22.2) | 83 (24.2) | 182 (52.0) | 166 (47.4) | 96 (40.2) | 115 (48.1) |
| One or more | 256 (77.8) | 260 (75.8) | 168 (48.0) | 184 (52.6) | 143 (59.8) | 124 (51.9) |
| **Age category of mothers (years)** | | | | | | |
| ≤30 | 266 (80.9) | 266 (77.6) | 322 (92.0) | 314 (89.7) | 206 (86.2) | 219 (91.6) |
| >30 | 63 (19.1) | 77 (22.4) | 28 (8.0) | 36 (10.3) | 33 (13.8) | 20 (8.4) |
| **Mother's education** | | | | | | |
| Illiterate | 88 (26.7) | 55 (16.0) | 81 (23.1) | 64 (18.3) | 32 (13.4) | 26 (10.9) |
| Up to 10th grade | 145 (44.1) | 157 (45.8) | 192 (54.8) | 209 (59.7) | 175 (73.2) | 176 (73.8) |
| Above 10th grade | 96 (29.2) | 131 (38.2) | 77 (22.0) | 77 (22.0) | 32 (13.4) | 38 (15.9) |
| **Father's education** | | | | | | |
| Illiterate | 42 (12.8) | 18 (5.2) | 58 (16.6) | 52 (15.7) | 47 (19.7) | 29 (12.1) |
| Up to 10th grade | 148 (45.0) | 158 (46.1) | 179 (51.1) | 153 (43.7) | 141 (59.0) | 139 (58.1) |
| Above 10th grade | 139 (42.2) | 167 (48.7) | 113 (32.3) | 142 (40.6) | 51 (21.4) | 71 (29.7) |
| **Father's occupation status** | | | | | | |
| Employed | 322 (97.9) | 322 (93.9) | 343 (98.0) | 345 (98.6) | 226 (94.6) | 231 (96.7) |
| Unemployed | 7 (2.1) | 21 (6.1) | 7 (2.0) | 5 (1.4) | 13 (5.4) | 8 (3.3) |
| **Symptoms** | | | | | | |
| Cough and cold | 327 (99.4) | 343 (100.0) | 342 (97.7) | 343 (98.0) | 228 (95.4) | 237 (99.2) |
| Fever | 296 (90.0) | 192 (56.0) | 153 (43.7) | 100 (28.6) | 77 (32.2) | 42 (17.6) |
| Breathing problem | 51 (15.5) | 3 (0.9) | 60 (17.1) | 34 (9.7) | 40 (16.7) | 8 (3.4) |
| Feeding problem | 16 (4.9) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Vomiting | 10 (3.0) | 2 (0.6) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Gastrointestinal problem | 5 (1.5) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Irritability | 5 (1.5) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Others | 2 (0.6) | 0 (0.0) | 11 (3.1) | 21 (6.1) | 12 (5.0) | 5 (2.0) |
| **Respiratory rate/minute (mean ± SD)** | | | | | | |
| 0–2 mo | 70.5±11.3 | 46.8±8.8 | 63.8±4.4 | 43.4±5.4 | 61[a] | 46.3±9.5 |
| 2–11 mo | 64.5±10.1 | 37.2±6.0 | 55.1±5.1 | 37.9±5.9 | 54.7±5.5 | 37.0±7.2 |
| 12–59 mo | 57.0±10.3 | 31.2±4.2 | 47.4±6.7 | 30.6±5.0 | 46.8±6.3 | 30.7±4.7 |
| **Temperature (°C)** | | | | | | |
| <37.5 | 236 (71.7) | 309 (90.1) | 216 (64.1) | 293 (85.4) | 201 (84.1) | 228 (95.6) |
| ≥37.5 | 93 (28.3) | 34 (9.9) | 121 (35.9) | 50 (14.6) | 38 (15.9) | 11 (4.4) |
| **Findings on auscultation** | | | | | | |
| No significant findings | 120 (36.5) | 334 (97.4) | 194 (55.4) | 320 (91.4) | 190 (79.5) | 233 (97.5) |
| Crepitations/wheeze | 209 (63.5) | 9 (2.6) | 156 (44.5) | 30 (8.5) | 49 (20.5) | 6 (2.5) |

Data are given as number (percent) unless otherwise indicated.
[a]There was only one child in the 0–2 age group at the Bangladesh rural site, hence a standard deviation value could not be calculated.
SD, standard deviation.
doi:10.1371/journal.pmed.1001422.t001

**Table 2.** Discriminative power of DHS/MICS questions about suspected pneumonia and of video for identifying childhood pneumonia (based on 2-wk and 4-wk recall).

| Recall Period/Site | Diagnostic Validity | DHS Questions | MICS Questions | Video |
|---|---|---|---|---|
| **2-wk recall period** | | | | |
| Pakistan (urban) | Sensitivity | 64.7 (58.4–70.9) | 63.8 (57.5–70.0) | 59.8 (53.3–66.2) |
| | Specificity | 68.5 (62.5–74.4) | 67.2 (61.1–73.2) | 78.0 (72.6–83.3) |
| Bangladesh (urban) | Sensitivity | 24.6 (17.5–32.9) | 25.4 (18.2–33.8) | 26.9 (19.5–35.4) |
| | Specificity | 81.7 (73.6–88.1) | 82.5 (74.5–88.8) | 82.5 (74.5–88.8) |
| Bangladesh (rural) | Sensitivity | 71.1 (61.0–79.9) | 70.1 (60.0–79.0) | 26.8 (18.3–36.8) |
| | Specificity | 56.5 (45.3–67.2) | 56.5 (45.3–67.2) | 77.6 (67.3–86.0) |
| **4-wk recall period** | | | | |
| Pakistan (urban) | Sensitivity | 71.4 (62.7–80.0) | 69.5 (60.6–78.3) | 64.8 (55.6–73.9) |
| | Specificity | 69.4 (55.4–77.9) | 67.6 (60.7–78.2) | 74.8 (66.7–82.8) |
| Bangladesh (urban) | Sensitivity | 23.2 (15.8–32.1) | 23.2 (15.8–32.1) | 28.6 (20.4–37.9) |
| | Specificity | 82.7 (74.3–89.3) | 83.6 (75.4–90.3) | 80.9 (72.3–87.8) |
| Bangladesh (rural) | Sensitivity | 72.3 (62.5–80.7) | 73.3 (63.5–81.6) | 29.7 (21.0–39.6) |
| | Specificity | 53.2 (43.4–62.7) | 53.2 (43.4–62.7) | 83.8 (75.6–90.1) |
| **Overall** | | | | |
| Pakistan (urban) | Sensitivity | 66.9 (61.8–71.9) | 65.7 (60.5–70.8) | 61.4 (56.1–66.6) |
| | Specificity | 68.8 (63.8–73.7) | 67.3 (62.3–72.2) | 77.0 (72.5–81.4) |
| Bangladesh (urban) | Sensitivity | 24.0 (18.7–29.9) | 24.4 (19.1–30.3) | 27.7 (22.1–33.8) |
| | Specificity | 82.2 (76.6–86.9) | 83.0 (77.6–87.7) | 81.7 (76.1–86.5) |
| Bangladesh (rural) | Sensitivity | 71.7 (64.9–77.9) | 71.7 (64.9–77.9) | 28.3 (22.1–35.1) |
| | Specificity | 54.6 (47.3–61.7) | 54.6 (47.3–61.7) | 81.1 (74.9–86.3) |

doi:10.1371/journal.pmed.1001422.t002

questions on the presence of fever made only modest improvements to the discriminative power. Thus, most children with "suspected pneumonia" (MICS) or "symptoms of acute respiratory infection" (DHS) who form the denominator for the "antibiotic treatment rate" do not truly have pneumonia. We conclude, therefore, that the addition of questions on the presence of these symptoms and signs would not make sufficiently large improvements in instrument validity for these to be adopted at present.

We found no significant difference in caregivers' recall at 2 and 4 wk, which is contrary to the findings in some previous studies. The recall sensitivity of caregivers' reports dropped when the interview took place after more than 2 wk in a respiratory questionnaire validation study conducted in Peru [6]. Some previous studies on maternal recall of breastfeeding duration showed a drop in mothers' recall accuracy with time [16,17]. A more recent study conducted in Kenya also assessed the accuracy of caregivers' recall over time and found that it decreased when

**Table 3.** Sensitivity and specificity of different tools (individually and in combination).

| Tools | Pakistan (Urban) | | Bangladesh (Urban) | | Bangladesh (Rural) | |
|---|---|---|---|---|---|---|
| | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| Fever[a] | 83.0 (72.9–87.3) | 31.8 (25.1–34.8) | 80.6 (75.0–85.4) | 26.5 (20.9–32.7) | 86.4 (80.8–90.8) | 23.0 (17.3–29.5) |
| LCI[a] | 42.6 (37.2–47.9) | 89.2 (85.9–92.4) | 22.7 (17.6–28.5) | 83.0 (77.6–87.7) | 41.4 (34.5–48.6) | 72.4 (65.6–78.6) |
| Chest pain[a] | 46.5 (41.1–51.8) | 77.8 (73.4–82.1) | 0.4 (0.0–2.3) | 97.0 (93.8–98.8) | 3.0 (1.12–6.5) | 97.4 (94.1–99.2) |
| DHS+fever | 61.2 (56.1–66.6) | 72.7 (68.2–77.6) | 19.4 (14.6–25.0) | 85.7 (80.4–89.9) | 62.6 (55.5–69.4) | 63.8 (56.6–70.5) |
| DHS+LCI | 33.6 (28.6–38.8) | 93.3 (90.6–95.9) | 11.6 (7.8–16.3) | 90.9 (86.4–94.3) | 34.8 (28.2–41.9) | 78.6 (72.2–84.1) |
| DHS+video | 44.7 (39.3–50.0) | 88.3 (84.9–91.7) | 14.5 (10.3–19.5) | 92.2 (87.9–95.3) | 25.8 (19.8–85.2) | 85.2 (79.4–89.9) |
| DHS+chest pain | 39.5 (34.2–44.7) | 86.6 (83.0–90.2) | 0.4 (0.0–2.8) | 99.6 (97.6–100.0) | 3.0 (1.1–6.5) | 98.5 (95.6–99.7) |
| DHS+fever+video | 42.6 (36.8–48.3) | 89.2 (85.9–92.4) | 10.7 (7.1–15.3) | 93.9 (90.0–96.6) | 21.7 (16.2–28.1) | 86.7 (81.2–91.1) |
| DHS+fever+LCI | 24.9 (20.2–29.5) | 95.6 (93.4–97.7) | 9.1 (5.8–13.4) | 93.5 (89.5–96.3) | 30.8 (24.5–37.7) | 84.7 (78.9–89.4) |

Data are given as percent (95% CI). Results for both recall periods combined.
[a]These variables from the pneumonia score were selected for this analysis because they had the strongest association with pneumonia in screening univariate analyses.
LCI, lower chest indrawing.
doi:10.1371/journal.pmed.1001422.t003

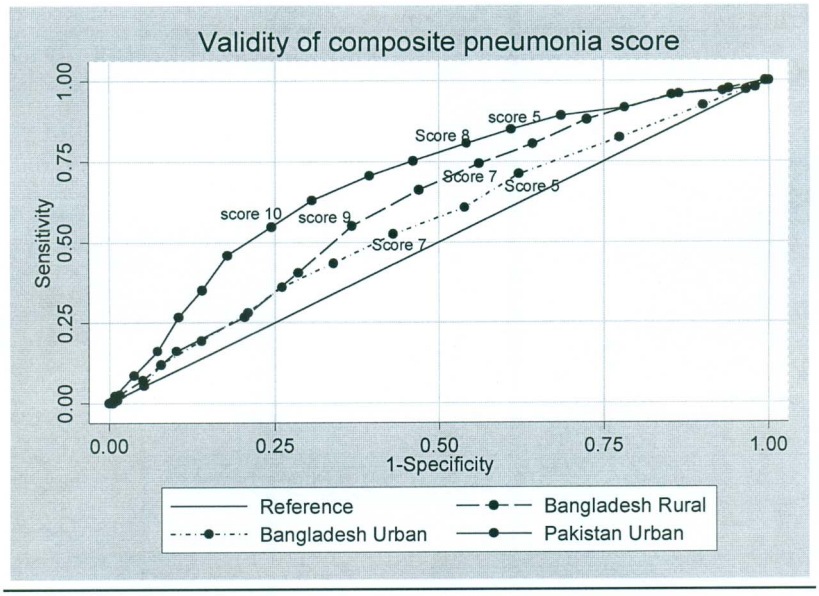


**Figure 4. Validity of composite pneumonia score.**
doi:10.1371/journal.pmed.1001422.g004

the recall period was increased to 2 wk from 3–4 d [18]. In our study, the lack of fall in performance with the longer recall period suggests that DHS/MICS surveys may be able to adopt a longer recall period for this question, thus increasing the period prevalence of true pneumonia cases detected in the survey. This possibility should be studied further, because since increasing the recall period, and thus the true period prevalence, can be expected to improve the validity of the indicator [11].

The sensitivities and specificities of the DHS/MICS survey instruments varied substantially by study site and by urban and rural setting (Table 2), reflecting different cultural factors, such as caregiver education, and making comparisons across countries or time periods difficult to interpret. These results have important implications for monitoring programs because these instruments are used to track progress over time (typically with serial surveys in different populations within a country) and to identify countries that have low intervention coverage in comparison to others. These findings urge caution in using DHS/MICS data for these purposes.

We show that the use of video material shows some potential to complement current DHS/MICS surveys and result in increased specificity for identification of episodes of pneumonia. However, substantial improvements in test specificity were found only in Pakistan, where the video was made, and were not found in Bangladesh. This finding is not unexpected and may relate to the fact that Pakistani children look somewhat different from Bangladeshi children, which is likely to have influenced caregiver responses. Moreover, we elected a priori to focus the video on showing fast breathing and lower chest indrawing. Ethnographic studies from this region have shown that mothers have varying perceptions and concepts regarding childhood pneumonia, and many caregivers do not associate fast breathing with pneumonia [19–21]. The fixed sequence of clips might also have influenced caregiver responses. The potential use of this video tool presents practical and logistical challenges, not least of which is that each individual country may be required to develop their own video. The improved discriminative power with the use of the video in Pakistan is promising, and we recommend that video presentation be explored further in an attempt to further improve its performance.

In children with pneumonia, antibiotic treatment was recalled correctly by two-thirds of caregivers using DHS or MICS questions in Pakistan and Bangladesh, with no significant difference between 2- and 4-wk recall. Correct recall increased with a drug flip chart/photo album or a computer-based drug chart. Thus, the validity of the DHS and MICS questions in correctly identifying antibiotic treatment for pneumonia improved with specific and more structured questions about the medicines or with the use of illustrations of common treatments. Although drug charts performed better than DHS/MICS questions at all three sites, using them in periodic household surveys may be challenging. There may be a very large number of available medicines in any given area, and new medicine brands are introduced often, so frequent revisions would be required.

## Study Limitations

The findings of the survey may have been biased by some aspects of the study design. DHS and MICS surveys are conducted in very diverse urban and rural settings. We attempted to reflect this through the urban and rural study sites selected for study. However, it is possible that the rural site in Bangladesh is not typical because of the ongoing research studies in that population. Furthermore, the urban site in Bangladesh included patients from an expanded urban and peri-urban catchment area, resulting in a relatively high rate of loss to follow-up and delayed follow-up. Thus, our findings may not be generalizable to some settings in low- and middle-income countries. Further studies, in particular in settings with high *Plasmodium falciparum* malaria transmission to explore the generalizability and to check the consistency of these findings, are recommended on this important issue. Nevertheless, the fact that our main study conclusions were confirmed in all three study settings suggests that the main findings may be widely relevant.

We selected study sites with investigators highly experienced in similar studies and employed a single technical expert to conduct the clinical training and oversee the monitoring of research staff in

an attempt to reduce the potential for artifactual differences in study implementation across sites. Nevertheless, some of the differences shown between study sites may reflect some variation in application of methods across sites.

Mothers who take their children to a secondary or tertiary care center do not represent a random sample of all mothers who take part in DHS and MICS surveys. This is likely to influence results. We expect that these mothers will be more educated and show better recognition of signs of pneumonia in their children. To the extent that this is true, we may have overestimated the sensitivity and specificity of DHS/MICS questions. Furthermore, it is also likely that attendance at a hospital makes the episodes more memorable (and thus subject to better recall by the caregiver) and thus may once again lead to overestimation of specificity.

It is possible that the "case mix" in the hospital setting is different from that found in community surveys. We attempted to minimize this by excluding cases of severe pneumonia (rarely found in community surveys due to very low 2-wk period prevalence) and by adopting a matched group who had a respiratory illness (cough and cold) rather than studying healthy children. Nevertheless, this may have influenced our estimates of sensitivity and specificity. Thus, study procedures may have influenced caregiver recall. However, our main finding is that specificity levels associated with DHS and MICS questions are much too low to provide a robust basis for a pneumonia treatment rate indicator. We believe that the biases noted above tend to lead to an overestimation of specificity, and so we consider that our main conclusions about poor discriminative ability should remain valid. In addition, the level of specificity that we found is much lower than would be required, suggesting moderate effects of bias.

### Conclusions

Monitoring antibiotic treatment of pneumonia is essential for national and global programs. Despite this, we believe this is the first study to assess the validity of current surveys to measure this program indicator. Given the very large investment of human and financial resources in these surveys and their importance to child health programs, it is remarkable that there has been so little published data from research on this key issue. In the context of a low true prevalence of pneumonia, the DHS/MICS questions to define children with "suspected pneumonia" (MICS term) has a low yield for pneumonia, and thus most of these children do not have true pneumonia. This finding reinforces DHS and MICS recommendations that these measures not be used as a proxy for pneumonia prevalence, as using these measures will result in substantial overestimates.

We found the discriminative power (in particular the specificity) of a construct of caregiver-reported symptoms and signs of acute respiratory infection to be low for true pneumonia, as measured by DHS and MICS questionnaires; therefore, the use of such data from these surveys can be misleading for measuring antibiotic treatment rates. Using such data could lead to incorrect policy decisions having programmatic implications. Although the alternative tools evaluated in this study did not perform markedly better than DHS and MICS questions, we have presented some options for improving their specificity. Furthermore, the sensitivities and specificities of DHS and MICS survey instruments varied substantially by site, reflecting different cultural factors, such as caregiver education, and making comparisons across countries or time periods difficult to interpret. These results suggest that data from these surveys should not be used for the purpose of monitoring antibiotic treatment rates in children with pneumonia at the present time.

## Supporting Information

**Text S1 Pneumonia module of DHS questionnaire used in Pakistan.**
(DOC)

**Text S2 Pneumonia module of MICS questionnaire used in Pakistan.**
(DOC)

**Text S3 Bangla version of follow-up questionnaire used in Bangladesh.**
(PDF)

**Text S4 STARD checklist for reporting of studies of diagnostic accuracy.**
(DOC)

**Text S5 Original study protocol.** Validation of caregiver report of childhood pneumonia and antibiotic treatment from DHS and MICS data and assessment of alternative approaches.
(DOC)

## Author Contributions

Conceived and designed the experiments: HC SAQ IR JB TH SEA. Performed the experiments: KB AMK SR NK MSK KMR SB MIK SKS ASMN TH SEA. Analyzed the data: HH NK SA ET QSR KB AMK TH SEA HC. Contributed to the writing of the manuscript: TH KB SEA AMK HH NK SA ET SAQ HC. ICMJE criteria for authorship read and met: TH KB SEA AMK HH NK SR SA QSR ET MSK KMR SB MIK SKS ASMN IR JB SAQ HC. Agree with manuscript results and conclusions: TH KB SEA AMK HH NK SR SA QSR ET MSK KMR SB MIK SKS ASMN IR JB SAQ HC.

## References


1. United Nations Children's Fund, World Health Organization, World Bank, United Nations (2012) Levels and trends in child mortality: Report 2012. New York: United Nations Children's Fund.
2. Graham SM, English M, Hazir T, Enarson P, Duke T (2008) Challenges to improving case management of childhood pneumonia at health facilities in resource-limited settings. Bull World Health Organ 86: 349–355.
3. Marsh DR, Gilroy KE, Van de Weerdt R, Wansi E, Qazi S (2008) Community case management of pneumonia: At a tipping point? Bull World Health Organ 86: 381–389.
4. Commission on Information and Accountability for Women's and Children's Health (2011) Keeping promises, measuring results. Available: http://whqlibdoc.who.int/publications/2011/9789241564328_eng.pdf. Accessed 13 March 2013.
5. Harrison LH, Moursi S, Guinena AH, Gadomski AM, el-Ansary KS, Khallaf N, et al. (1995) Maternal reporting of acute respiratory infection in Egypt. Int J Epidemiol 24: 1058–1063.
6. Lanata CF, Quintanilla N, Verastegui HA (1994) Validity of a respiratory questionnaire to identify pneumonia in children in Lima, Peru. Int J Epidemiol 23: 827–834.
7. World Health Organization (1990) Acute respiratory infections in children: Case management in small hospitals in developing countries. WHO/ARI/90.5. Geneva: World Health Organization.
8. Liu L, Johnson HL, Cousens S, Perin J, Scott S, et al. (2012) Global, regional, and national causes of child mortality: An updated systematic analysis for 2010 with time trends since 2000. Lancet 379: 2151–2161.
9. Theodoratou E, Al-Jilaihawi S, Woodward F, Ferguson J, Jhass A, et al. (2010) The effect of case management on childhood pneumonia mortality in developing countries. Int J Epidemiol 39 (Suppl 1): i155–i171.
10. Selwyn BJ (1990) The epidemiology of acute respiratory tract infection in young children: comparison of findings from several developing countries. Rev Infect Dis 12 (Suppl 8): S870–S888.
11. Campbell H, el Arifeen S, Hazir H, O'Kelly J, Bryce J, et al. (2013) Measuring coverage in MNCH: Challenges in monitoring the proportion of young children

with pneumonia who receive antibiotic treatment. PLoS Med 10: e1001421. doi:10.1371/journal.pmed.1001421
12. National Institute of Population Studies, Macro International (2008) Pakistan Demographic and Health Survey 2006-07. Islamabad (Pakistan): National Institute of Population Studies, Macro International.
13. Nizami SQ, Bhutta ZA, Hasan R (2006) Incidence of acute respiratory infections in children 2 months to 5 years of age in periurban communities in Karachi. J Pak Med Assoc 56: 163–167.
14. Castro AV, Nascimento-Carvalho C, Ney-Oliveria F, Araujo-Neto CA, Andrade SCS, et al. (2005) Additional markers to refine the World Health Organization algorithm for diagnosis of pneumonia. Indian Pediatr 42: 773–781.
15. Cardoso MR, Nascimento-Carvalho CM, Ferrero F, Alves FM, Cousens SN (2011) Adding fever to WHO criteria for diagnosing pneumonia enhances the ability to identify pneumonia cases among wheezing children. Arch Dis Child 96: 58–61.
16. Bland RM, Rollins NC, Solarsh G, Van den Broeck J, Coovadia HM, et al. (2003) Maternal recall of exclusive breast feeding duration. Arch Dis Child 88: 778–783.
17. Li R, Scanlon KS, Serdula MK (2005) The validity and reliability of maternal recall of breastfeeding practice. Nutr Rev 63: 103–110.
18. Feikin DR, Audi A, Olack B, Bigogo GM, Polyak C, et al. (2010) Evaluation of the optimal recall period for disease symptoms in home based morbidity surveillance in rural and urban Kenya. Int J Epidemiol 39: 450–458.
19. Mull DS, Mull JD (1994) Insights from community-based research on child pneumonia in Pakistan. Med Anthropol 15: 335–352.
20. Rehman GN, Qazi SA, Mull DS, Khan MA (1994) ARI concepts of mothers in Punjabi villages: A community-based study. J Pak Med Assoc 44: 185–188.
21. Hussain R, Lobo MA, Inam B, Khan A, Qureshi AF, et al. (1997) Pneumonia perceptions and management: An ethnographic study in urban squatter settlements of Karachi, Pakistan. Soc Sci Med 45: 991–1004.

## Editors' Summary

**Background.** Pneumonia is a major cause of death in children younger than five years across the globe, with approximately 1.2 million children younger than five years dying from pneumonia every year. Pneumonia can be caused by bacteria, fungi, or viruses. It is possible to effectively treat bacterial pneumonia with appropriate antibiotics; however, only about 30% of children receive the antibiotic treatment they need. The Millennium Development Goals (MDGs) are eight international development goals that were established in 2000. The fourth goal (MDG 4) aims to reduce child mortality, specifically, to reduce the under-five mortality rate by two-thirds, between 1990 and 2015. Given that approximately 18% of all deaths in children under five are caused by pneumonia, providing universal coverage with effective treatments for pneumonia is an important part of MDG 4.

To ensure that MDG 4 targets are met, it is important to measure progress in providing effective treatments. For pneumonia, one of the key indicators for measuring progress is the proportion of children with pneumonia in a population who receive antibiotic treatment, also known as the antibiotic treatment rate. The antibiotic treatment rate is often measured using surveys, such as the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS), which collect nationally representative data about populations and health in developing countries.

**Why Was This Study Done?** Concerns have been raised about whether information collected from DHS and MICS is able to accurately identify cases of pneumonia. In a clinical setting, pneumonia is typically diagnosed based on a combination of physical symptoms, including coughing, rapid breathing, or difficulty breathing, and a chest X-ray. The surveys rely on information collected from interviews of mothers and primary caregivers using structured questions about whether the child has experienced physical symptoms in the past two weeks and whether these were chest-related. The DHS survey labels this condition as ''symptoms of acute respiratory infection,'' while the MICS survey uses the term ''suspected pneumonia.'' Thus, these surveys provide a proxy measure for pneumonia that is limited by the reliance on the recall of symptoms by the mother or caregiver. Here the researchers have evaluated the use of these surveys to discriminate physician-diagnosed pneumonia and to provide accurate recall of antibiotic treatment in urban and rural settings in Pakistan and Bangladesh.

**What Did the Researchers Do and Find?** The researchers identified caregivers of 950 children under five years with pneumonia and 980 who had a cough or cold but did not have pneumonia from urban and rural settings in Pakistan and Bangladesh. Cases of pneumonia were identified based on a physician diagnosis using World Health Organization guidelines. They randomly assigned caregivers to be interviewed using DHS and MICS questions with either a two- or four-week recall period. They then assessed how well the DHS and MICS questions were able to accurately diagnose pneumonia and accurately recall antibiotic use. In addition, they asked caregivers to complete a pneumonia score questionnaire and showed them a video tool showing children with and without pneumonia, as well as a medication drug chart, to determine if these alternative measures improved the accuracy of pneumonia diagnosis or recall of antibiotic use. They found that both surveys, the pneumonia score, and the video tool had poor ability to discriminate between children with and without physician-diagnosed pneumonia, and there were no differences between using two- or four-week recall. The sensitivity (proportion of pneumonia cases that were correctly identified) ranged from 23% to 72%, and the specificity (the proportion of ''no pneumonia'' cases that were correctly identified) ranged from 53% to 83%, depending on the setting. They also observed that prescribed antibiotics for pneumonia were correctly recalled by about two-thirds of caregivers using DHS questions, and this increased to about three-quarters of caregivers when using a drug chart and detailed enquiry.

**What Do These Findings Mean?** The findings of this study suggest that the current use of questions from DHS and MICS based on mother or caregiver recall are not sufficient for accurately identifying pneumonia and antibiotic use in children. Because these surveys have poor ability to identify children who have true pneumonia, reported antibiotic treatment rates for children with pneumonia based on data from these surveys may not be accurate, and these surveys should not be used to monitor treatment rates. These findings should be interpreted cautiously, given the relatively high rate of loss to follow-up and delayed follow-up in some of the children and because some of the settings in this study may not be similar to other low-income settings.

**Additional Information.** Please access these websites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.1001422.

- More information is available on the United Nations goal to reduce child mortality (MDG 4)
- The World Health Organization provides information on pneumonia, its impact on children, and the global action plan for prevention and control of pneumonia
- More information is available on Demographic and Health Surveys and Multiple Indicator Cluster Surveys
- KidsHealth, a resource maintained by the Nemours Foundation (a not-for-profit organization for children's health) provides information for parents on pneumonia (in English and Spanish)
- MedlinePlus provides links to additional information on pneumonia (in English and Spanish)

PLOS | MEDICINE

# Measuring Coverage in MNCH: Accuracy of Measuring Diagnosis and Treatment of Childhood Malaria from Household Surveys in Zambia

**Thomas P. Eisele[1]*, Kafula Silumbe[2], Josh Yukich[1], Busiku Hamainza[3], Joseph Keating[1], Adam Bennett[1], John M. Miller[2]**

**1** Department of Global Health Systems and Development, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana, United States of America, **2** Malaria Control and Evaluation Partnership in Africa, Program for Appropriate Technology in Health, Lusaka, Zambia, **3** National Malaria Control Centre, Lusaka, Zambia

## Abstract

***Background:*** To assess progress in the scale-up of rapid diagnostic tests and artemisinin-based combination therapies (ACTs) across Africa, malaria control programs have increasingly relied on standardized national household surveys to determine the proportion of children with a fever in the past 2 wk who received an effective antimalarial within 1–2 d of the onset of fever. Here, the validity of caregiver recall for measuring the primary coverage indicators for malaria diagnosis and treatment of children <5 y old is assessed.

***Methods and Findings:*** A cross-sectional study was conducted in five public clinics in Kaoma District, Western Provence, Zambia, to estimate the sensitivity, specificity, and accuracy of caregivers' recall of malaria testing, diagnosis, and treatment, compared to a gold standard of direct observation at the health clinics. Compared to the gold standard of clinic observation, for recall for children with fever in the past 2 wk, the sensitivity for recalling that a finger/heel stick was done was 61.9%, with a specificity of 90.0%. The sensitivity and specificity of caregivers' recalling a positive malaria test result were 62.4% and 90.7%, respectively. The sensitivity and specificity of recalling that the child was given a malaria diagnosis, irrespective of whether a laboratory test was actually done, were 76.8% and 75.9%, respectively. The sensitivity and specificity for recalling that an ACT was given were 81.0% and 91.5%, respectively.

***Conclusions:*** Based on these findings, results from household surveys should continue to be used for ascertaining the coverage of children with a fever in the past 2 wk that received an ACT. However, as recall of a malaria diagnosis remains suboptimal, its use in defining malaria treatment coverage is not recommended.

*Please see later in the article for the Editors' Summary.*

**Citation:** Eisele TP, Silumbe K, Yukich J, Hamainza B, Keating J, et al. (2013) Measuring Coverage in MNCH: Accuracy of Measuring Diagnosis and Treatment of Childhood Malaria from Household Surveys in Zambia. PLoS Med 10(5): e1001417. doi:10.1371/journal.pmed.1001417

**Academic Editor:** David Osrin, Wellcome Trust Senior Research Fellow in Clinical Science, UCL Reader in International Child Health, Honorary Consultant, Great Ormond Street Hospital for Children, United Kingdom

**Received** October 9, 2012; **Accepted** February 22, 2013; **Published** May 7, 2013



**Funding:** This work was funded by the Malaria Control and Evaluation Partnership in Africa (MACEPA), a PATH project, from funding from The Bill & Melinda Gates Foundation. No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** ACT, artemisinin-based combination therapy; CI, confidence interval; DHS, Demographic and Health Surveys; MICS, Multiple Indicator Cluster Surveys; MIS, Malaria Indicator Surveys; RDT, rapid diagnostic test.

* E-mail: teisele@tulane.edu

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection.*

## Introduction

New rapid diagnostic tests (RDTs) and artemisinin-based combination therapies (ACTs) are being scaled up across Africa in the fight against malaria [1,2]. Nearly all endemic countries in Africa have now adopted a policy requiring laboratory confirmation of suspected malaria cases with either RDTs or microscopy [3]. Once cases are confirmed, ACTs are the current first-line drugs for treating uncomplicated *Plasmodium falciparum* malaria infection. These policies are critical to ensure that malaria cases are promptly and properly diagnosed and treated, especially as malaria transmission falls because of increasingly high coverage of effective vector control measures [4].

To assess progress in the scale-up of RDTs and ACTs across Africa, malaria control programs have increasingly relied on standardized national household surveys to assess the proportion of children with a fever in the past 2 wk who received an affective antimalarial within 1–2 d of the onset of fever [5]. This indicator makes no distinction between treatment of a suspected malaria case and one that was laboratory confirmed. Because of the availability and scale-up of RDTs in many African countries, caregivers and mothers are also now asked in national surveys if the child received a heel or finger stick (for an assumed test of

malaria) and if they sought treatment at a health facility [6]. However, caregivers are not typically asked if they recall the result of any malaria diagnostic test given. Information on the test result would be needed to construct a better indicator of the proportion of children with a fever in the past 2 wk with a laboratory-confirmed malaria diagnosis who received an effective and appropriate antimalarial within 1–2 d of the onset of fever. This indicator would capture current diagnosis and treatment policies in most African countries better than the current standard.

Individuals' recall of autobiographical events, including past health events, is a complex process and can be biased by multiple external and internal factors, including the frequency of the event, time since the event, cues provided by interviewers or questionnaire design, and individuals' emotional perception of related events [7]. Survey questions to caregivers of children related to fever in the past 2 wk, treatment-seeking behavior, and malaria diagnosis and treatment are particularly subject to sources of error and bias, which may result in biased coverage estimates for diagnosis and treatment [8]. However, these indicators and their means of measurement to our knowledge have yet to be validated against a gold standard to assess the validity of caregivers' recall.

The objective of this study was to assess the validity of the primary coverage indicators for malaria diagnosis and treatment in children <5 y old, constructed from data collected during household surveys. To accomplish this, a cross-sectional study was conducted in five health clinics in Kaoma District, Western Provence, Zambia, to estimate the sensitivity, specificity, and accuracy of caregivers' recall of malaria testing, diagnosis, and antimalarial treatment compared to a gold standard of direct observation at the health clinics. The relationship between the accuracy of caregiver recall for malaria diagnosis and treatment and socio-demographic characteristics were also assessed.

## Methods

### Ethics Statement

Ethical approval for this study was obtained from the Institutional Review Boards of Tulane University, the University of Zambia, and the Program for Appropriate Technology in Health.

### Study Site

The study was conducted in Kaoma District (population of approximately 183,000), located in Western Province about 460 km west of the capital city of Lusaka (Figure 1). Kaoma is a rural district, with an economy based primarily on commercial and subsistence farming. Kaoma, covering 23,000 $km^2$, has a tropical climate, with seasonal rains lasting from November to April and a corresponding peak in malaria transmission from April to June.

This study took place in four rural public health facilities and one urban public health facility; all were purposefully selected for adequate patient flow of suspected malaria cases in children that received clinical and laboratory diagnoses. The estimated combined catchment area population of the five clinics was 42,100. Malaria burden in the district fluctuates widely during the course of the year with the transmission season. As a percentage of outpatient attendance, malaria diagnoses can range from 2% during the dry season to more than 60% during the transmission season. Malaria test positivity rates among clinic attendees similarly can fluctuate from less than 5% during the dry season to as high as 60%–70% during the peak of the transmission season in April–June. As per national policy in Zambia, artemether/lumefantrine (ACT) was the first-line drug combination used in all five clinics for treating uncomplicated malaria [3]. Although the national policy in Zambia calls for 100% laboratory diagnosis of suspected malaria cases, the clinic in Mulamba primarily relied on clinical diagnosis (diagnosis based on symptoms) because of a stock-out of RDTs throughout the study period; Mulamba does not have microscopy. The remaining four clinics almost exclusively used laboratory diagnosis with RDTs.

### Study Design, Participants, and Data Collection Procedures

A cross-sectional study design was used to assess the validity of the primary coverage indicators for malaria diagnosis and treatment from data collected using household surveys. Data collection at the clinics and at home-based follow-ups occurred in May–June 2012. All caregivers at least 18 y old of children <5 y old presenting for treatment for fever in the five health centers during the study period were eligible for inclusion in this study.

Details of the clinic visit for the episode, including the child's age, presence of fever, diagnostic used (if any), test result, diagnoses made, and treatment provided, were recorded in specially designed clinic visit sheets. These forms were completed for each child at the conclusion of the visit by the attending health professional. While the health professionals received training in completing the forms by the study team prior to the start of the study, they were not informed explicitly that this study aimed to assess caregiver recall of the diagnoses and treatments received at the child's visit. The clinic visit sheets captured all relevant demographic, diagnosis, and treatment information that is typically captured in the standard clinic logbook.

At the conclusion of the clinic visit, caregivers were asked if they would like to participate in a malaria study in their community by the attending health professional. If they agreed, a study representative at the clinic obtained their contact information so that a follow-up visit at their house could be scheduled. To replicate national surveys that ask a series of questions to mothers (Demographic and Health Surveys [DHS] and Malaria Indicator Surveys [MIS]) and caregivers (Multiple Indicator Cluster Surveys [MICS]) of children <5 y old with a fever in the past 2 wk related to malaria diagnosis and treatment, follow-up home visits were scheduled 0–14 d following the clinic visit. Follow-up visit dates were randomly assigned 0–14 d using a random number generator at the time of participant recruitment, and a follow-up appointment time and date were arranged with the caregiver for the follow-up interview. Clinic visit sheets were stapled to the contact information sheet for each caregiver that agreed to a follow-up interview.

Informed consent was obtained from all caregivers at the start of the home-based follow-up interviews. For the majority of children, caregivers were identified as the mother of the child identified at the clinic visit, and as listed on the contact information sheet. If the mother of the child who sought treatment at the clinic did not live in the house, the caregiver of the child was identified and interviewed. If the caregiver did not consent to be in the study at the time of the interview, the contact sheet and clinic visit sheet were destroyed and participation in the study was discontinued. For consenting caregivers, a standardized questionnaire based on the DHS survey [9] and MIS survey (and similar to the MICS survey) was used to obtain information on whether the child had a fever in the past 2 wk; whether treatment was sought, including when and where; whether a diagnostic test was given, based on recall of a finger or heel stick; and details of the treatment given. Questions about recall of the results of a malaria test (if a finger/heel stick was recalled) and about whether any malaria diagnosis was made were added to the questionnaire. No props or cues were

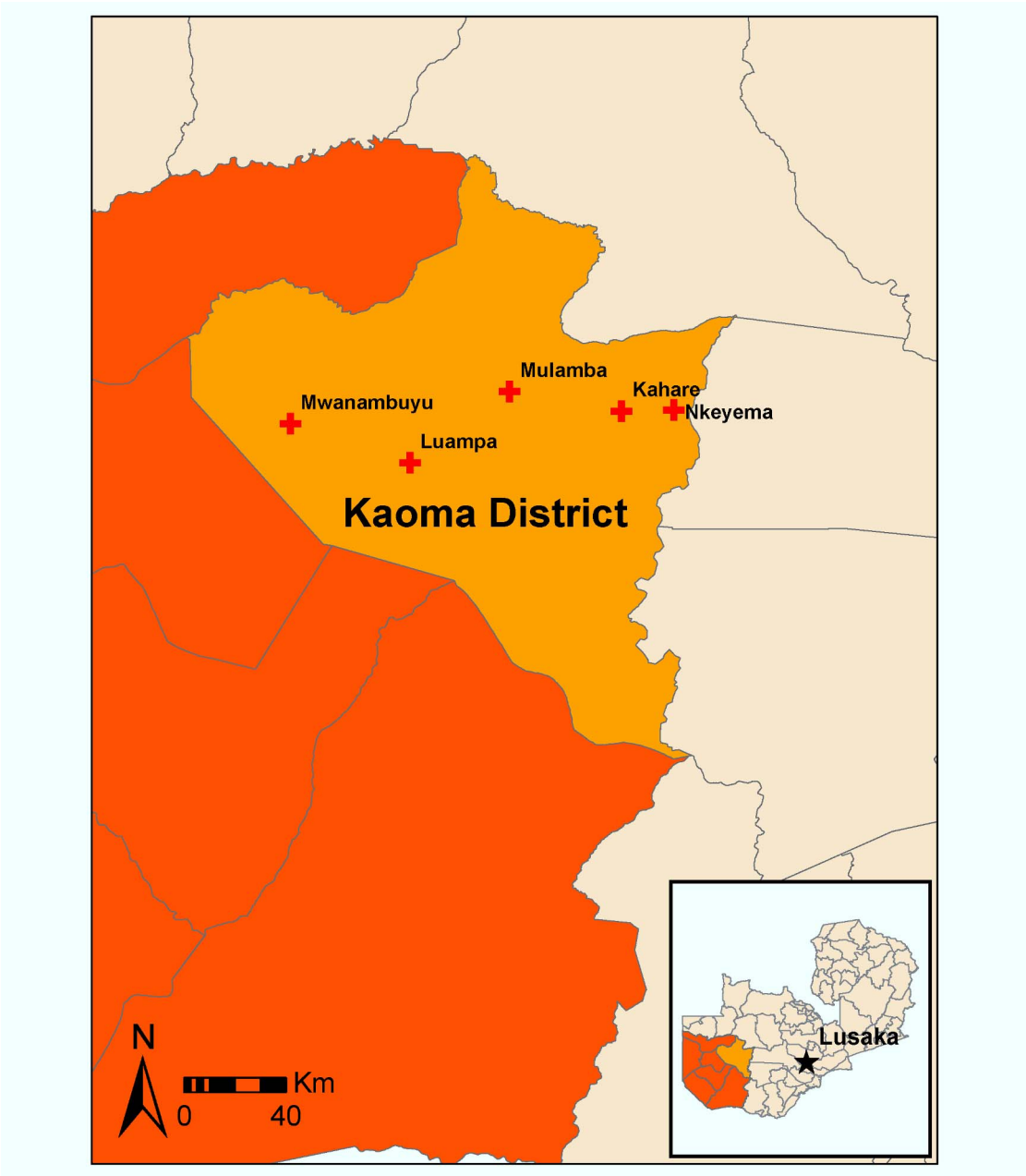


**Figure 1. Map of study clinics in Kaoma District, Western Provence, Zambia.**
doi:10.1371/journal.pmed.1001417.g001

used in asking caregivers about the diagnosis and treatment the child may have received. Information on household assets, as well as the caregiver's sex, age, education level, and literacy, was also obtained based on the questions used in the DHS and MIS surveys. Data were entered into specially designed personal digital assistants used for the Zambia national MIS survey.

## Primary Outcome and Explanatory Variables

The primary outcomes were the sensitivity, specificity, and accuracy of caregivers' recall of the child having received a finger/heel stick (for malaria laboratory diagnostic), result of the malaria diagnostic test, diagnosis for the fever episode, and type of treatment received. Recall of events related to the treatment of the child's fever was ascertained from the follow-up questionnaire. Data recoded on the clinic visit sheet served as the gold standard against which caregivers' recall was assessed. Sensitivity represents the percent of caregivers that correctly recalled the child's receiving a finger/heel stick, a test result, a malaria diagnosis, and a particular antimalarial treatment, of those that actually received them based on clinic observation. Specificity represents the percent of caregivers that correctly recalled the child's not receiving a finger/heel stick, a test result, a malaria diagnosis, and a particular antimalarial treatment, among those that truly did not receive them based on clinic observation. Accuracy represents the percent of caregivers whose recall of the diagnosis and treatment received for their child's fever episode agreed with the clinic observation.

Household socioeconomic status was derived from a principal components analysis of household assets using previously estab-

lished methods, divided into wealth quintiles [10]. Caregiver education was categorized as none, at least some primary, or secondary or higher. Caregiver age was categorized as 18–24, 25–34, 35–44, or 45 y or older. All caregiver demographic and social data were obtained from the follow-up questionnaire. Days to follow-up was the difference between the clinic attendance date and the home follow-up interview date, disaggregated by 0–6 and 7–14 d.

### Sample Size and Stratification

A total sample size of 600 caregivers of children presenting with fever at the five study clinics was sought, stratified equally by those with a laboratory-confirmed malaria diagnosis, those with a laboratory-negative malaria diagnosis, and those without a laboratory diagnosis. The sample size for estimating the sensitivity and specificity of caregiver recall of diagnostic result, malaria diagnosis, and antimalarial treatment was based on the probability of committing a type 1 error of 5% (two-tailed test), a sensitivity of 70%–80%, a specificity of 80%–90%, a precision of ±7.0%, and an interview refusal rate of 10%, resulting in a sample size of 200 caregivers of children who tested positive for malaria and 200 caregivers of children who tested negative for malaria. Children across the five study clinics presenting with fever were selected accordingly, based on stratification by malaria diagnosis. To assess the sensitivity and specificity of caregiver recall of whether the child received a finger/heel stick for malaria diagnosis, an additional 200 caregivers of children who did not receive a finger/heel stick were selected, based on the same parameters outlined above.

### Statistical Analysis

For the primary outcomes related to measuring the coverage of malaria diagnosis and treatment, SAS 9 was used to obtain estimates of sensitivity, specificity, and accuracy, along with 95% confidence intervals (CIs). Standard errors and accompanying 95% CIs for all descriptive point estimates and outcomes were estimated with the Huber–White Sandwich estimator to account for correlated data at the health facility level.

Differences in sensitivity, specificity, and accuracy of malaria diagnosis and treatment between clinics, child characteristics, and caregiver socio-demographic characteristics were tested using chi-square ($\chi^2$). Fisher's exact test was used where any cell count was below ten. Separate logistic regression models were used to assess the association of child and caregiver characteristics with the outcomes of sensitivity, specificity, and accuracy for malaria diagnosis and treatment. As there was considerable heterogeneity across the five clinics with respect to malaria diagnosis, logistic regression models were constructed using both clinic as a random effect and clinic as an independent cluster in generalized estimating equation models. The models with clinic as a random effect yielded more conservative estimates than the generalized estimating equation models in all instances, and were thus chosen to account for heterogeneity across the five health facilities. The final models included child age in years, sex, days to follow-up, caregiver's age category, caregiver's relationship with child (mother or not), caregiver's education, and household wealth quintile, with clinic included as a random effect. Diagnostic test (laboratory versus clinical) was included in models for the outcomes of recall of a positive malaria diagnosis and whether the child was given ACT.

Using the observed sensitivity and specificity in this study for caregiver recall of finger/heel stick, positive diagnostic result, positive malaria diagnosis (laboratory confirmed or clinical), and ACT given, estimates for the coverage of these interventions that one would expect from a household survey of caregivers' recall were modeled across true intervention coverages (observed at clinic) ranging from 0% to 100% as follows: estimated coverage from caregiver recall = (true coverage at clinic × sensitivity) + [(1 − true coverage at clinic) × [1 − specificity)]. Thus, at zero intervention coverage at a clinic (e.g., no finger/heel stick available/offered), the estimated coverage from caregiver recall is equal to the observed 1 − specificity, while at 100% coverage at the clinic, the estimated coverage from caregiver recall is equal to the observed sensitivity. Random sensitivity and specificity (i.e., 50% each) would be expected to yield 50% intervention coverage from caregiver recall, irrespective of the true coverage at a clinic.

## Results

A total of 644 caregivers of children seeking treatment at the five outpatient study clinics were asked to participate in the study. A total of 601 caregivers agreed to be followed up, consented to take part in the study at the time of the follow-up interview, and completed the questionnaire, resulting in a nonresponse rate of 6.7%.

Nearly all (96.5%) of the 115 children diagnosed with malaria in the Mulamba clinic received a diagnosis based on clinical examination rather than on the results of a laboratory test (Table 1). Nearly all (98.3%) of malaria diagnoses in children in the other four study clinics were based on an RDT (one based on microscopy). Among children <5 y old seeking treatment across the five study clinics, if they received a malaria diagnosis, it was based on a laboratory test 67.4% of the time. Of children with a malaria diagnosis, nearly all (97.7%) received an antimalarial, of which 92.1% received the ACT Coartem.

The vast majority (93.3%) of caregivers interviewed at follow-up were the mother of the child, with nearly all (95.7%) being the same person who took the child to the clinic visit (Table 1). Nearly all (95.3%) caregivers interviewed at their homes for the follow-up visit were female, the majority were less than 34 y old, and most had at least some primary school, with a third having attended secondary school or higher (Table 1).

Differences in socio-demographic characteristics of children and caregivers across the five clinics are shown in Table S1. Children were similar across the five study clinics with respect to age and sex (Table S1). Caregivers were similar across study clinics in age, but varied slightly across clinics by sex, relationship to child, education, and household wealth.

Nearly all (96.0%) of the 601 children included in the study were reported to have had a fever in the past 2 wk by the caregiver at the follow-up interview (Table 2). Based on clinic observation, two-thirds (66.9%) of all children were given either an RDT or microscopy (only one child received microscopy) to test for malaria. Of those tested ($n = 402$), over half (58%) had a positive test result. Of all children ($n = 601$), over half had a diagnosis of malaria at the clinic visit, either laboratory confirmed or clinically diagnosed.

During the home interviews, only two-thirds (67.6%) of caregivers recalled that the child was taken to a health facility for a fever in the past 2 wk (among those children with a fever in the past 2 wk, $n = 577$; Table 2). This recall did not vary by whether the caregiver was the mother of the child or someone else in the household ($\chi^2 = 0.70$, df = 1, $p = 0.4033$). However, the caregiver was significantly more likely to have correctly recalled that the child was taken for treatment for the fever if they were the same person who took the child to the clinic ($\chi^2 = 5.0$, df = 1, $p = 0.0279$).

**Table 1.** Characteristics of the children, caregivers, and clinics included in the study, Western Province, Zambia, 2012.

| Characteristic | Number | Point Estimate (Percent) | 95% CI | Sample Size |
|---|---|---|---|---|
| **Child age in years[a]** | | | | 601 |
| 0 | 145 | 24.1 | 20.8–27.4 | |
| 1 | 168 | 28.0 | 17.9–38.0 | |
| 2 | 121 | 20.1 | 17.4–22.8 | |
| 3 | 94 | 15.6 | 11.6–19.6 | |
| 4 | 73 | 12.1 | 6.5–17.8 | |
| **Female child[a]** | 304 | 50.6 | 45.8–55.4 | 601 |
| **Age of caregiver in years[a]** | | | | 601 |
| 18–24 | 247 | 41.1 | 31.6–50.6 | |
| 25–34 | 241 | 40.1 | 34.6–45.6 | |
| 35–44 | 97 | 16.1 | 13.0–19.3 | |
| 45 or older | 16 | 2.7 | 0.0–6.7 | |
| **Female caregiver[a]** | 573 | 95.3 | 89.2–100.0 | 601 |
| **Mother of child as caregiver[a]** | 561 | 93.3 | 82.5–100.0 | 601 |
| **Caregiver interviewed same as who took child to clinic visit[a]** | 575 | 95.7 | 88.9–100.0 | 601 |
| **Caregiver education[a]** | | | | 601 |
| None | 42 | 7.0 | 1.1–12.9 | |
| At least some primary | 346 | 57.6 | 39.2–75.9 | |
| Secondary or higher | 213 | 35.4 | 15.2–55.7 | |
| **Mean number of days between clinic visit and survey interview[a]** | | 5.2 | 2.8–7.5 | 601 |
| **Proportion of observations from each clinic** | | | | 601 |
| Kahare | 114 | 19.0 | 15.8–22.1 | |
| Luampa | 101 | 16.8 | 13.8–19.8 | |
| Mulamba | 149 | 24.8 | 21.3–24.1 | |
| Mwanambuyu | 125 | 20.8 | 17.5–24.1 | |
| Nkeyema | 112 | 18.6 | 15.5–21.8 | |
| **Proportion of malaria diagnoses that were laboratory confirmed** | | | | |
| Kahare | 48 | 100 | — | 48 |
| Luampa | 42 | 97.7 | 93.0–100 | 43 |
| Mulamba | 4 | 3.5 | 0.1–6.9 | 115 |
| Mwanambuyu | 61 | 95.3 | 90.0–100 | 64 |
| Nkeyema | 83 | 100 | — | 83 |

[a]Standard errors estimated using the Huber–White Sandwich estimator to account for correlated data at the facility level.
doi:10.1371/journal.pmed.1001417.t001

The sensitivity, specificity, and accuracy of recalling whether the child received a finger/heel stick, a positive laboratory test result, a positive malaria diagnosis, and treatment with an ACT varied significantly across the five study clinics (Tables S2 and S3). While no clear pattern emerged, caregivers of children seen at the Mwanambuyu clinic had the most accurate recall of these events, while caregivers from the Kahare, Luampa, and Mulamba clinics had the least accurate recall of these events.

Of all children sampled, caregivers' recall of their child receiving a finger or heel stick was nearly identical to their recall that the child was tested for malaria with a laboratory test (43.7% and 43.3%, respectively; Table 2). Of children tested ($n = 260$), 90.0% of their caregivers reported that the results were shared with them, with two-thirds (67.3%) reporting that the test was positive. Among all children, over half the caregivers (53.4%) recalled that a malaria diagnosis (laboratory confirmed or clinical) was made, and over half (50.6%) reported that the child received an ACT.

Of children with fever in the past 2 wk ($n = 577$), the sensitivity for recalling that a finger/heel stick was done was 61.9% (95% CI 58.1%–67.7%), with a specificity of 90.0% (95% CI 85.7%–94.2%); results of recalling whether a malaria laboratory test was performed were nearly identical. Based on the random effects logistic regression (Table S4), there was no association between caregiver recall of a finger/heel stick and child and caregiver characteristics.

Compared to the gold standard of clinic observation, for children with fever in the past 2 wk who received a laboratory diagnosis at the clinic, the sensitivity for recalling that it was positive was 62.4% (95% CI 56.1%–68.7%), with a specificity of 90.7% (95% CI 86.3%–95.2%). Table S4 presents the random effects logistic regression models assessing factors related to the

**Table 2.** Characteristics of treatment recall by caregivers and observation at study clinics, Western Province, Zambia, 2012.

| Characteristic | Number | Point Estimate (Percent) | 95% CI | Sample Size |
|---|---|---|---|---|
| **Caregiver recall for household interview** | | | | |
| Child with fever in past 2 wk | 577 | 96.0 | 86.8–100.0 | 601 |
| Child taken for treatment for fever, of those with fever in past 2 wk | 390 | 67.6 | 33.8–100.0 | 577 |
| Child received finger/heel stick | 263 | 43.7 | 0.0–89.5 | 601 |
| Child tested for malaria with RDT | 259 | 43.1 | 0.0–89.0 | 601 |
| Child tested for malaria with microscopy[a] | 1 | 0.2 | 0–0.6 | 601 |
| Of those for whom caregiver recalled child being tested, caregiver had result shared with them | 234 | 90.0 | 79.6–100.0 | 260 |
| Of those for whom caregiver recalled child being tested, child with positive test | 175 | 67.3 | 39.9–94.7 | 260 |
| Child with malaria diagnosis[b] | 321 | 53.4 | 40.3–66.5 | 601 |
| Child given any antimalarial | 328 | 54.6 | 19.6–89.5 | 601 |
| Child given ACT | 304 | 50.6 | 12.2–89.0 | 601 |
| **Observed at clinic** | | | | |
| Child tested for malaria with RDT | 401 | 66.7 | 12.2–100.0 | 601 |
| Child tested for malaria with microscopy[a] | 1 | 0.2 | 0–0.6 | 601 |
| Of tested, child with positive test | 233 | 58.0 | 40.2–75.7 | 402 |
| Child with malaria diagnosis[b] | 353 | 58.7 | 37.0–80.4 | 601 |
| Child given any antimalarial | 381 | 63.4 | 43.2–83.6 | 601 |
| Child given ACT | 351 | 58.4 | 34.2–82.6 | 601 |

All standard errors estimated using the Huber–White Sandwich estimator to account for correlated data at the facility level.
[a]Only one child tested with microscopy.
[b]Includes laboratory-confirmed malaria and clinical diagnosis based on symptoms.
doi:10.1371/journal.pmed.1001417.t002

sensitivity, specificity, and accuracy of a caregivers' recall. The accuracy of recall of a positive test result was also significantly higher in caregivers with secondary or higher education, as compared to those with just primary education.

Among all children with a fever in the past 2 wk, the sensitivity of recalling that the child was diagnosed with malaria, irrespective of whether a laboratory test was actually done, was 76.8% (95% CI 72.4%–81.3%), with a specificity of 75.9% (95% CI 70.4%–81.4%). Based on logistic regression (Table S4), the sensitivity of recalling that a malaria diagnosis was made was significantly better when a laboratory diagnosis was used, as compared to clinical diagnosis (adjusted odds ratio = 1.4, 95% CI 1.3–1.6).

Compared to clinic observation, among children with a fever in the past 2 wk, the sensitivity and specificity for recalling that an ACT was given were 81.0% (95% CI 76.8%–85.2%) and 91.5% (95% CI 87.9%–95.1%), respectively, resulting in an accuracy of 85.3% (95% CI 82.4%–88.2%; Table 3). Recall that any antimalarial was given was nearly identical. Logistic regression showed that the accuracy of recalling that an ACT was given was significantly better when the child received a laboratory diagnostic test (adjusted odds ratio = 1.2, 95% CI 1.1–1.3), as compared to clinical diagnosis (Table S4).

The modeled coverage of finger/heel stick ascertained from caregiver recall at various coverage scenarios, based on the sensitivity and specificity observed in this study, shows that at low coverage levels, caregiver recall would tend to only slightly overestimate the true coverage of finger/heel sticks at clinics (Figure 2). However, where about 80% of children with a fever actually receive a finger/heel stick at clinics, the low sensitivity of caregiver recall of this happening would lead to a substantial underestimation of the coverage of this indicator. Modeled estimates of the recall of a positive malaria diagnosis based on the accuracy of caregivers' recall show this indicator to perform poorly, with substantial overestimation when the true level of malaria diagnosis in clinics is low, and substantial underestimation when positive malaria diagnoses are common (i.e., above 50%). The modeled estimates of the recall that a child with a fever received an ACT based on the accuracy of caregivers' recall performed much better, with only a slight overestimation when actual prescription of ACT is low (i.e., below 20%), while yielding a slight underestimation when prescription of ACT is more common.

## Discussion

A cross-sectional study was conducted in five rural health clinics in Kaoma District, Zambia, to estimate the sensitivity, specificity, and accuracy of caregivers' recall of malaria testing, malaria diagnosis, and antimalarial treatment. Observation of malaria diagnosis and treatment at the clinics served as the gold standard.

Overall, the sensitivity of caregiver recall that a malaria diagnostic test was performed on a child reported to have had a fever in the past 2 wk was poor in this setting, irrespective of whether the questionnaire asked if a finger or heel stick was done (62.9%) or whether a blood test for malaria was done (61.9%). Estimates of specificity for these questions were higher, at 90.0% and 89.4%, respectively. These results mean that about four out of ten respondents do not recall that a malaria diagnostic test was done when in fact the child received one, while relatively few (one in ten) incorrectly recall that a diagnostic was done when in reality one was not done. These results suggest that while there is substantial recall error (i.e., forgetting a child received a test), there is limited information bias related to stating a child received something he or she did not really receive. These results further

**Table 3.** Accuracy of caregiver recall of key questions of diagnosis and treatment of malaria, Western Province, Zambia, 2012.

| Caregiver Recall | TP | TP+FN | Sensitivity (Percent) | 95% CI | TN | TN+FP | Specificity (Percent) | 95% CI | TP+TN | TP+TN+FP+FN | Accuracy (Percent) | (95% CI) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Recall of fever in past 2 wk | 577 | 601 | 96.0 | (86.8–100.0) | 0 | 0 | 100.0 | — | 577 | 601 | 96.0 | (86.8–100.0) |
| Recall of finger/heel stick[a] | 244 | 388 | 62.9 | (17.8–100.0) | 170 | 189 | 90.0 | (85.8–94.1) | 414 | 577 | 71.8 | (41.0–100.0) |
| Recall of malaria laboratory diagnosis performed[a] | 240 | 388 | 61.9 | (16.6–100.0) | 170 | 189 | 89.4 | (84.2–95.7) | 410 | 577 | 70.9 | (39.9–100.0) |
| Recall of positive malaria test result (of those tested at clinic)[a] | 141 | 226 | 62.4 | (14.2–100.0) | 147 | 162 | 90.7 | (75.6–100.0) | 288 | 388 | 74.2 | (49.6–98.9) |
| Recall that malaria diagnosis was made[b] | 265 | 345 | 76.8 | (55.1–98.5) | 176 | 232 | 75.9 | (48.3–100.0) | 441 | 577 | 76.4 | (55.2–97.6) |
| Recall of any antimalarial given[a] | 305 | 372 | 82.0 | (53.2–100.0) | 182 | 205 | 88.8 | (76.5–100.0) | 487 | 577 | 84.4 | (69.5–99.3) |
| Recall of ACT given[a] | 277 | 342 | 81.0 | (50.8–100.0) | 215 | 235 | 91.5 | (80.3–100.0) | 492 | 577 | 85.3 | (72.7–97.8) |

All standard errors estimated using the Huber–White Sandwich estimator to account for correlated data at the facility level.
[a]Among children reported by caregiver to have a fever in the past 2 wk.
[b]Includes laboratory-confirmed malaria and clinical diagnosis based on symptoms.
CI, confidence interval; FN, false negative; FP, false positive; TN, true negative; TP, true positive.
doi:10.1371/journal.pmed.1001417.t003

suggest that the indicator used for measuring the coverage of laboratory diagnosis of malaria among children with a fever in the past 2 wk will yield substantial underestimates of RDT and microscopy coverage, within similar settings. This finding is consistent with the 2010 Zambia MIS, which showed a lower estimate of children receiving a finger or heel stick than what would be expected based on the programmatic data on the availability and use of RDTs across public facilities in Zambia [6].

Of those children that actually received a laboratory diagnosis, the sensitivity (62.4%) and specificity (90.7%) of recalling a positive test result were very similar to those of recalling that the test was done. This suggests that many caregivers fail to recall a positive test result, while relatively few state the child had a positive test result when they actually did not.

Among children with a fever in the past 2 wk, irrespective of whether a laboratory or clinical diagnosis was used at the clinic, there was relatively poor sensitivity (76.8%) and specificity (75.9%) of caregiver recall that a positive malaria diagnosis was made. The sensitivity and specificity of caregiver recall of a positive malaria diagnosis were significantly lower in the Mulamba clinic, which primarily based malaria diagnosis on clinical symptoms because of a stock-out of RDTs. This coincides with the finding that a laboratory diagnosis results in significantly better sensitivity (86.6%) of caregiver recall of a positive malaria diagnosis, as compared to a clinical diagnosis (57.0%). While the sensitivity of caregiver recall of a malaria diagnosis may improve as RDTs are scaled up across Africa and more children receive laboratory diagnoses, these results clearly suggest that more can be done to ensure physicians and nurses share the child's diagnosis with the caregiver at the visit.

Within this setting, the sensitivity (81.0%) and specificity (91.5%) of caregiver recall that the child received an ACT was reasonably high, resulting in an overall accuracy of 85.3%. These results suggest that the indicator for measuring the coverage of children with a fever in the past 2 wk who received an ACT may yield reasonable estimates. However, this study included only public clinics in Zambia where ACTs are supplied free of charge and where a concerted effort has been made to limit the use of monotherapies, such as sulfadoxine-pyrimethamine and chloroquine, for treating uncomplicated malaria in children [3,11]. The estimates of sensitivity and specificity of ACT recall observed in this setting may be considerably higher than in other settings where ACTs are not free and where substantial use of antimalarial monotherapies persists; use of antimalarial monotherapies may be especially high in settings where a large share of children are taken to private facilities [12].

The Mwanambuyu clinic had consistently higher accuracy of recall of a finger/heel stick, a positive malaria test result, malaria diagnosis, and whether the child received an ACT (Tables S2 and S3). We surmise this was likely due to the population of the clinic having a higher education level (Table S1).

The standard recall period for asking about details of treatment of fevers in children in the DHS and MICS surveys is 2 wk. In this study setting, there was no observed drop-off in the sensitivity, specificity, and accuracy of caregiver recall of malaria diagnosis and treatment at a clinic visit that occurred 0–6 d prior to the survey interview compared to one that occurred 7–14 d prior. This suggests that there would be no benefit of improving the accuracy of caregiver recall by limiting survey questions to children that had a fever within a shorter time frame than the previous 2 wk.

The DHS survey asks details about a child's fever episode from the mother, while the MICS survey asks these details from the mother, if available, or, if the mother has died or does not live at

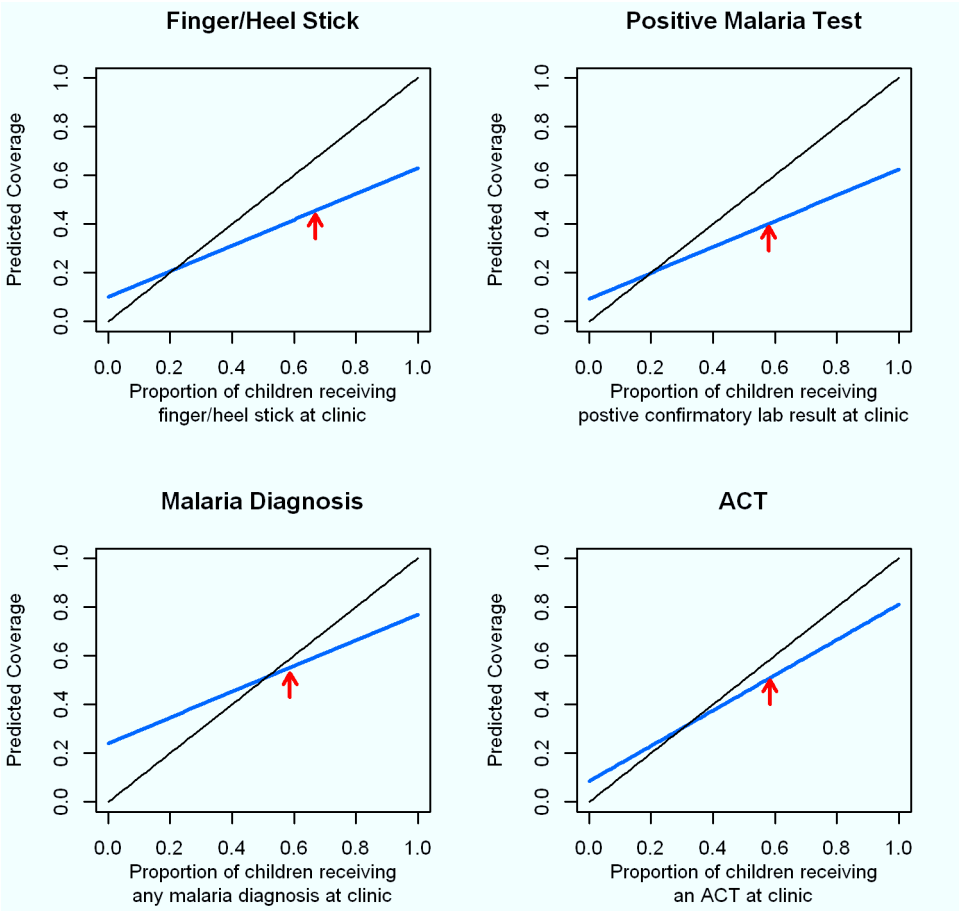


**Figure 2. Modeled diagnosis and treatment coverage based on sensitivity and specificity of caregiver recall across actual intervention coverages in a given community.** Proportions of patients actually experiencing each event at the study clinics are illustrated with red arrows. The solid black line at a 45° angle represents 100% sensitivity and specificity. Estimates for the coverage of these interventions expected from a household survey from caregiver recall with the sensitivity and specificity observed in this study (blue line) were modeled for true intervention coverages (observed at clinic) ranging from 0% to 100% as follows: estimated coverage from caregiver recall = (true coverage at clinic × sensitivity) + [(1 − true coverage at clinic) × [1 − specificity)].
doi:10.1371/journal.pmed.1001417.g002

the house, from a caregiver who is present in the house [13]. It is helpful that results from this study show little if any difference between the accuracy of recall of malaria diagnosis and treatment based on whether the caregiver is the mother or someone else in the house that cares for the child. This reinforces the comparability between the DHS and MICS surveys for measuring malaria treatment coverage indicators.

Several similar studies published as companion papers in this Collection measured the validity of indicators from household coverage surveys in estimating the prevalence of health events [14,15]. Coinciding with our results of poor recall of malaria diagnosis, results from the study of childhood pneumonia prevalence and antibiotic treatment prevalence generally showed poor sensitivity and specificity for the classification of "true pneumonia" as opposed to "suspected pneumonia" or "symptoms of acute respiratory infection" and, subsequently, appropriate antibiotic treatment [14] While there is some evidence that caregiver recall of past health events may be highly sensitive and specific in some instances, including mothers' recall of child vaccination status [16], several studies have shown recall of previous health events to be quite poor and often biased. A study of recall of essential drug treatment dosage and duration in rural Burkina Faso showed that while approximately 70% of patients interviewed after initial consultation could remember the dosage of drugs, only 30% recalled the duration of appropriate treatment [17]. While our study showed that caregivers recalled a child receiving an ACT with relatively high sensitivity and specificity, the results on other indicators, including whether or not the child received a malaria diagnostic test and the results of the test, were much less promising. This indicates that it will be difficult to estimate accurately the proportion of children receiving appropriate malaria treatment with household survey methods. Though our study found high sensitivity and specificity of fever recall, this may be due to interviewer bias or to participation in the study positively influencing recall. Measures of fever, clinic attendance, and antimalarial prescription recall conducted in Demographic Surveillance System settings in western Kenya indicated that fever recall over a 2-wk period underestimates the true prevalence because of recall bias, and that the bias tends to increase when the time from the symptom to the interview increases; such changes were more pronounced among caregivers than among adult patients [18]. That study also showed similar results for documented clinic attendances, antimalarial prescriptions, and antibiotic prescriptions. Furthermore, differences in questionnaire

design may also influence the results for recall of past health events; a study of antimalarial drug utilization recall conducted in Mozambique indicated that recall of which antimalarial drug was used to treat a past episode of malaria could be influenced by the ordering of the drugs as read to the respondent [19].

Results from this study should be treated with caution for several reasons. First, this study may have limited broad external validity across Zambia and other African settings for several reasons, and results may actually represent a best-case scenario for the accuracy of caregiver recall of malaria diagnosis and treatment events. This study was conducted among only five public health clinics in a single rural district of Zambia selected by convenience; they are not representative of all of Zambia. Zambia has achieved substantial scale-up of malaria control measures with considerable resources from the government and international donors [20]. One hundred percent laboratory confirmation of suspected malaria is encouraged under the current national diagnosis and treatment policy, with substantial resources allocated to scaling up use of RDTs since 2007 [3]. Compared to many other African settings, there may have been fewer stock-outs of ACTs. Second, while details of the study aims and objectives were not explicitly presented to the clinic staff, it is possible their behaviors were influenced by a "malaria study" being conducted, which may have influenced caregiver recall. Similarly, while caregivers recruited into the study were not informed of the exact details of the study until the informed consent was obtained at the start of the follow-up interview, it is possible that by agreeing to a follow-up at their home for a "malaria study," their recall of events at the clinic may have been biased. Third, while the information in the clinic visit sheets was entered by the attending health professionals themselves, we were not able to validate this information against a secondary direct observation. And last, props, such as the box that Coartem (artemether/lumefantrine) most commonly comes in or commonly used RDT cassettes, were not used during the survey interviews. It is possible the lack of such props, which are commonly used in many national surveys, may have hindered the accuracy of caregiver recall [7].

### Conclusions and Recommendations

While specificity of caregiver recall was reasonable, compared to a gold standard of direct observation at clinics, results from this study show there to be relatively poor sensitivity of caregiver recall of receiving a malaria diagnostic test and having the diagnostic test result shared, for children with a fever in the past 2 wk. These results suggest that at the current low coverage levels of malaria diagnosis across most African countries, estimates of diagnostic coverage may be reasonable (with slight overestimation due to less than perfect specificity), but as coverage in the population increases, survey estimates will yield increasingly larger underestimates of the true coverage because of the poor sensitivity of caregiver recall. Using household surveys to measure trends in the population coverage of laboratory diagnostics would therefore mask the true program success where high coverage of access to laboratory diagnoses is achieved. Recall of malaria diagnosis for children with a fever, irrespective of whether a laboratory diagnosis was made, had poor sensitivity and specificity, rendering this a very poor indicator for program managers to assess trends over time for this indicator. However, the accuracy of caregiver recall that a child received an ACT was relatively high in this setting, suggesting that the current indicator for measuring the coverage of children with a fever in the past 2 wk who received an ACT may yield reasonable estimates in similar settings, and may prove useful for measuring trends over time.

Based on these findings, results from surveys should continue to be used for ascertaining the coverage of children with a fever in the past 2 wk that received an ACT. However, as recall of a malaria diagnosis remains suboptimal, its use in defining malaria treatment coverage is not recommended.

As laboratory diagnosis is scaled up, these results suggest recall of any malaria diagnosis may improve; continued research to assess changes in malaria diagnosis recall as RDTs are scaled up is recommended. In the meantime, better communication between health professionals and caregivers should be promoted through additional training, with a focus on communicating diagnostics used, the child's diagnosis, and the treatment provided. For tracking progress towards targets for prompt, effective treatment of malaria, it is recommended that program managers and policy makers use household survey data only for measuring coverage of treatment seeking for fevers and access to antimalarial drugs. If possible, these data should then be supplemented with data from health system programs on the proportion of suspected malaria cases that receive a laboratory malaria diagnostic test, and the proportion of suspected and laboratory-confirmed malaria cases that receive the appropriate antimalarial. Where possible, studies consisting of exit interviews with caregivers following fever consultations would also prove useful for estimating the proportion of children receiving appropriate malaria diagnosis and treatment at the health system level.

## Supporting Information

**Table S1 Characteristics of the children, caregivers, and households, by clinic, Western Province, Zambia, 2012.**
(DOC)

**Table S2 Accuracy of caregiver recall of key questions of diagnosis and treatment of malaria for children with reported fever in the past 2 wk, by follow-up and social-demographic characteristics, Western Province, Zambia, 2012.**
(DOC)

**Table S3 Bivariate statistical test for differences of accuracy of caregiver recall of key questions of diagnosis and treatment of malaria for children with reported fever in the past 2 wk, by follow-up and social-demographic characteristics, Western Province, Zambia, 2012.**
(DOC)

**Table S4 Random effects logistic regression models of sensitivity, specificity, and accuracy of caregiver recall of key questions of diagnosis and treatment of malaria for children with reported fever in the past 2 wk: associations with follow-up and social-demographic characteristics, Western Province, Zambia, 2012.**
(DOC)

## Acknowledgments

We wish to thank the community members and health care providers from Kaoma District, Western Province, Zambia. We thank the Kaoma District Health Office, the District Medical Officer (Dr. Mwepa Mwewa), the Malaria Focal Person (Shabalala Lawrence), and the District Information Officer (Myambango Mwenda). We also thank the Zambia National Malaria Control Centre for their support to conduct this research. Benjamin Kayungwa is thanked for his support in programming the questionnaire in

the personal digital assistants. Dr. Harry Campbell and Dr. Jennifer Bryce are thanked for their valuable comments on the draft manuscript.

## Author Contributions

Conceived and designed the experiments: TPE KS JM. Performed the experiments: TPE KS BH JM. Analyzed the data: TPE JY AB. Wrote the first draft of the manuscript: TPE. Contributed to the writing of the manuscript: TPE KS JY BH JK AB JM. ICMJE criteria for authorship read and met: TPE KS JY BH JK AB JM. Agree with manuscript results and conclusions: TPE KS JY BH JK AB JM. Enrolled patients: KS BH.

## References

1. Roll Back Malaria Partnership (2008) Global malaria action plan: for a malaria free world. Geneva: Roll Back Malaria Partnership.
2. Roll Back Malaria Partnership (2011) Rolling back malaria: a decade of progress in malaria control. Geneva: Roll Back Malaria Partnership.
3. World Health Organization (2010) World malaria report 2010. Geneva: World Health Organization.
4. Lim SS, Fullman N, Stokes A, Ravishankar N, Masiye F, et al. (2011) Net benefits: a multicountry analysis of observational data examining associations between insecticide-treated mosquito nets and health outcomes. PLoS Med 8: e1001091. doi:10.1371/journal.pmed.1001091
5. Roll Back Malaria Partnership (2008) Guidelines for core population-based indicators. Geneva: Roll Back Malaria Partnership.
6. Zambia Ministry of Health (2010) Zambia national malaria indicator survey 2010. Lusaka: Zambia Ministry of Health.
7. Bradburn NM, Rips LJ, Shevell SK (1987) Answering autobiographical questions: the impact of memory and inference on surveys. Science 236: 157–161.
8. Thwing J, Eisele T, Steketee R (2011) Protective efficacy of malaria case management and intermittent preventive treatment for preventing malaria mortality in children: a systematic review for the Lives Saved Tool. BMC Public Health 11: S14.
9. MEASURE DHS Demographic and Health Surveys. Available: http://www.measuredhs.com/. Accessed 14 March 2013.
10. Rutstein SO, Johnson K (2004) The DHS wealth index. Calverton (Maryland): Macro International.
11. Sipilanyambe N, Simon JL, Chanda P, Olumese P, Snow RW, et al. (2008) From chloroquine to artemether-lumefantrine: the process of drug policy change in Zambia. Malar J 7: 25.
12. O'Connell KA, Gatakaa H, Poyer S, Njogu J, Evance I, et al. (2011) Got ACTs? Availability, price, market share and provider knowledge of anti-malarial medicines in public and private sector outlets in six malaria-endemic countries. Malar J 10: 326.
13. Hancioglu A, Arnold F (2013) Measuring coverage in MNCH: Tracking progress in health for women and children using DHS and MICS household surveys. PLoS Med 10: e1001391. doi:10.1371/journal.pmed.1001391
14. Hazir T, Bequem K, el Arifeen S, Khan AM, Huque MH, et al. (2013) Measuring coverage in MNCH: A prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia. PLoS Med 10: e1001422. doi:10.1371/journal.pmed.1001422
15. Stanton CK, Rawlins B, Drake M, Anjos M, Cantor D, et al. (2013) Measuring coverage in MNCH: Testing the validity of women's self-report of key maternal and newborn health Interventions during the peripartum period in Mozambique. PLoS ONE 8: e60694. doi:10.1371/journal.pone.0060694
16. Ndirangu J, Bland R, Barnighausen T, Newell ML (2011) Validating child vaccination status in a demographic surveillance system using data from a clinical cohort study: evidence from rural South Africa. BMC Public Health 11: 372.
17. Krause G, Benzler J, Heinmuller R, Borchert M, Koob E, et al. (1998) Performance of village pharmacies and patient compliance after implementation of essential drug programme in rural Burkina Faso. Health Policy Plan 13: 159–166.
18. Feikin DR, Audi A, Olack B, Bigogo GM, Polyak C, et al. (2010) Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. Int J Epidemiol 39: 450–458.
19. Lunet N, Bastos J, Cumaio F, Silva P, Dias E, et al. (2008) Recall of drug utilization depends on subtle structural questionnaire characteristics. Pharm World Sci 30: 175–181.
20. Chizema-Kawesha E, Miller JM, Steketee RW, Mukonka VM, Mukuka C, et al. (2010) Scaling up malaria control in Zambia: progress and impact 2005–2008. Am J Trop Med Hyg 83: 480–488.

## Editors' Summary

**Background.** The World Health Organization estimates that there are over 200 million cases of malaria each year, with nearly 1 million deaths. The majority of these deaths are among children living in sub-Saharan Africa, and *Plasmodium falciparum* is the parasite responsible. Malaria transmission can be prevented by insect control measures, and current treatment regimens use antimalarial drugs. Recently, the use of highly effective artemisinin-based combination treatments (ACTs) has significantly reduced the deaths and disability caused by malaria. To avoid drug overuse and the development of parasite resistance to ACTs, the World Health Organization recommends that before treatment with ACTs, a laboratory test to confirm malaria should be performed. Rapid diagnostic tests (RDTs) allow health workers to diagnose malaria in settings lacking laboratory facilities, thus providing a method for improving malaria diagnosis and reducing the overuse of ACTs.

**Why Was This Study Done?** The success of RDTs and ACTs across Africa in combating malaria is measured by standardized national household surveys. These surveys assess the proportion of children with a fever in the past two weeks who have received an antimalarial treatment within 1–2 days of the onset of fever. The surveys do not distinguish between treatment of a suspected malaria case and one that was laboratory confirmed. Due to the availability and scale-up of RDTs in many African countries, caregivers and mothers are also now asked in national surveys if the child was tested for malaria, but are not usually asked for the result of any malaria diagnostic test given. Knowing whether a child has been diagnosed with malaria is necessary to construct a better indicator of what proportion of children receive an effective and appropriate antimalarial within the appropriate treatment time frame. This indicator is important because it provides more insight into the current diagnosis and treatment policies in most African countries. Biased coverage estimates for diagnosis and treatment may result from these types of surveys because survey questions to caregivers of children concerning fever in the past two weeks, treatment-seeking behavior, and malaria diagnosis and treatment can be particularly subject to sources of error and bias. Despite this possibility, these indicators and surveys have not been checked against direct observation of the children to assess the validity of caregivers' recall and household surveys to gauge appropriate treatment of malaria in children.

**What Did the Researchers Do and Find?** In this study, the authors investigated the validity of caregiver recall of malaria diagnosis and treatment in children under five years old. The authors did a cross-sectional study of five public clinics in Kaoma District, Western Provence, Zambia, to estimate the sensitivity, specificity, and accuracy of caregivers' recall of malaria testing, malaria diagnosis, and antimalarial treatment, and compared the surveys to direct observation at the health clinics. The results from this study demonstrate low sensitivity of caregiver recall of malaria diagnostic use, test results, and malaria diagnosis among children who had a fever in the past two weeks. However, the accuracy of caregiver recall that a child received an ACT was relatively high in this setting. This suggests that the current indicator for measuring the coverage of children with a fever in the past two weeks who received an ACT can be applied in similar settings, and may be useful for estimating infection and treatment over time.

**What Do These Findings Mean?** These findings suggest that results from household surveys are accurate for obtaining information about the coverage of children with a fever in the past two weeks that receive an ACT. However, as caregiver recall of a malaria diagnosis is not highly sensitive, the authors suggest that malaria diagnosis from caregiver recall in household surveys is not recommended for defining malaria treatment coverage.

**Additional Information.** Please access these websites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.1001417.

- More information about malaria is available from MedlinePlus, the World Health Organization, and the US Centers for Disease Control and Prevention
- The Roll Back Malaria Partnership brings numerous organizations together to combat malaria around the globe

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: Challenges in Monitoring the Proportion of Young Children with Pneumonia Who Receive Antibiotic Treatment

**Harry Campbell[1]*, Shams el Arifeen[2], Tabish Hazir[3], James O'Kelly[1], Jennifer Bryce[4], Igor Rudan[1], Shamim Ahmad Qazi[5]**

**1** Centre for Population Health Sciences, University of Edinburgh, Edinburgh, United Kingdom, **2** International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh, **3** Children's Hospital, Pakistan Institute of Medical Sciences, Islamabad, Pakistan, **4** Department of International Health, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States of America, **5** Department of Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland

**Abstract:** Pneumonia remains a major cause of child death globally, and improving antibiotic treatment rates is a key control strategy. Progress in improving the global coverage of antibiotic treatment is monitored through large household surveys such as the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS), which estimate antibiotic treatment rates of pneumonia based on two-week recall of pneumonia by caregivers. However, these survey tools identify children with reported symptoms of pneumonia, and because the prevalence of pneumonia over a two-week period in community settings is low, the majority of these children do not have true pneumonia and so do not provide an accurate denominator of pneumonia cases for monitoring antibiotic treatment rates. In this review, we show that the performance of survey tools could be improved by increasing the survey recall period or by improving either overall discriminative power or specificity. However, even at a test specificity of 95% (and a test sensitivity of 80%), the proportion of children with reported symptoms of pneumonia who truly have pneumonia is only 22% (the positive predictive value of the survey tool). Thus, although DHS and MICS survey data on rates of care seeking for children with reported symptoms of pneumonia and other childhood illnesses remain valid and important, DHS and MICS data are not able to give valid estimates of antibiotic treatment rates in children with pneumonia.

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection.*

## Introduction

Pneumonia has been the largest single cause of child death over the 2000–2015 Millennium Development Goal period [1,2], and despite large falls in global under-five mortality, pneumonia remains the major single cause of child death in the post-neonatal period [3]. Recent estimates suggest that pneumonia accounted for about 0.32 million deaths in the first month of life and 1.1 million post-neonatal child deaths in 2010—over 14% of all child deaths under five years of age. These estimates, together with an analysis of the rate of fall of cause-specific child mortality, suggest that efforts to reduce child deaths from pneumonia will have to be accelerated if the target of Millennium Development Goal 4—to reduce child mortality by two-thirds between 1990 and 2015—is to be met [3]. Moreover, pneumonia accounts for a substantial percentage of all paediatric out-patient attendances, in-patient admissions, and antibiotic prescriptions in health services in low- and middle-income countries (bacterial pneumonia is the major cause of severe episodes of pneumonia and death from pneumonia) and so places a large burden on these health services and on the families involved [4].

Several effective interventions are available to tackle the challenge of childhood pneumonia, as recently summarised in a series of review articles commissioned as part of the Global Action Plan for the Prevention and Control of Pneumonia [4,5]. Key amongst these interventions are immunisation and correct case management of young children with pneumonia who present to trained health workers. Both of these strategies have been shown to be effective through controlled trials [6,7]. However, although there are robust and accurate mechanisms to measure the performance and progress of immunisation programmes, there are no well-established monitoring methods for (community) case management programmes.

The two major interventions that reduce mortality from bacterial pneumonia are antibiotic treatment (for all cases) and

**Citation:** Campbell H, el Arifeen S, Hazir T, O'Kelly J, Bryce J, et al. (2013) Measuring Coverage in MNCH: Challenges in Monitoring the Proportion of Young Children with Pneumonia Who Receive Antibiotic Treatment. PLoS Med 10(5): e1001421. doi:10.1371/journal.pmed.1001421

**Academic Editor:** David Osrin, Wellcome Trust Senior Research Fellow in Clinical Science, UCL Reader in International Child Health, Honorary Consultant, Great Ormond Street Hospital for Children, United Kingdom

**Published** May 7, 2013



**Funding:** This work was conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF, with financial support from The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in study design, data collection and analysis, or preparation of the manuscript. The funders supported the decision to publish.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** DHS, Demographic and Health Surveys; MICS, Multiple Indicator Cluster Surveys

* E-mail: Harry.Campbell@ed.ac.uk

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

oxygen therapy (in those children who have hypoxaemia). Thus, programme monitoring of case management programmes at (sub-) national, regional, and global levels requires the accurate measurement of the percentage of children with pneumonia in a defined population who receive antibiotic therapy.

Unfortunately, caregivers of sick children with pneumonia often do not seek antibiotic treatment from trained health providers. Findings from eight studies from seven low- and middle-income countries that interviewed caregivers of young children with pneumonia suggest that caregivers sought care from an appropriate health provider who could give correct antibiotic treatment in only about half of pneumonia episodes [8–15]. These findings highlight the need for community-based rather than hospital-based studies to measure antibiotic treatment rates in children with pneumonia if a true population-based estimate is to be made (rather than a biased estimate based on the unrepresentative subgroup of children whose caregivers seek care). Furthermore, the availability of first and second line antibiotics within first level health facilities in low- and middle-income countries may be poor and may vary markedly across institutions, again suggesting that a population-based survey is required.

Programme monitoring therefore needs to identify a representative group of children who have recently had an episode of pneumonia and then investigate what proportion of these children received (correct) antibiotic treatment. In recent years, the Demographic and Health Surveys (DHS) and the United Nations Children's Fund Multiple Indicator Cluster Surveys (MICS) have undertaken these processes in low- and middle-income countries. Specifically, these surveys ask caregivers to recall symptoms and signs of pneumonia in their children, which provides the denominator of pneumonia cases, and collect data from caregivers on antibiotic treatments received, which allows the antibiotic treatment rate to be estimated. In this review, which is part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, we consider how our current understanding of the epidemiology of childhood pneumonia can inform the design and interpretation of surveys seeking to monitor antibiotic treatment rates.

## Disease Prevalence and Survey Sample Size

It is reasonable to assume that the data used to monitor antibiotic treatment rates will not be widely available from cohort studies but will be measured in cross-sectional surveys. The power of such surveys can be roughly estimated by assuming that there is no seasonality of pneumonia incidence and that the duration of pneumonia symptoms is one week on average. A cross-sectional survey of 1,000 young children that enquires about symptoms of pneumonia that were present in the past two weeks should identify approximately 18 caregivers whose child truly had an episode of pneumonia at some point within this recall period. This two-week period prevalence is calculated from a summary estimate of pneumonia incidence of 300 cases per 1,000 children per year in low- and middle-income countries based on a systematic review of population-based cohort studies using standard case definitions that are consistent with the World Health Organization Integrated Management of Childhood Illness case definition of pneumonia [16]. This summary estimate equates to about six new episodes per week per 1,000 children. Children with new episodes that arise over a three-week period will have pneumonia symptoms falling within the two-week period if the child is symptomatic for one week.

Efforts can then be made to establish whether the children with pneumonia received appropriate antibiotic treatment. To estimate a treatment rate of 50% with a precision of ±5% (95% confidence interval of 45%–55%) would require 385 children with pneumonia to be surveyed, and this in turn would require a survey sample size of about 20,000 children. Thus, surveys of many thousand caregivers of young children are required to generate a sufficiently large denominator of "children with pneumonia" from which to measure antibiotic treatment rate. The only such surveys that are conducted widely in low- and middle-income countries at this scale are the DHS and MICS surveys. Both types of survey collect information on children with cough accompanied by rapid or difficult breathing that is due to a problem in the chest. DHS refers to these children as having "symptoms of acute respiratory infection"; MICS refers to them as "suspected" cases of pneumonia. Neither of these survey programmes claims that actual cases of pneumonia are measured through the questions included in the surveys.

## How Well Do Reported Symptoms and Signs of Pneumonia Indicate the Presence of True Pneumonia?

We have previously reported that in conditions of low disease prevalence it is very challenging to obtain an accurate estimate of the true number of episodes of a condition from a screening test—in this case eliciting caregiver report of symptoms of pneumonia or "suspected pneumonia" based on clinical signs recognised by the caregiver (as in DHS and MICS surveys) [17]. For simplicity, throughout the rest of this paper, we will refer to these as cases of "suspected pneumonia".

Table 1 shows a schematic distribution of cases of "true pneumonia" (true disease) according to caregiver report of "suspected pneumonia" (reported symptoms) and true disease status. The sensitivity, specificity, positive and negative predictive values of the survey (test characteristics), and disease prevalence can all be calculated from the numbers of children in cells *a–d* of this table (see Box 1). As we will now discuss, plausible values of pneumonia prevalence and of sensitivity and specificity can be inserted into these 2×2 tables, and the findings can be used to better understand the output of DHS and MICS surveys.

Our first example (Table 2) considers a scenario where 1,000 caregivers are surveyed, the test sensitivity is 80%, and the test specificity is 85%. This corresponds to roughly the level of sensitivity that has been reported for a trained health worker assessment of pneumonia at a health centre based on clinical signs

**Table 1.** Distribution of cases of "true pneumonia" (true disease) according to caregiver report of "suspected pneumonia" (reported symptoms) and true disease status.

| Reported Symptoms | True Disease | | |
|---|---|---|---|
| | Present | Absent | Total |
| Present | *a* | *b* | |
| Absent | *c* | *d* | |
| Total | | | |

Cell *a* represents children with "true pneumonia" (true positives) whose caregiver gave a report of "suspected pneumonia" (test positive). Cell *b* represents children without pneumonia (true negatives) whose caregiver gave a report of "suspected pneumonia" (test positive). Cell *c* represents children with "true pneumonia" (true positives) whose caregiver did not give a report of "suspected pneumonia" (test negative). Cell *d* represents children without pneumonia (true negatives) whose caregiver did not give a report of "suspected pneumonia" (test negative).
doi:10.1371/journal.pmed.1001421.t001

## Box 1. Calculation of Test Characteristics

If we consider the caregiver report of reported symptoms and signs of pneumonia (referred to below as "suspected pneumonia") as defined by DHS and MICS survey guidelines to be a test of true pneumonia status in the child, then:

- The sensitivity of the caregiver report of "suspected pneumonia" (test) is given by $a/(a+c)$;
- The specificity of the caregiver report of "suspected pneumonia" (test) is given by $d/(b+d)$;
- The positive predictive value of the caregiver report of "suspected pneumonia" (test) is given $a/(a+b)$;
- The negative predictive value of the caregiver report of "suspected pneumonia" (test) is given by $d/(c+d)$; and
- The disease prevalence is given by $(a+c)/(a+b+c+d)$.

but assumes a higher level of specificity than is usually achieved in these circumstances. It is unlikely that the performance of caregiver report of "suspected pneumonia" will reach these levels of sensitivity and specificity, and this scenario is therefore likely to give an over-optimistic picture of the ability of surveys based on caregiver report to discriminate true pneumonia. Table 2 illustrates that in this scenario there would be 161 reports of "suspected pneumonia" and 18 cases of true pneumonia among the children of the 1,000 caregivers surveyed—a ratio of reported "suspected pneumonia" to true pneumonia of 8.9:1 (161/18). Thus, "suspected pneumonia" in this setting is a very inaccurate measure of true pneumonia frequency. Furthermore, of the 161 cases of "suspected pneumonia" for which the caregiver would be questioned about receipt of antibiotic treatment, only 14/161 (8.7%, the positive predictive value of the survey tool) would have true pneumonia, making "suspected pneumonia" a very unreliable denominator from which to calculate antibiotic treatment rate. If all 14 children with true pneumonia correctly received antibiotic treatment and all the cases without true pneumonia correctly did not receive antibiotic treatment, then this would be recorded as an antibiotic treatment rate of about 9%, and programme efforts to substantially increase this rate would only serve to promote the over-prescription of antibiotics.

It is possible that, due to a different "case mix" (a different proportion of completely well children) at the household and at the health centre level, the specificities of "suspected pneumonia" in household surveys could be higher than those based on clinical signs recorded by trained health workers at the health centre. Thus, in Tables 3 and 4, we illustrate examples with higher test specificities. As the test specificity rises to very high levels (95% and 99%, respectively), the ratio of reported "suspected pneumonia" to true pneumonia falls to 3.5:1 (63/18) and 1.3:1 (24/18), respectively—still overestimates of pneumonia prevalence but much less inaccurate. In addition, as the specificity increases, the proportion of "suspected pneumonia" that is true pneumonia (the positive predictive value of the tool) rises to 22% (14/63) and 58% (14/24), respectively. Thus, maximising test specificity has the potential to make large improvements to the validity of the denominator that is used to measure pneumonia treatment rates. An increase in specificity from 80% to 99% increases the proportion of "suspected pneumonia" that is truly pneumonia (the positive predictive value) from less than 9% (Table 2) to approximately 58% (Table 4).

By contrast, decreasing test sensitivity from 80% to 60% with a fixed specificity of 95% has only a modest impact on the ratio of reported "suspected pneumonia" to true pneumonia, which falls only slightly from 3.5:1 (63/18) (Table 3) to 3.3:1 (60/18) (Table 5). Moreover, the positive predictive value falls only slightly from 22% (4/63) (Table 3) to 18% (11/60) (Table 5). Since increasing test specificity is usually linked to falling test sensitivity, it is clear that maximising test specificity should be prioritised in the test design. We will briefly discuss how test specificity might be maximised at the end of this review.

There are actually very few published reports of the sensitivity and specificity of caregiver reports of symptoms and signs of pneumonia for the discrimination of true pneumonia. However, using the mean estimate of sensitivity (31%) and specificity (91%) of caregiver report of fast or difficult breathing for prediction of true pneumonia (diagnosed by a study physician) from two community-based studies in Gambia [18] and Pakistan [19] would yield 94 reported cases of "suspected pneumonia" among 1,000 children. This represents a ratio of "suspected pneumonia" to true pneumonia of 5.2:1 (94/18) and a positive predictive value of only 6.4% (6/94). As discussed elsewhere in this Collection, studies undertaken in Pakistan and Bangladesh provide additional data from both hospital-based and community-based studies to further assess the important issue of detection of true pneumonia [20].

Importantly, as noted earlier, the data on the sensitivity and specificity of pneumonia reporting that is needed to determine the prevalence of true pneumonia should ideally come from community-based studies because hospital-based clinic studies may overestimate sensitivity and underestimate specificity due to the different case mix at the hospital level (pneumonia cases tend to be more severe on average in hospitals than in the community, and non-pneumonia cases tend to be less healthy and more likely to have other causes of difficulty breathing). Moreover, because the discriminative power of caregiver reports depends largely on caregiver recognition of symptoms and signs of pneumonia, it is also likely to be influenced greatly by contextual factors such as the

**Table 2.** Distribution of cases of "true pneumonia" according to caregiver report of "suspected pneumonia" (test) and true disease status when test sensitivity is 80% and specificity is 85%.

| Reported Symptoms | True Disease | | |
|---|---|---|---|
| | Present | Absent | Total |
| Present | 14 | 147 | 161 |
| Absent | 4 | 835 | 839 |
| Total | 18 | 982 | 1,000 |

doi:10.1371/journal.pmed.1001421.t002

**Table 3.** Distribution of cases of "true pneumonia" according to caregiver report of "suspected pneumonia" (test) and true disease status when test sensitivity is 80% and specificity is 95%.

| Reported Symptoms | True Disease | | |
|---|---|---|---|
| | Present | Absent | Total |
| Present | 14 | 49 | 63 |
| Absent | 4 | 933 | 937 |
| Total | 18 | 982 | 1,000 |

doi:10.1371/journal.pmed.1001421.t003

**Table 4.** Distribution of cases of "true pneumonia" according to caregiver report of "suspected pneumonia" (test) and true disease status when test sensitivity is 80% and specificity is 99%.

| Reported Symptoms | True Disease | | |
|---|---|---|---|
| | Present | Absent | Total |
| Present | 14 | 10 | 24 |
| Absent | 4 | 972 | 976 |
| Total | 18 | 982 | 1,000 |

doi:10.1371/journal.pmed.1001421.t004

level of maternal education and prior exposure to relevant health education messages. Thus, the interpretation of trends in antibiotic treatment data over time will be complicated in settings where there have been temporal trends in education levels or health education interventions or where serial surveys have been conducted in different populations.

## Another Strategy for Improving the Positive Predictive Value of Surveys

In addition to increasing test specificity, another strategy that should improve the proportion of "suspected pneumonia" reports that truly represent cases of pneumonia (positive predictive value) is to increase the disease prevalence detected in DHS and MICS surveys. This could be achieved by conducting surveys during the peak pneumonia season. Alternatively, if the recall period were to be increased from two weeks to four weeks (or eight weeks), then 30 (or 54) reports of pneumonia symptoms and signs would be expected, rather than 18 reports. Tables 6 and 7 show that with the same levels of sensitivity (80%) and specificity (95%) as in the example in Table 3, these longer recall periods would yield 72 (or 90) reported cases of "suspected pneumonia" and 30 (or 54) cases of true pneumonia among 1,000 children. Thus, the ratio of "suspected pneumonia" to true pneumonia would be about 3:1 (or 1.7:1) (compared to 3.5:1 in Table 3), and the proportion of "suspected pneumonia" that is true pneumonia would be 33% (or 48%) rather than 22% as in Table 3.

If we use the data from Gambia [18] and Pakistan [19] on test specificity and sensitivity combined with longer recall periods, the proportion of "suspected pneumonia that is true pneumonia (positive predictive value) would rise from 6.4% to 9.8% (or 16.7%) based on four-week (or eight-week) recall, respectively. Because these predictions are based on the assumption that test performance would not change with the longer recall period, it will be important to test this assumption (see [20]) before any such increase in recall period is introduced into surveys.

**Table 5.** Distribution of cases of "true pneumonia" according to caregiver report of "suspected pneumonia" (test) and true disease status when test sensitivity is 60% and specificity is 95%.

| Reported Symptoms | True Disease | | |
|---|---|---|---|
| | Present | Absent | Total |
| Present | 11 | 49 | 60 |
| Absent | 7 | 933 | 940 |
| Total | 18 | 982 | 1,000 |

doi:10.1371/journal.pmed.1001421.t005

**Table 6.** Distribution of cases of "true pneumonia" according to caregiver report of "suspected pneumonia" (test) and true disease status when test sensitivity is 80% and specificity is 95% with a four-week recall period.

| Reported Symptoms | True Disease | | |
|---|---|---|---|
| | Present | Absent | Total |
| Present | 24 | 48 | 72 |
| Absent | 6 | 922 | 928 |
| Total | 30 | 970 | 1,000 |

doi:10.1371/journal.pmed.1001421.t006

## A Combined Strategy to Improve the Identification of True Pneumonia Episodes from Maternal Report of "Suspected Pneumonia"

Finally, we can consider the effect of combining an improved test specificity and an increased recall period. Using the published Gambia [18] and Pakistan [19] data, if study recall could be increased from two to four or eight weeks and test specificity increased from 91% to 97%, then the proportion of "suspected pneumonia" that is true pneumonia (the positive predictive value) would rise by almost an order of magnitude from 6.4% to 23.7% or 56.9% for four- or eight-week recall, respectively. The field studies in Pakistan and Bangladesh [20] provide further data from which to estimate the improvements that could be expected from attempts to improve survey instruments.

## Are DHS and MICS Surveys Suitable Tools for Monitoring Antibiotic Treatment of Childhood Pneumonia?

A research priority setting exercise that involved a large number of doctors from low- and middle-income countries recently listed improving the community case management of pneumonia and identifying barriers to and improving access to antibiotic treatment as among the top research priorities likely to contribute to achievement of Millennium Development Goal 4 [21]. The identification of these research priorities reflects the need to reduce childhood pneumonia if childhood mortality is going to be reduced, and supports the self-evident need to develop a robust

**Table 7.** Distribution of cases of "true pneumonia" according to caregiver report of "suspected pneumonia" (test) and true disease status when test sensitivity is 80% and specificity is 95% with an eight-week recall period.

| Reported Symptoms | True Disease | | |
|---|---|---|---|
| | Present | Absent | Total |
| Present | 43 | 47 | 90 |
| Absent | 11 | 899 | 910 |
| Total | 54 | 946 | 1,000 |

doi:10.1371/journal.pmed.1001421.t007

## Key Points

- Large household surveys are required to identify recent cases of pneumonia as a denominator from which antibiotic treatment rates for pneumonia can be estimated.
- At the low levels of pneumonia prevalence found in household surveys, most of the children identified with "suspected pneumonia" will not have true pneumonia, even when survey tools with very high sensitivity and specificity are used.
- This inflation of the denominator of the antibiotic treatment rate will make the treatment rate appear falsely low and could lead to incorrect programme decision making.
- In theory, the performance of DHS/MICS survey tools can be improved by increasing test specificity and/or by increasing pneumonia period prevalence (by increasing the recall period or by conducting the survey in the peak pneumonia season), but this prediction needs testing.
- Alternate approaches to measuring the antibiotic treatment rate should also be considered, including those that make use of digital formats to facilitate pneumonia recognition and recall of antibiotic treatment by caregivers.

programme indicator—the antibiotic treatment rate—to monitor local, national, and global progress in increasing the coverage of this essential pneumonia intervention. Measurement of this indicator has to be community-based (to capture the many pneumonia cases that do not attend health services for treatment) and large (to include enough pneumonia cases to give precise estimates). DHS and MICS surveys are the only tools that fit these requirements and that are conducted widely in developing countries. It is therefore important to assess their suitability for this purpose, since estimates of pneumonia prevalence and antibiotic treatment coverage based on these surveys will be influential in guiding national and international decisions about programmes to control pneumonia deaths.

### The Validity of DHS and MICS "Suspected Pneumonia" Survey Data for the Estimation of the Prevalence of True Pneumonia

In circumstances of low pneumonia prevalence (such as found with the two-week recall of pneumonia episodes included in current surveys), even when the sensitivity and specificity of "suspected pneumonia" as a test for true pneumonia is very high (>90%), the estimates of the number of pneumonia episodes based on DHS and MICS survey data will be greatly inflated, and most reported episodes will be false positives [17]. DHS and MICS survey questions were not designed to estimate pneumonia prevalence, and current DHS and MICS guidelines advise against the use of data in this way. Our review reinforces these recommendations.

### The Validity of DHS and MICS "Suspected Pneumonia" Survey Data for the Estimation of the Antibiotic Treatment Rate for Pneumonia

The findings we present in this review (Tables 1–7) have important implications for the use of existing DHS and MICS survey data in monitoring antibiotic treatment rates. The DHS/MICS antibiotic treatment indicator for "suspected pneumonia" is often used as a proxy indicator for a pneumonia treatment indicator. For this indicator to give accurate information that is useful for programme planning, we need to have a denominator as close to true pneumonia as possible. A denominator of "suspected pneumonia" in which most cases are not true pneumonia makes interpretation of this indicator problematic, and action based on adoption of this indicator alone could drive over-prescription of antibiotics.

It is clear from the examples we work through in this review and from underlying theory based on known epidemiology of pneumonia [17] that the discriminative power of survey instruments needs to be improved. This discriminative power is constrained by the inherent limited ability of caregivers to correctly recognise and report symptoms and signs of pneumonia. Within this constraint, it is possible to increase test specificity (at the expense of test sensitivity) by adding a few additional symptoms or signs to survey questionnaires that show good predictive power or by employing a "pneumonia score", as reported elsewhere in this Collection [20]. In this latter approach, a survey would include questions about a series of signs or symptoms of pneumonia, and one mark would be awarded for each sign or symptom that is reported. A threshold score level defined by a favourable combination of sensitivity and specificity levels could then be selected for use. However, it is possible that the underlying discriminative power of this approach will remain constrained since it relies on caregiver recognition. One way to overcome this constraint and to improve overall discriminative power might be to adopt a new approach for questioning caregivers that relies on video recognition [20]. Such an approach would operate via visual recognition memory rather than via auditory recognition memory, which may promote better recognition and recall [22]. This argument could also be applied to the measurement of the antibiotic treatment rate indicator. Thus, "pill boards" or digital formats could be shown to caregivers that illustrate a range of local drugs, to promote correct recall of antibiotic prescriptions.

It could be argued that although the interpretation of the absolute value of the antibiotic treatment rate indicator is very problematic, there may still be considerable utility in using its relative value to track trends over time (and to compare across countries) for programme planning purposes. However, these data should be interpreted with caution, since contextual factors are likely to influence results significantly, and these may vary by time and place independent of trends in antibiotic treatment rates.

### The Validity of DHS and MICS "Suspected Pneumonia" Survey Data for the Estimation of Appropriate Care Seeking for Pneumonia

The care-seeking indicator for "suspected pneumonia" that is found in DHS/MICS surveys remains valid. "Suspected pneumonia" based on simple signs that caretakers can understand and that programmes can use is an appropriate denominator for this indicator, as the aim is to encourage all these children to be assessed by a health provider whether or not they actually have pneumonia.

## Future Research and Prospects

Given the importance of antibiotic treatment rates as a programme indicator, there is an urgent need for more research to measure the sensitivity and specificity of "suspected pneumonia" as defined in DHS and MICS surveys for the identification of true pneumonia episodes. The sensitivity and specificity of "suspected pneumonia" needs to be measured in a range of

settings and with questions based on a range of recall periods from two weeks to several months, and must also be compared to the performance of new approaches such as those described above. Hazir et al. provide some first estimates of these parameters based on two- and four-week recall and estimate the impact on test performance of some new survey methods [20]. Consideration should also be given to the feasibility of including in future surveys some assessment of whether the prescribed antibiotic was actually taken correctly by the child.

In addition to optimising existing means of determining antibiotic treatment rates for pneumonia, recent developments in eHealth and mHealth (health care supported by electronic processes and communication and by mobile devices, respectively) applications in low- and middle-income countries and their use, for example, in surveillance of influenza episodes [23] may mean that novel real-time measurement of child health programme indicators will soon be feasible in some settings. Finally, in the short term, digital illustrations of local treatments or of children with signs of pneumonia that are recognised by local caregivers should be technically feasible and could facilitate accurate data capture, storage, and transmission for analysis [24], thereby helping to improve the way we monitor antibiotic treatment coverage among young children with pneumonia.

## Acknowledgments

The authors would like to thank the Child Health Epidemiology Reference Group Working Group on Improving Coverage Measurement for comments and feedback during the preparation of this review.

## Author Contributions

Analyzed the data: HC SA TH JO JB IR SQ. Wrote the first draft of the manuscript: HC SA TH JO JB IR SQ. Contributed to the writing of the manuscript: HC SA TH JO JB IR SQ. ICMJE criteria for authorship read and met: HC SA TH JO JB IR SQ. Agree with manuscript results and conclusions: HC SA TH JO JB IR SQ.

## References

1. Black RE, Cousens S, Johnson H, Lawn J, Rudan I, et al. (2010) Global, regional and national causes of child mortality, 2008. Lancet 375: 1969–1987.
2. Theodoratou E, Zhang JSF, Kolcic I, Bhopal S, Nair H, et al. (2011) Estimating the global, regional and national burden of childhood pneumonia deaths in 2008. PLoS ONE 6: e25095. doi:10.1371/journal.pone.0025095
3. Liu L, Johnson H, Perin J, Cousens S, Scott S, et al. (2012) Global, regional, and national causes of child mortality in 2000–2010: an updated systematic analysis. Lancet 379: 2151–2161.
4. Rudan I, Boschi-Pinto C, Biloglav Z, Mulholland K, Campbell H (2008) The epidemiology and etiology of childhood pneumonia. Bull World Health Organ 86: 408–416.
5. Greenwood BM (2008) A global action plan for the prevention and control of pneumonia. Bull World Health Organ 86: 322–322A.
6. Theodoratou E, Johnson S, Jhass A, Madhi SA, Clark A, et al. (2010) The effect of Haemophilus influenzae type b and pneumococcal conjugate vaccines on childhood pneumonia incidence, severe morbidity and mortality. Int J Epidemiol 39 (Suppl 1): i172–i185.
7. Theodoratou E, Al-Jilaihaw S, Woodward F, Ferguson J, Jhass A, et al. (2010) The effect of case management on childhood pneumonia mortality in developing countries. Int J Epidemiol 39 (Suppl 1): i155–i171.
8. Tupasi TE, Miguel CA, Tallo VL, Bagasao TM, Natividad JN, et al. (1989) Child care practices of mothers: implications for intervention in acute respiratory infections. Ann Trop Paediatr 9: 82–88.
9. Amarasiri de Silva MW, Wijekoon A, Hornik R, Martines J (2001) Care seeking in Sri Lanka: one possible explanation for low childhood mortality. Soc Sci Med 53: 1363–1372.
10. Nagaraj A, Nagaraj VK, Lobo J (2004) Perception of severity of acute respiratory tract infections and diarrhoea in children under the age of five in rural India: utilization pattern of health services. Trop Doct 34: 222–223.
11. Mishra S, Kumar H, Sharma D (1994) How do mothers recognize and treat pneumonia at home? Indian Pediatr 31: 15–18.
12. Luque JS, Whiteford LM, Tobin GA (2008) Maternal recognition and health care-seeking behavior for acute respiratory infection in children in a rural Ecuadorian county. Matern Child Health J 12: 287–297.
13. Jordan HT, Prapasiri P, Areerat P, Anand S, Clague B, et al. (2009) A comparison of population-based pneumonia surveillance and health-seeking behavior in two provinces in rural Thailand. Int J Infect Dis 13: 355–361.
14. Goldman N, Pebley AR, Gragnolati M (2002) Choices about treatment for ARI and diarrhoea in rural Guatemala. Soc Sci Med 55: 1693–1712.
15. Kristiansson C, Reilly M, Gotuzzo E, Rodriguez H (2008) Antibiotic use and health seeking behaviour in an underprivileged area of Peru. Trop Med Int Health 13: 434–442.
16. Rudan I, Tomaskovic L, Boschi-Pinto C, Campbell H, WHO Child Health Epidemiology Reference Group (2004) Global estimate of the incidence of clinical pneumonia among children under five years of age. Bull World Health Organ 82: 895–903.
17. Campbell H, Biloglav Z, Rudan I (2008) Reducing bias from test misclassification in burden of disease studies: use of test to actual positive ratio—new test parameter. Croat Med J 49: 402–414.
18. Campbell H, Byass P, Greenwood BM (1990) Acute lower respiratory infections in Gambian children: maternal perception of illness. Ann Trop Paediatr 10: 45–51.
19. Mull DS, Mull JD, Kundi MZ, Anjum M (1994) Mothers' perceptions of severe pneumonia in their own children: a controlled study in Pakistan. Soc Sci Med 38: 973–987.
20. Hazir T, Bequem K, el Arifeen S, Khan AM, Huque MH, et al. (2013) Measuring coverage in MNCH: A prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia. PLoS Med 10: e1001422. doi:10.1371/journal.pmed.1001422
21. Rudan I, El Arifeen S, Bhutta ZA, Black RE, Brooks A, et al. (2011) Setting research priorities to reduce global mortality from childhood pneumonia by 2015. PLoS Med 8: e1001099. doi:10.1371/journal.pmed.1001099
22. Cohen MA, Horowitz TS, Wolfe JM (2009) Auditory recognition memory is inferior to visual recognition memory. Proc Natl Acad Sci U S A 106: 6008–6010.
23. Rajatonirina S, Heraud JM, Randrianasolo L, Orelle A, Razanajatovo NH, et al. (2012) Short message service sentinel surveillance of influenza-like illness in Madagascar, 2008–2012. Bull World Health Organ 90: 385–389.
24. Brock TP, Smith SR (2007) Using digital videos displayed on personal digital assistants (PDAs) to enhance patient education in clinical settings. Int J Med Inform 76: 829–835.

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: Design, Implementation, and Interpretation Challenges Associated with Tracking Vaccination Coverage Using Household Surveys

**Felicity T. Cutts[1]*, Hector S. Izurieta[2], Dale A. Rhoda[3]**

**1** Independent consultant, La Londe les Maures, France, **2** Food and Drug Administration, Rockville, Maryland, United States of America, **3** Battelle Memorial Institute, Columbus, Ohio, United States of America

**Abstract:** Vaccination coverage is an important public health indicator that is measured using administrative reports and/or surveys. The measurement of vaccination coverage in low- and middle-income countries using surveys is susceptible to numerous challenges. These challenges include selection bias and information bias, which cannot be solved by increasing the sample size, and the precision of the coverage estimate, which is determined by the survey sample size and sampling method. Selection bias can result from an inaccurate sampling frame or inappropriate field procedures and, since populations likely to be missed in a vaccination coverage survey are also likely to be missed by vaccination teams, most often inflates coverage estimates. Importantly, the large multi-purpose household surveys that are often used to measure vaccination coverage have invested substantial effort to reduce selection bias. Information bias occurs when a child's vaccination status is misclassified due to mistakes on his or her vaccination record, in data transcription, in the way survey questions are presented, or in the guardian's recall of vaccination for children without a written record. There has been substantial reliance on the guardian's recall in recent surveys, and, worryingly, information bias may become more likely in the future as immunization schedules become more complex and variable. Finally, some surveys assess immunity directly using serological assays. Sero-surveys are important for assessing public health risk, but currently are unable to validate coverage estimates directly. To improve vaccination coverage estimates based on surveys, we recommend that recording tools and practices should be improved and that surveys should incorporate best practices for design, implementation, and analysis.

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection.*

## Introduction

The percentage of a population that has been vaccinated—vaccination coverage—is an imperfect but helpful measure of the effectiveness of vaccination programs and of public health more broadly [1]. Vaccination coverage is a tracer condition for results-based financing [2], an indicator of eligibility for Millennium Challenge Account assistance [3], and a criterion for support from the GAVI Alliance for the introduction of new vaccines [4]. Making funding decisions contingent on coverage potentially incentivizes inflation of coverage figures, and there is wide recognition of the need to improve the data [5–8].

Ideally, vaccination coverage should be monitored continuously using registries or administrative reports [9]. Electronic immunization registries aim to document all vaccinations of each individual in each birth cohort [10,11]. Denominators may derive from the same registry [11] or from a separate vital statistics system. When well implemented, electronic immunization registries can provide data for coverage measurement and for program management activities such as monitoring vaccine supply and requisitions and sending vaccination reminders. However, challenges facing such registries include accounting for migration within and between countries, ensuring complete birth registration and vaccination reporting, avoiding record duplication [12], and ensuring continuity after organizational changes [13]. Although pilot studies of electronic registries are ongoing in low- and middle-income countries including Albania, Guatemala, India, and Viet Nam [14], these challenges currently limit their use. Therefore, in most low- and middle-income countries, "administrative coverage" is calculated using aggregate reported data on the number of doses of each vaccine administered to children in the target age group in a given time period and target population estimates from censuses [1]. Health workers at each health facility typically compile data manually each month from clinic records such as

**Citation:** Cutts FT, Izurieta H, Rhoda DA (2013) Measuring Coverage in MNCH: Design, Implementation, and Interpretation Challenges Associated with Tracking Vaccination Coverage Using Household Surveys. PLoS Med 10(5): e1001404. doi:10.1371/journal.pmed.1001404

**Academic Editor:** David Osrin, Wellcome Trust Senior Research Fellow in Clinical Science, UCL Reader in International Child Health, Honorary Consultant, Great Ormond Street Hospital for Children, United Kingdom

**Published** May 7, 2013



**Funding:** FTC and DAR were contracted by The Bill & Melinda Gates Foundation to undertake this work. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed in this paper are those of the individual authors and not necessarily those of the US Food and Drug Administration.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** BCG, Bacille Calmette Guerin vaccine against tuberculosis; DHS, Demographic and Health Survey/s; EPI, Expanded Programme on Immunization; HBR, home-based record; LQAS, Lot Quality Assurance Sampling; MICS, Multiple Indicator Cluster Survey/s; UNICEF, United Nations Children's Fund; WHO, World Health Organization.

* E-mail: felicity.cutts@lshtm.ac.uk

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

immunization registers or tally sheets. At each vaccination visit, the health worker records vaccinations on clinic records and on a vaccination card, child health card, or other home-based record (HBR) that the mother keeps. The HBR serves as an educational tool for the mother and is also an important data source in household surveys. In many countries, however, the quality of primary recording of vaccinations, of transcription and compilation of data, and of reporting is low, and numerators may be either inflated (e.g., because doses outside the recommended age range are included) or too low (e.g., if private practitioners do not report). Moreover, denominators are often grossly inaccurate [5,7]. Hence, wherever possible other data sources such as surveys are still considered in the World Health Organization (WHO)–United Nations Children's Fund (UNICEF) estimates of national immunization coverage [15].

Given current problems with coverage estimates based on administrative reports in many countries, we believe that surveys will continue to provide important information in the short-to-medium term, at national and sub-national levels. It is therefore critical that surveys are conducted rigorously. In this review, which is part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, we discuss the survey methods used to estimate vaccination coverage in low- and middle-income countries, highlight potential pitfalls, and propose strategies to improve coverage measurement. Our review aims to inform public health practitioners and the researchers who design and implement surveys as well as Ministry of Health officials and donors who interpret and use data from surveys.

## Survey Methods Used to Measure Vaccination Coverage

Four types of surveys are commonly employed to estimate vaccination coverage (Table 1). The Demographic and Health Surveys (DHS) [16] and Multiple Indicator Cluster Surveys (MICS) [17] are probability sample surveys, in which each household has a known and nonzero probability of being selected in the sample. There have been about 10–15 DHS and 20 MICS per year since 1995. These large, important, and generally well-conducted household surveys, which are used to collect data about many aspects of health, are described in detail in a companion paper in this Collection [18].

The Expanded Programme on Immunization (EPI) cluster survey was developed by the WHO and was described in 1982 as a practical tool to quickly estimate coverage to within $\pm 10$ percentage points of the point estimate [19]. The original EPI survey method selects 30 clusters from which seven children in each cluster are selected using the "random start, systematic search" method. Specifically, a starting dwelling is chosen by starting at a central location in the village or town, selecting a direction at random, counting the dwellings lying in that direction up to the edge of the village, and selecting one of them randomly;

**Table 1.** Characteristics of common surveys used to measure vaccination.

| Survey Characteristic | DHS | MICS | EPI | LQAS |
|---|---|---|---|---|
| **Primary objectives** | Collection of information on a wide range of population, health, and nutrition topics, plus additional optional modules | Collection of information on population health, child protection, and child development | Estimation of vaccination coverage | Classification of lots (catchment areas) into two groups: those with adequate coverage and those with inadequate coverage |
| **Sampling scheme** | Stratified cluster sampling; clusters selected using PPES; clusters are usually census enumeration areas | Stratified cluster sampling; clusters selected using PPES; clusters are usually census enumeration areas | Cluster sampling with or without stratification; clusters are usually villages or urban neighborhoods, selected using PPES | Classic method uses simple random sampling within a lot; when lots are large, cluster sampling is sometimes employed |
| **Household selection** | Household selected randomly based on a complete household listing and mapping in the sample clusters | Current practice is random selection of households based on a complete listing and mapping of enumeration areas | Varies; usually non-probability; the first household is selected randomly, then neighboring households are selected until seven children can be enrolled | When cluster sampling is used, the first household is selected randomly before moving in a consistent direction, sampling every *k*th household |
| **Total sample size** | Based on desired precision for key indicators at the regional level; the number of children aged 12–23 months covered in recent surveys is typically around 1,800 at the national level | Based on desired precision of key indicators selected by implementing agencies; usually >2,000 women and several hundred children aged 12–23 months | Usually 30 clusters of seven children aged 12–23 months; sized to yield estimate of $\pm 10\%$ assuming design effect of two | Varies greatly; 19 respondents per lot is a common size with simple random sampling; 50 or 60 is common when using cluster sampling |
| **Respondents** | All men and women aged 15–49 years; vaccination data on children <5 years if biological mother is interviewed, and on women of childbearing age | All women aged 15–49 years; vaccination data on children <5 years if primary caretaker is interviewed, and on women of childbearing age | Mother or primary caretaker of children aged 12–23 months | Varies; field workers interview caretaker and when possible substantiate response with vaccination record or sometimes indelible ink finger mark on child |
| **Questionnaire length** | Household: 25 pages; woman's questionnaire: about 70 pages | Household: 18 pages; woman's: 38 pages; children under 5 years: 18 pages | 1–2 pages | Often 1 page |
| **Implementers** | Usually National Statistical Office or equivalent, with capacity-building from MEASURE DHS | Usually National Statistical Office, with support from UNICEF and other partners | Varies; often national- or district- level Ministry of Health employees | Varies; usually independent from vaccination team |
| **Duration** | 12 months or more to plan, implement, analyze, and report | 12 months or more to plan, implement, analyze, and report | Several months to plan; weeks to implement, analyze, and report | Varies; 1–2 days per lot to implement and analyze |

PPES, probability proportional to estimated population size.
doi:10.1371/journal.pmed.1001404.t001

adjacent households are then visited until seven children aged 12–23 months have been enrolled [20,21]. The central starting location may bias the method to include households with good access to vaccination, so it is difficult to assign unbiased probabilities of selection to the households using this method, which does not meet the above criteria for a probability sample and is, therefore, a "non-probability sampling" survey method [22]. EPI surveys are widely used at national and sub-national levels, but there is no central database of results, so the total number of surveys conducted is unknown. Adaptations of the EPI survey have incorporated probability sampling at the final stage of sample selection [22–26], and the updated WHO guidelines [21] as well as a recent companion manual on hepatitis B immunization surveys emphasize the need for probability sampling for scientifically robust estimates of coverage [27].

The main design differences between EPI surveys (if probability sampling is used) and DHS or MICS surveys is that EPI surveys focus specifically on vaccination data while DHS and MICS surveys cover a wide range of population and health topics and include a much larger sample size. In addition, field implementation of EPI surveys is variable and often done without external technical assistance, while the DHS and MICS are highly standardized and have substantial technical assistance and quality control.

A final household survey method commonly used to estimate health intervention coverage in low- and middle-income countries is Lot Quality Assurance Sampling (LQAS). LQAS surveys use a stratified sampling approach to classify "lots," which might be districts, health units, or catchment areas, as having either "adequate" or "inadequate" coverage of various public health interventions. For vaccination coverage measurement, LQAS is "nested" within a cluster survey to evaluate neonatal tetanus elimination [28], coverage of yellow fever vaccination [29], and coverage of meningococcal vaccine campaigns [30], and to monitor polio vaccination coverage after supplementary immunization activities [31].

## Survey Design and Implementation

Surveys used to estimate vaccination coverage should have a sample size that results in an acceptable sampling error, and implementation should minimize non-sampling errors, including selection bias and information bias (Table 2). In DHS and MICS surveys, the sample size is determined by the estimated number of households required for the desired precision of key indicators (not vaccination coverage), and all children in the eligible age groups in those households are included. In recent DHS surveys this design has given sample sizes of around 1,800 children aged 12–23 months. EPI surveys traditionally included 210 children aged 12–23 months in 30 clusters, but the sample size and number of clusters should be calculated according to assessments of the likely coverage, intra-cluster correlation, and desired precision of the vaccination coverage estimate [21,27].

Selection bias may occur due to use of an outdated or nonrepresentative sampling frame, use of non-probability sampling, or poor field worker practices such as substituting a selected household with one that is easier to reach. The "random start, systematic search" method used in traditional EPI surveys has intrinsic geographic bias. It allows field workers to select households rather than this being part of the initial sampling process, does not document reasons for nonparticipation, and cannot adjust for biases resulting from out-of-date size estimates for selection of clusters using probability proportional to estimated size sampling. Moreover, in EPI and LQAS surveys, teams are likely to replace households where no one is home or where eligible respondents refuse to participate. If respondents are not selected randomly and if the same forces that influence participation in the survey also influence participation in vaccination (e.g., families missed by interviewers because they work in the fields all day may also lack time to attend vaccination clinics), replacement is likely to result in bias, probably upwards. Finally, surveys of the vaccination status of living individuals are inherently subject to selection bias since death is more likely in unvaccinated than in vaccinated children. In settings where there is a high infant mortality rate, this bias may be substantial.

There are multiple potential sources of information error and bias in measuring the vaccination status of each child in surveys (Figure 1), many of which also affect data included in administrative reports. Mistakes can occur during primary data recording each time a child attends a vaccination point or when survey interviewers transcribe birth and vaccination dates onto a paper or digital questionnaire. If a paper questionnaire is used, further errors can occur during digital data entry. Data on source documents can also be incomplete [32] or inaccurate [33] and, when new vaccines are introduced, old HBRs may remain in circulation, requiring health workers to improvise in their recording (Figure 2). There is further confusion regarding recording of vaccines administered during campaigns such as "vaccination weeks" on HBRs [34].

When the HBR is not available (it may be lost or locked away, or mothers may not be given enough time to find it), interviewers question the child's parent or guardian to construct a verbal vaccination history. The reliability of such histories may vary with the information received or understood by mothers at the time of vaccination; the interviewer's skills, carefulness, neutral demeanor, and use of appropriate language; the recall period; and the length of the questionnaire and resulting interview fatigue [35–37]. The complexity of the vaccination schedule can also affect the reliability of a verbal vaccination history. When the EPI survey was introduced in the 1980s, the infant EPI schedule comprised five visits, which lent themselves to straightforward questions to the mother (Table 3). Because current schedules are much more complex (Table 4) and vary over time and between countries, constructing a verbal history of vaccinations received is now much more difficult and likely to become increasingly so. Thus, the questions included in surveys need substantial and continuous adaptation.

## Data Analysis and Reporting Issues

Traditionally, surveys report the proportion of persons who have been vaccinated as recorded "by card only" and by "card plus history," both by age 12 months and by age at the time of the survey. EPI surveys also calculate and report separately on coverage of "valid" doses among children with cards, such as a minimum interval of 28 days between doses of diphtheria-tetanus-pertussis-containing vaccines and a minimum age of 270 days for measles vaccination [21]. As coverage increases, evaluation of the timeliness of vaccination among children with documented dates of birth, and of each vaccine dose, provides additional information to guide program performance. Timeliness can be illustrated through graphs of the distribution of age at receipt of each dose compared to the national schedule [32,38] or by time-to-event curves of the cumulative coverage by age [32,39]. The mean number of extra days or weeks that children remain under-vaccinated and at risk of disease [38,40,41] and risk factors for delay in vaccination can be assessed [39,40].

Surveys that use probability proportional to estimated size sampling without stratification assume that each cluster has equal weight in the analysis. EPI surveys do not collect data on the number of eligible households in each cluster, and cannot validate this assumption. Consequently, if outdated or inaccurate sampling

**Table 2.** Main potential sources of error and strategies to minimize them in population-based surveys measuring vaccination coverage.

| Source of Error | Effect of Error on Results | Strategies to Minimize Error |
|---|---|---|
| **Random error** | | |
| Sampling error | Reduces precision | Choose optimum sample design (e.g., number and size of clusters) and adjust sample size to achieve desired precision while retaining budgetary and logistical practicality |
| **Systematic error** | | |
| Selection bias—sampling frame | Depends on size of excluded population and difference in vaccination uptake between those excluded and included | Use most recent census data available |
| | | Assess likelihood of census projections reflecting reality and update census if necessary |
| | | If large populations have been excluded (e.g., security constraints at time of census), consider special efforts to include them |
| Selection bias—sampling procedures | Non-probabilistic sampling may lead to bias in either direction | Use probability sampling method (plan time for listing of households within selected clusters) |
| | | Use appropriate weighting in analysis |
| Selection bias—poor field procedures | Most likely to lead to upward bias in coverage results | Preselect households and ensure strict supervision |
| | | Conduct survey at time of year and of day when people most likely to be available |
| | | Work with communities to enhance survey participation rates |
| | | Conduct revisits as necessary to locate caregivers and HBRs |
| | | Do not substitute households |
| Information bias—lack of HBR or poorly filled HBR | Bias in coverage results may underestimate or overestimate coverage depending on how missing data are handled and how HBRs are read by enumerators | Public health programs need to educate families to retain HBRs and improve primary recording of vaccination data |
| | | Publicize reminders about HBRs prior to survey (e.g., during household listing step) |
| | | Allow time for mothers to look for HBR, revisit if necessary |
| | | Include younger age groups in surveys and measure age-appropriate vaccination coverage |
| | | Include questions as to condition of HBR and checks for errors |
| | | Seek health facility–based records on children without HBR |
| Information bias—inaccurate verbal history | Most likely to bias infant coverage upwards as mothers may feel pressure to say their children have been vaccinated; for tetanus toxoid in adult women, verbal history usually underestimates percent of women protected | Ensure interviewers maintain neutral attitude |
| | | Give time to mothers to respond |
| | | Shorter questionnaires likely to have less interviewee fatigue |
| | | Standardize questions, use visual aids, conduct close supervision |
| | | For tetanus toxoid, ask careful questions about *all* doses received in previous and current pregnancies and in campaigns (but this still does not account for diphtheria-tetanus-pertussis vaccination received in infancy); sero-surveys play a useful role in the measurement of the prevalence of protection |
| Data transcription and data entry errors | May increase data classed as missing; can bias coverage results | Conduct close supervision |
| | | Conduct range and consistency checks; enumerators can revisit household if necessary to correct data |
| Missing data | If nonrandom, biases result, often upwards | Conduct high-quality planning, training, and supervision to reduce missing data |
| | | Include appropriate statistical adjustment for missing data |

doi:10.1371/journal.pmed.1001404.t002

frames are used, the resulting point estimate may be biased. If, however, a household listing step is included in the survey preparation and sampling stages, appropriate weights can be calculated and used to derive national estimates and confidence intervals, as is recommended nowadays by DHS and MICS protocols [18].

The standard error of the coverage estimate is traditionally used to calculate and report a 95% confidence interval around the point estimate. The confidence interval is affected by the sample size, the sampling design, and the underlying proportion itself. Because individuals living in one cluster of a population tend to be more similar to each other than persons from different clusters,

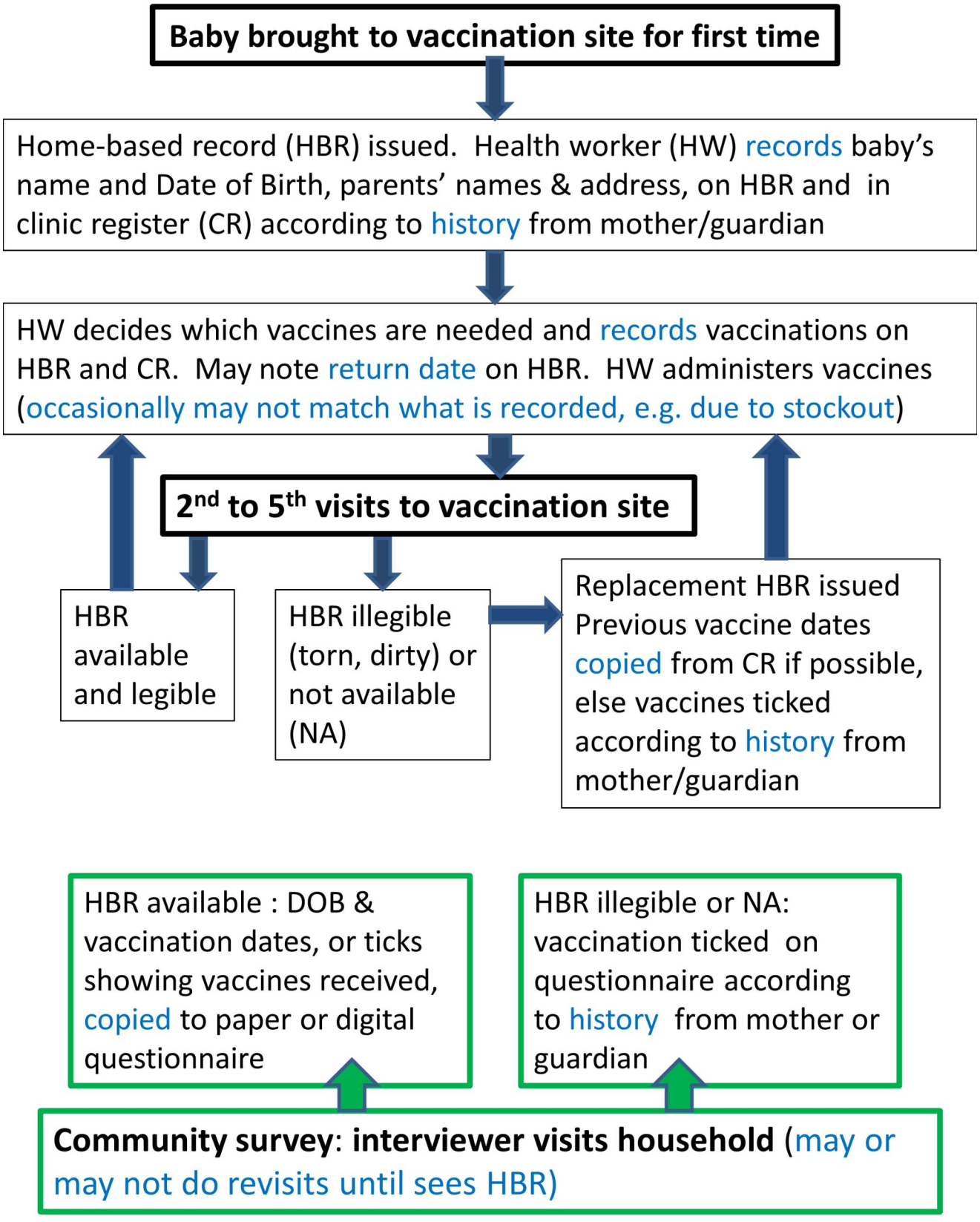


**Figure 1. Schematic of recording of vaccination data at the time of vaccination and during community surveys.** Recording at the time of vaccination (primary recording) is indicated in black boxes; recording during surveys is indicated in green boxes. Main potential sources of information error and bias are highlighted in blue. DOB, date of birth.
doi:10.1371/journal.pmed.1001404.g001


**Figure 2. Several instances of improvisation on a vaccination card.** (Photo courtesy of Carolina Danovaro, Pan American Health Organization.)
doi:10.1371/journal.pmed.1001404.g002

respondents in a cluster sample each contribute less independent information about the overall population than respondents in a simple random sample. This positive intra-cluster correlation causes cluster samples to have a wider confidence interval around the point estimate of the population parameter than a simple random sample of the same size. DHS, MICS, and EPI surveys all provide guidance on estimation of confidence intervals for key indicators, but the degree to which confidence intervals are reported and used varies widely, as discussed elsewhere in this Collection [42].

The application of standard statistical techniques to estimate confidence intervals has been challenged for surveys that use non-probability sampling of households within each cluster [43], although simulations of results from EPI surveys have shown that confidence intervals in these surveys are generally within the desired precision of $\pm 10$ percentage points [44]. Some variations on the EPI survey method take a probability sample (e.g., a systematic random sample in the final stage) [22–26], which makes it possible to calculate sampling weights and construct meaningful confidence intervals.

LQAS surveys inevitably have a central range of coverage (the gray area) that is not excluded by either the "adequate" or the "inadequate" classification. That is, neither classification excludes the medium category. For fixed values of alpha and beta (the probability of type I and II errors, respectively), a larger sample size per lot will result in a narrower gray area and a correspondingly more confident conclusion about whether coverage is likely to be adequate (Figure 3). When data are combined across numerous lots, it is possible to estimate a region-wide proportion and confidence interval using formulae from stratified sampling and applying strata and cluster weights. However, at the level of the individual lot, the user does not obtain a precise coverage estimate from a LQAS survey, but only an assurance that coverage in populations where there is very low coverage is very likely to be classified as inadequate and that coverage in populations where there is very high coverage is very likely to be classified as adequate.

In the analysis phase, survey analyses are usually restricted to respondents with complete data. However, analyses of DHS and

**Table 3.** Illustrative questions used in the past to elicit a verbal history of vaccination according to the EPI schedule in the 1980s.

| Recommended Age for Vaccination | Vaccines and How Administered | Example Questions to Mother to Elicit Verbal History |
|---|---|---|
| Birth | BCG (intradermally, usually in the upper arm) | Did the child receive an injection in the upper arm soon after birth? (check for scar) |
| 6 weeks | First dose of DTP (subcutaneous or intramuscular injection, usually in the thigh) and OPV (oral) | Did the child receive an injection in the thigh (the "triple vaccine")? If yes, how many times? Did the child also receive drops in the mouth? If yes, how many times? |
| 10 weeks | DTP, OPV 2 | Same as for 6 weeks |
| 14 weeks | DTP, OPV 3 | Same as for 6 weeks |
| 9 months | Measles (subcutaneous injection, usually in the upper arm) | Did the child receive an injection in the arm against [use local term for measles], after he/she was old enough to sit up or crawl? |

DTP, diphtheria toxoid, tetanus toxoid, and whole cell pertussis vaccine combination; OPV, oral polio vaccine.
doi:10.1371/journal.pmed.1001404.t003

**Table 4.** World Health Organization–recommended EPI schedule, 2012.

| Age of Infant | Parenteral Vaccines | Oral Vaccines |
|---|---|---|
| Birth | BCG, HBV[a] | OPV[b] |
| 6 weeks (some countries give this dose at 8 weeks) | DTP, Hib, HBV, usually administered as pentavalent combination[a]; 10- or 13-valent PnCV[c] | OPV; rotavirus vaccine (Rotateq or Rotarix) |
| 10 weeks (some countries give this dose at 16 weeks) | Pentavalent combination[a]; 10- or 13-valent PnCV (3p+0 schedule)[c] | OPV; rotavirus vaccine (Rotateq or Rotarix) |
| 14 weeks (some countries give this dose at 24 weeks) | Pentavalent combination[a]; 10- or 13-valent PnCV[c] | OPV; rotavirus vaccine[d] (Rotateq) |
| 9–12 months | Measles[e] (rubella[f], with measles); 10- or 13-valent PnCV (2p+1 schedule)[c]; yellow fever (endemic countries)[g]; Japanese encephalitis (endemic countries)[h] | |

Adapted from [67].

[a]Since perinatal or early postnatal transmission is an important cause of chronic infections globally, all infants should receive their first dose of hepatitis B vaccine as soon as possible (<24 hours) after birth even in low-endemicity countries. The primary hepatitis B immunization series conventionally consists of three doses of vaccine (one monovalent birth dose followed by two monovalent or combined vaccine doses at the time of DTP1 and DTP3 vaccine doses). However, four doses may be given for programmatic reasons (e.g., one monovalent birth dose followed by three monovalent or combined vaccine doses with DTP vaccine doses), according to the schedules of national routine immunization programs.

[b]OPV alone, including a birth dose, is recommended in all polio-endemic countries and those at high risk for importation and subsequent spread. A birth dose is not considered necessary in countries where the risk of polio virus transmission is low, even if the potential for importation is high/very high.

[c]For infants, three primary doses (the 3p+0 schedule) or, as an alternative, two primary doses plus a booster (the 2p+1 schedule). If the 3p+0 schedule is used, vaccination can be initiated as early as 6 weeks of age with an interval between doses of 4–8 weeks. If the 2p+1 schedule is selected, the two primary doses should ideally be completed by 6 months of age, starting as early as 6 weeks of age with a minimum interval of 8 weeks between the two doses (for infants aged ≥7 months a minimum interval of 4 weeks between doses is possible). One booster dose should be given at 9–15 months of age.

[d]If Rotarix is used, only two doses are administered.

[e]In countries that have achieved a high level of control of measles, the initial dose of measles vaccine can be administered at 12 months of age. All children are currently expected to receive a second dose of measles vaccine. In the least developed countries this is often administered through mass immunization campaigns.

[f]Rubella vaccine, administered in combination with measles vaccine, is recommended for countries that reliably administer two doses of measles vaccine and have achieved a high level of measles control.

[g]Yellow fever should be co-administered at the infant visit when measles vaccine is administered.

[h]Japanese encephalitis vaccines may be given at age 12 months for children living in highly endemic areas.

DTP, diphtheria toxoid, tetanus toxoid, and whole cell pertussis vaccine combination; HBV, hepatitis B vaccine; Hib, *Haemophilus Influenzae* type b conjugate vaccine; OPV, oral polio vaccine; pentavalent combination, DTP+HBV+Hib formulated to be administered in combination as a single injection; PnCV, pneumococcal conjugate vaccine containing either 10 or 13 separate conjugates of different capsular serotypes.

doi:10.1371/journal.pmed.1001404.t004

MICS surveys prior to 2002 showed that maternally recalled vaccination data were internally consistent, and that inclusion of a verbal history of vaccination in results was preferable to other options such as restricting analyses to children with HBRs or assuming that coverage among those without HBRs was the same as those with HBRs [45]. Inclusion of children without HBRs is only possible for calculation of percentage coverage, however, and not for assessing timeliness of vaccination.

When measuring vaccination coverage from survey data, if data are missing for reasons related to the likelihood of vaccination, restricting analyses to those with complete data will bias results. Analysts can impute hypothetical values for missing responses, but imputing a single value fails to account correctly for the uncertainty associated with selecting an arbitrary (though perhaps plausible) value to impute. More sophisticated methods include integrating over a likelihood function, or imputing numerous values for each missing datapoint (multiple imputation), and are preferred but rare in vaccination coverage measurement [46,47]. DHS and MICS surveys adjust sampling weights for nonresponse, and may impute a single value when a HBR vaccination date is invalid (e.g., lists day 32 of the month) or missing, but do not currently employ multiple imputation in vaccination coverage estimates [48], although this could be introduced in the future as the technique of multiple imputation becomes more accessible.

## How Can the Challenges Facing the Measurement of Vaccination Coverage Be Addressed?

Vaccination coverage is an important indicator that is used to monitor not only immunization programs but also health system performance at national and global levels. Coverage surveys can include questions on reasons for receiving, or not receiving, vaccines, and investigate demographic and other factors associated with coverage [49]. In addition, in the specific context of measles vaccination, coverage is a key component of the cohort analyses that estimate the build-up of susceptible children after vaccination campaigns and that identify when follow-up campaigns are needed [50]. Inflated coverage data contribute to delays in implementing follow-up campaigns, which may lead to measles outbreaks [51]. Finally, coverage data can also be used in field evaluation of vaccine effectiveness [52].

Despite these important public health applications of vaccination coverage data, many challenges face the collection of accurate coverage data [5,6] that must be addressed to guide program implementation and to ensure that funding decisions are based on real performance. For example, although survey data are often preferred to administrative reports for the reasons that we outlined at the start of this article, surveys are also subject to errors (Table 2). Groups that may be omitted from sampling frames—for example, those living in conflict-affected or otherwise inaccessible areas or in rapidly growing urban areas [53], and recent migrants—may have lower vaccination coverage [1]. Depending on the size of these population groups and the purposes of the evaluation, substantial extra effort may be needed to update the sampling frame or to conduct special surveys among these groups.

Among potential sources of non-sampling error, one of the most critical challenges is the availability of HBRs, which has been low in surveys of some of the countries contributing most to global estimates of vaccination coverage [54]. In India, WHO and UNICEF coverage estimates since 2009 have been based on a

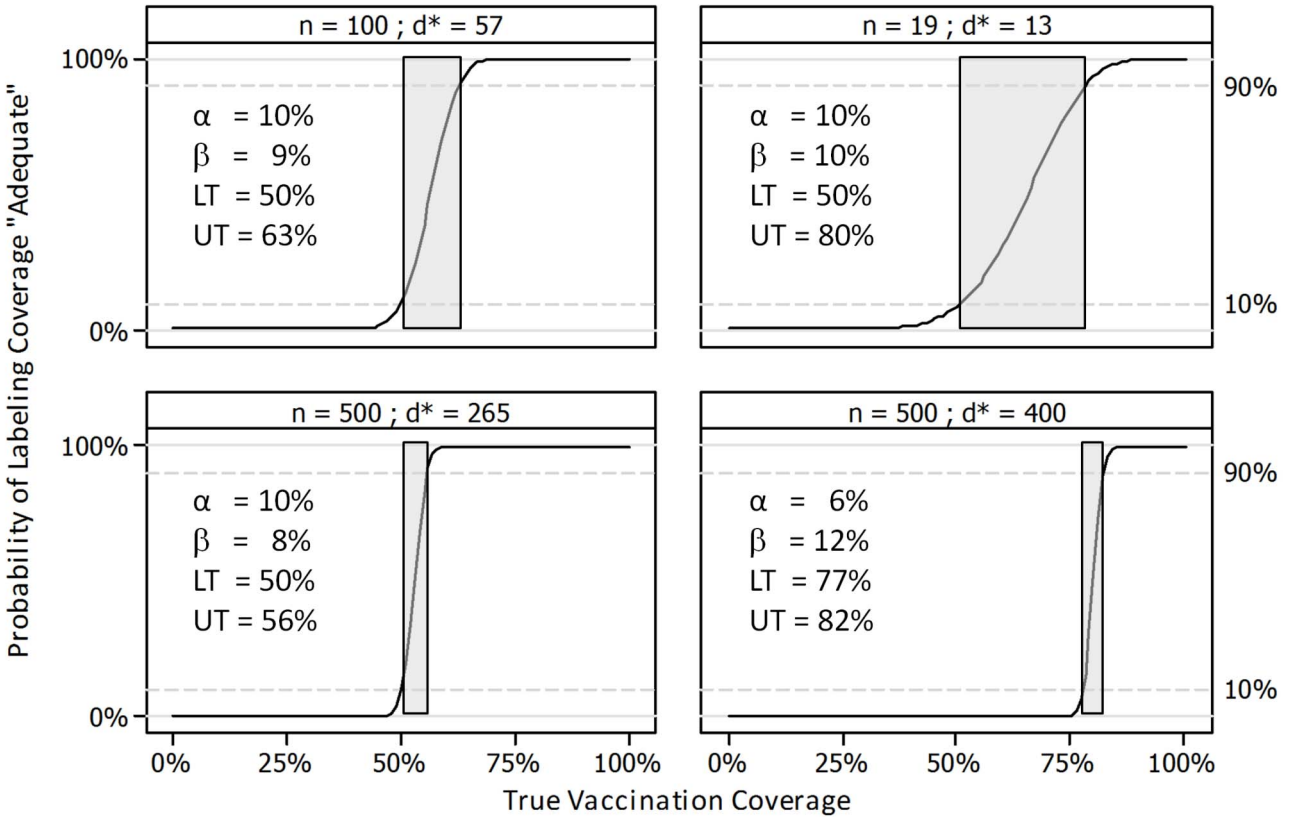


**Figure 3. Operating characteristic curves for four LQAS sampling plans.** In each panel, the curve indicates the probability of finding *d** or more vaccinated children in a random sample of size *n*. Lots with coverage≤lower threshold (LT) will be classified as having inadequate coverage with probability ≥(1 − α). Lots with coverage≥upper threshold (UT) will be classified as having adequate coverage with probability ≥(1 − β). The gray area is the region where LT<coverage<UT; lots with coverage in the gray area may be labeled either adequate or inadequate. The gray area includes the region of coverage, for instance, where there is a 50/50 probability of being classified adequate or inadequate. Neither classification (adequate or inadequate) rules out the strong possibility that the true coverage lies in the gray area. The gray area may be made larger or smaller and may be moved to regions of higher or lower coverage by manipulating LT, UT, α, and β to arrive at different values of *n* and *d**.
doi:10.1371/journal.pmed.1001404.g003

2008 coverage evaluation survey in which vaccine cards were seen for only 52% of children [55]; in Nigeria, HBRs were seen in 40% of children in the 2009 national immunization coverage survey and in only 26% in the 2007 DHS [56]. Obtaining records of vaccines administered after infancy, including booster doses and vaccines targeted to older age groups such as tetanus toxoid, which is administered to pregnant women, is even more difficult. Records of administration of vaccines containing tetanus toxoid may be available for the most recent pregnancy but not for previous pregnancies, mass campaigns, or early childhood. Serological surveys show that vaccination coverage surveys tend to underestimate the prevalence of protection against neonatal tetanus [57]. Thus, unless primary recording of vaccination data on HBRs and clinic records is improved, investment in other strategies to improve survey methods or administrative estimates will have limited effect.

When choosing a survey methodology, decision makers should consider both the specific information needed and the speed with which the information is required. The timing and frequency of DHS and MICS surveys are not decided by immunization program managers, and thus data are used by these health professionals as and when they become available (which is several months after survey completion even for the preliminary results). In the years between DHS and MICS surveys, program managers may need additional information on vaccination coverage to improve planning and may prefer the quicker and less expensive EPI surveys; we propose ways to improve these below. Sometimes, vaccination coverage estimates are needed very rapidly (e.g., during campaigns, while the vaccination team is still in the area), and a purposive sampling method focused on areas most likely to have low coverage has been used [9,58]. Recent experience with rapid monitoring of polio campaigns using purposive sampling, however, has not been positive, and LQAS surveys are now promoted for polio campaign monitoring [31].

Although biomarkers are under consideration for validating coverage, they have several limitations for this purpose and are not currently included routinely in surveys measuring childhood vaccination coverage. For most vaccines, the presence of antibody following vaccination cannot be distinguished from that following "natural" infection. Exceptions are the presence of tetanus antibody (infection does not generate lasting immunity) and the presence of antibodies to antigens included in subunit vaccines such as hepatitis B but not to other antigens found in the whole organism (an indication of vaccine-induced immunity) [59]. Even for these vaccines, detection of antibodies does not indicate reliably how many doses have been received [60]. Furthermore, absence of antibody means either that the child is unvaccinated or that vaccines have lost their potency. Biomarkers are therefore potentially useful to estimate population-level protection [61] but not necessarily to validate coverage measurements or vaccination

program performance. The development of antibody assays on oral fluid samples for tetanus [60] and measles [62] may make surveys with repeated sample collection more acceptable and may allow evaluation of vaccination campaigns [63].

## Recommendations and Conclusions

To reduce bias in coverage measurement by any method, we recommend that primary recording of vaccination data be improved. In the long term, we recommend further investment in the development, evaluation, and roll-out of effective systems for digital recording and data transmission. In the short-to-medium term in low- and middle-income countries, paper-based recording must be improved. Record design will need to evolve rapidly to accommodate the introduction of new and underutilized vaccines and to allow recording of doses administered through campaigns and across the life course. We recommend that research be conducted to improve the design of primary records to reduce recording errors and facilitate compilation of data. We also recommend that health workers educate mothers so that they value HBRs, understand the information therein, and retain them safely.

Whenever surveys are conducted, all efforts to ensure high-quality data should be made (Table 2). In particular, probability samples should be used, the sample size and number of clusters should be selected appropriately for the survey objectives, efforts should be made to encourage mothers to have HBRs ready at the time of the survey, households should be revisited if necessary to interview a suitable respondent and see the HBR, health facilities should be visited to look for records of children whose HBRs were unavailable, and strict quality control measures should be implemented. Analyses should incorporate internal consistency checks as part of quality control, including assessment of the validity of the verbal history of vaccination. These checks include comparison of prevalence of a Bacille Calmette Guerin vaccine against tuberculosis (BCG) scar among children with documented versus verbal history of BCG [32], and comparison of coverage among those who reported receiving a vaccination card but did not present it and those who presented a card [45]. In multi-indicator surveys, vaccination coverage could be cross-tabulated against coverage of other interventions for persons with and without a HBR; if the verbal history is reliable, the same associations should be found in both groups. Large-scale survey programs should also be evaluated periodically to sustain attention to quality control [64].

Technical expertise for collecting high-quality data and interpreting and using results needs to be further developed at national and sub-national levels. Program managers need to use coverage data with other program indicators to improve program planning and management. Identification of low-coverage areas should trigger action to reach underserved children, who are often those at highest risk of dying should they acquire a vaccine-preventable infection [65]. As a measure of population protection, coverage is currently limited by assumptions about vaccine effectiveness and thus is helpful but not sufficient. Additional information from vaccine management assessments, surveillance, outbreak investigations, and, where available, special studies such as case control studies of vaccine effectiveness should be reviewed together with coverage data to obtain a fuller picture of program success. Sero-surveys and vaccination coverage surveys are likely to complement each other for the foreseeable future, and we recommend that further research into the use of sero-surveys and the development of new biomarkers be undertaken.

Sources of uncertainty in surveys must be considered before drawing strong conclusions about their results. It is common to focus on the point estimate of coverage from the survey, but it is also important to consider uncertainty due to sampling design, which is usually expressed using a confidence interval, and potential biases, which should be assessed by reviewing information about the survey protocol and its implementation. The recent inclusion of a "grade of confidence" in national immunization coverage estimates produced by WHO-UNICEF is an important first step towards improving the usefulness of these estimates, and we recommend that grading of the quality of surveys should also be done. In particular, when LQAS surveys are done, we recommend that their results be interpreted with due recognition of the gray area [66].

Finally, as we mentioned at the start of this review, there is currently tension between financing systems, which reward high coverage, and efforts to improve the quality of coverage measurement. We believe and strongly recommend that it is time to reward actions that improve the quality of data, particularly those discussed in this review, rather than rewarding a country's apparent coverage achievements.

### Key Points

- Vaccination coverage is an important indicator of public health if measured accurately; at present, well-designed and executed surveys provide more accurate and comparable results than administrative reports, which are subject to incomplete and inaccurate reporting of the numerator and inaccurate estimation of the denominator.
- To reduce bias in coverage measurement based on surveys and on administrative reports, primary recording of vaccination data on home-based records and clinic records must be improved. In the long term, this will involve digital recording and data transmission. In the short-to-medium term in low-income countries, paper-based recording must be improved.
- Whenever surveys are done, to minimize selection bias and information bias, the sample size should be selected according to program needs, probability sampling should be used, and strict quality control measures should be implemented for data collection and analysis.
- The potential magnitude of bias in surveys must be assessed before results are interpreted, quality assessment criteria should be developed and endorsed by partners, and partners should consider uncertainty in coverage estimates before basing decisions such as those involving performance-based financing on coverage.
- To improve program performance, national immunization programs and their partners should take action to improve the collection, interpretation, and use of vaccination coverage data together with data on other indicators.


## Acknowledgments

We thank Fred Arnold, Tony Burton, Thomas Cherian, Thom Eisele, and Attila Hancioglu for comments on earlier drafts of this work.


## Author Contributions

Conducted literature review: FTC. Analyzed the data: FTC DAR. Wrote the first draft of the manuscript: FTC DAR. Contributed to the writing of the manuscript: FTC HSI DAR. ICMJE criteria for authorship read and met: FTC HSI DAR. Agree with manuscript results and conclusions: FTC HSI DAR.

## References


1. Hadler SC, Dietz V, Okwo-Bele JM, Cutts FT (2008) Vaccination programs in developing countries. In: Plotkin SA, Orenstein WA, editors. Vaccines, 5th edition. Philadelphia: Saunders.
2. Brenzel L, Measham A, Naimoli J, Batson A, Bredenkamp C, et al. (2009) Taking stock: World Bank experience with results-based financing (RBF) for health. Washington: The World Bank. Available: http://www.rbfhealth.org/rbfhealth/library/doc/taking-stock-world-bank-experience-results-based-financing-rbf-health. Accessed 29 August 2012.
3. Millenium Challenge Corporation (2011) Report on the criteria and methodology for determining the eligibility of candidate countries for Millennium Challenge Account assistance in fiscal year 2012. Available: https://www.mcc.gov/documents/reports/report-2011001066201-fy12-selection-criteria.pdf. Accessed 29 August 2012.
4. GAVI Alliance (2012) Country eligibility policy. Available: http://www.gavialliance.org/about/governance/programme-policies/country-eligibility/. Accessed 29 August 2012.
5. Lim SS, Stein DB, Charrow A, Murray CJL (2008) Tracking progress towards universal childhood immunisation and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage. Lancet 372: 2031–2046.
6. Murray CJL, Shengelia B, Gupta N, Moussavi S, Tandon A, et al. (2003) Validity of reported vaccination coverage in 45 countries. Lancet 362: 1022–1027.
7. Bosch-Capblanch X, Ronveaux O, Doyle V, Remedios V, Bchir A (2009) Accuracy and quality of immunization information systems in forty-one low income countries. Trop Med Int Health 14: 2–10.
8. Burton T, Neil M, Okwo-Bele JM, Salama P, Wardlaw T (2009) Measurement of immunisation coverage. Lancet 373: 210–211.
9. Dietz V, Venczel L, Izurieta H, Stroh G, Zell ER, et al. (2004) Assessing and monitoring vaccination coverage levels: lessons from the Americas. Rev Panam Salud Publica 16: 432–442.
10. Hull B, Dey A, Mahajan D, Menzies R, McIntyre PB (2011) Immunisation coverage annual report, 2009. Commun Dis Intell Q Rep 35: 132–148.
11. Ronveaux O, Arrieta F, Curto S, Laurani H, Danovaro-Holliday M (2009) Assessment of the quality of immunization data produced by the national individual registration system in Uruguay, 2006. Rev Panam Salud Publica 26: 153–160.
12. Luhm KR, Cardoso MRA, Waldman EA (2011) Vaccination coverage among children under two years of age based on electronic immunization registry in Southern Brazil. Rev Saude Publica 45: 90–98.
13. Crowcroft NS (2009) Action on immunisation: no data, no action. Arch Dis Child 94: 829–830.
14. PATH (2012) Rethinking the vaccine supply chain. Available: http://www.path.org/projects/project-optimize. Accessed 29 August 2012.
15. Burton A, Monasch R, Lautenbach B, Gacic-Dobo M, Neill M, et al. (2009) WHO and UNICEF estimates of national infant immunization coverage: methods and processes. Bull World Health Organ 87: 535–541.
16. MEASURE DHS Demographic and Health Surveys. Available: http://www.measuredhs.com/. Accessed 1 June 2012.
17. UNICEF. Statistics and Monitoring: Multiple Indicator Cluster Survey. Available: http://www.unicef.org/statistics/index_24302.html. Accessed 29 June 2012.
18. Hancioglu A, Arnold F (2013) Measuring coverage in MNCH: Tracking progress in health for women and children using DHS and MICS household surveys. PLoS Med 10: e1001391. doi:10.1371/journal.pmed.1001391
19. Henderson RH, Sundaresan T (1982) Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. Bull World Health Organ 60: 253–260.
20. World Health Organization (1991) Facilitator guide for the EPI coverage survey. Available: http://www.who.int/immunization_monitoring/routine/Facilitator_guide_EPI_coverage_survey.pdf. Accessed 3 January 2013.
21. World Health Organization (2005) Immunization coverage cluster survey—reference manual. Available: http://whqlibdoc.who.int/hq/2005/WHO_IVB_04.23.pdf. Accessed 29 October 2012.
22. Grais RF, Rose AMC, Guthmann JP (2007) Don't spin the pen: two alternative methods for second-stage sampling in urban cluster surveys. Emerg Themes Epidemiol 4: 8.
23. Luman ET, Worku A, Berhane Y, Martin R, Cairns L (2007) Comparison of two survey methodologies to assess vaccination coverage. Int J Epidemiol 36: 633–641.
24. Milligan P, Alpha N, Bennett S (2004) Comparison of two cluster sampling methods for health surveys in developing countries. Int J Epidemiol 33: 469–476.
25. Myatt M, Feleke T, Sadler K, Collins S (2005) A field trial of a survey method for estimating the coverage of selective feeding programmes. Bull World Health Organ 83: 20–26.
26. Turner AG, Magnani RJ, Shuaib M (1996) A not quite as quick but much cleaner alternative to the Expanded Programme on Immunization (EPI) cluster survey design. Int J Epidemiol 25: 198–203.
27. World Health Organization (2012) Sample design and procedures for hepatitis B immunization surveys: a companion to the WHO cluster survey reference manual. Available: http://whqlibdoc.who.int/hq/2011/WHO_IVB_11.12_eng.pdf. Accessed 29 August 2012.
28. Cotter B, Bremer V, Stroh G, Msambichaka K, Mabuzane E, et al. (2003) Assessment of neonatal tetanus elimination in an African setting by lot quality assurance cluster sampling (LQA-CS). Epidemiol Infect 130: 221–226.
29. Pezzoli L, Pineda S, Halkyer P, Crespo G, Andrews N, et al. (2009) Cluster-sample surveys and lot quality assurance sampling to evaluate yellow fever immunisation coverage following a national campaign, Bolivia, 2007. Trop Med Int Health 14: 355–361.
30. Kim SH, Pezzoli L, Yacouba H, Coulibaly T, Djingarey MH, et al. (2012) Whom and where are we not vaccinating? Coverage after the introduction of a new conjugate vaccine against group A meningococcus in Niger in 2010. PLoS ONE 7: e29116. doi:10.1371/journal.pone.0029116
31. World Health Organization (2012) Assessing vaccination coverage levels using clustered Lot Quality Assurance Sampling: field manual—version edited for the Global Polio Eradication Initiative (GPEI). Available: http://www.polioeradication.org/Portals/0/Document/Research/OPVDelivery/LQAS.pdf. Accessed 29 August 2012.
32. Jahn A, Floyd S, Mwinuka V, Mwafilaso J, Mwagomba D, et al. (2008) Ascertainment of childhood vaccination histories in northern Malawi. Trop Med Int Health 13: 129–138.
33. Cutts FT, Smith PG, Colombo S, Mann G, Ascherio A, et al. (1990) Field evaluation of measles vaccine efficacy in Mozambique. Am J Epidemiol 131: 349–355.
34. World Health Organization (2012) World immunization week 2012. Available: http://www.who.int/immunization/newsroom/events/immunization_week/en/index.html. Accessed 15 May 2012.
35. Gareaballah ET, Loevinsohn BP (1989) The accuracy of mother's reports about their children's vaccination status. Bull World Health Organ 67: 669–674.
36. Langsten R, Hill K (1998) The accuracy of mothers' reports of child vaccination: evidence from rural Egypt. Soc Sci Med 46: 1205–1212.
37. Valadez JJ, Weld LH (1992) Maternal recall error of child vaccination status in a developing nation. Am J Public Health 82: 120–122.
38. Clark A, Sanderson C (2009) Timing of children's vaccinations in 45 low-income and middle-income countries: an analysis of survey data. Lancet 373: 1543–1549.
39. Dayan GH, Shaw KM, Baughman AL, Orellana LC, Forlenza R, et al. (2006) Assessment of delay in age-appropriate vaccination using survival analysis. Am J Epidemiol 163: 561–570.
40. Luman ET, Barker LE, Shaw KM, McCauley MM, Buehler JW, et al. (2005) Timeliness of childhood vaccinations in the United States. JAMA 293: 1204–1211.
41. Babirye JN, Engebretsen IMS, Makumbi F, Fadnes LT, Wamani H, et al. (2012) Timeliness of childhood vaccinations in Kampala Uganda: a community-based cross-sectional study. PLoS ONE 7: e35432. doi:10.1371/journal.pone.0035432
42. Eisele T, Rhoda DA, Cutts FT, Keating J, Ren R, et al. (2013) Measuring coverage in MNCH: Total survey error and the interpretation of intervention coverage estimates from household surveys. PLoS Med 10: e1001386. doi:10.1371/journal.pmed.1001386
43. Bennett S, Radalowicz A, Vella V, Tomkins A (1994) A computer simulation of household sampling schemes for health surveys in developing countries. Int J Epidemiol 23: 1282–1291.
44. Lemeshow S, Tserkovnyi AG, Tulloch JL, Dowd JE, Lwanga SK, et al. (1985) A computer simulation of the EPI survey strategy. Int J Epidemiol 14: 473–481.
45. Brown J, Monasch R, Bicego G, Burton A, Boerma JT (2002) An assessment of the quality of national child immunization coverage estimates in population-based surveys. Chapel Hill (North Carolina): MEASURE Evaluation, Carolina Population Center, University of North Carolina at Chapel Hill. Available: http://www.cpc.unc.edu/measure/publications/wp-02-53/at_download/document. Accessed 29 August 2012.
46. Little RJA, Rubin DB (2002) Statistical analysis with missing data. Hoboken (New Jersey): John Wiley & Sons.
47. Schafer JL (1997) Analysis of incomplete multivariate data. Boca Raton (Florida): Chapman & Hall/CRC.
48. Rutstein SO, Rojas G (2006) Guide to DHS statistics. Calverton (Maryland): Demographic and Health Surveys, ORC Macro.
49. Bosch-Capblanch X, Banerjee K, Burton A (2012) Unvaccinated children in years of increasing coverage: how many and who are they? Evidence from 96 low- and middle-income countries. Trop Med Int Health 17: 697–710.
50. de Quadros CA, Olive JM, Hersh BS, Strassburg MA, Henderson DA, et al. (1996) Measles elimination in the Americas. Evolving strategies. JAMA 275: 224–229.
51. Simons E, Ferrari M, Fricks J, Wannemuehler K, Anand A, et al. (2012) Assessment of the 2010 global measles mortality reduction goal: results from a model of surveillance data. Lancet 379: 2173–2178.
52. Orenstein WA, Bernier RH, Dondero TJ, Hinman AR, Marks JS, et al. (1985) Field evaluation of vaccine efficacy. Bull World Health Organ 63: 1055–1068. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2536484/pdf/bullwho00089-0084.pdf. Accessed 29 August 2012.
53. Bharti N, Tatem AJ, Ferrari MJ, Grais RF, Djibo A, et al. (2011) Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. Science 334: 1424–1427.

54. Brown DW (2012) Child immunization cards: essential yet underutilized in national immunization programmes. Open Vaccine J 5: 1–7.
55. World Health Organization (2011) India: WHO and UNICEF estimates of immunization coverage: 2011 revision. Available: http://www.who.int/immunization_monitoring/data/ind.pdf. Accessed 29 June 2012.
56. World Health Organization (2011) Nigeria: WHO and UNICEF estimates of immunization coverage: 2011 revision. Available: http://www.who.int/immunization_monitoring/data/nga.pdf. Accessed 29 August 2012.
57. World Health Organization (2006) Tetanus vaccine WHO position paper. Wkly Epidemiol Rec 85: 197–208.
58. Izurieta H, Venczel L, Dietz V, Tambini G, Barrezueta O, et al. (2003) Monitoring measles eradication in the region of the Americas: critical activities and tools. J Infect Dis 187: S133–S139.
59. Cutts F, Hall A (2004) Vaccines for neonatal viral infections: hepatitis B vaccine. Expert Rev Vaccines 3: 349–352. Accessed 29 August 2012.
60. Tapia MD, Pasetti MF, Cuberos L, Sow SO, Doumbia MN, et al. (2006) Measurement of tetanus antitoxin in oral fluid: a tool to conduct serosurveys. Pediatr Infect Dis J 25: 819–825.
61. Fortuin M, Maine N, Mendy M, Hall A, George M, et al. (1995) Measles, polio and tetanus toxoid antibody levels in Gambian children aged 3 to 4 years following routine vaccination. Trans R Soc Trop Med Hyg 89: 326–329.
62. Nigatu W, Nokes D, Enquselassie F, Brown D, Cohen B, et al. (1999) Detection of measles specific IgG in oral fluid using an FITC/anti-FITC IgG capture enzyme linked immunosorbent assay (GACELISA). J Virol Methods 83: 135–144.
63. Nigatu W, Samuel D, Cohen B, Cumberland P, Lemma E, et al. (2008) Evaluation of a measles vaccine campaign in Ethiopia using oral-fluid antibody surveys. Vaccine 26: 4769–4774.
64. Ahmed S, Ali D, Bisharat L, Hill A, LaFond A, et al. (2009) Evaluation of UNICEF Multiple Indicator Cluster Surveys round 3 (MICS3) final report. Boston (Massachusetts): John Snow. Available: http://www.unicef.org/evaldatabase/files/MICS.pdf. Accessed 29 August 2012.
65. Rheingans R, Atherly D, Anderson J (2012) Distributional impact of rotavirus vaccination in 25 GAVI countries: estimating disparities in benefits and cost-effectiveness. Vaccine 30 (Suppl 1): A15–A23. Available: http://ac.els-cdn.com/S0264410X12000333/1-s2.0-S0264410X12000333-main.pdf?_tid=41e1f1f9a28236cbe2409273fa8fecc4&acdnat=1340737914_ce1e78df52b569cf4a045350c373885c. Accessed 29 August 2012.
66. Rhoda DA, Fernandez SA, Fitch DJ, Lemeshow S (2010) LQAS: user beware. Int J Epidemiol 39: 60–68.
67. World Health Organization (2012) Table 2: summary of WHO position papers—recommended routine immunizations for children. Available: http://www.who.int/immunization/policy/Immunization_routine_table2.pdf. Accessed 23 July 2012.

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: Current Indicators for Measuring Coverage of Diarrhea Treatment Interventions and Opportunities for Improvement

**Christa L. Fischer Walker[1]*, Olivier Fontaine[2], Robert E. Black[1]**

**1** Johns Hopkins Bloomberg School of Public Health, Department of International Health, Baltimore, Maryland, United States of America, **2** World Health Organization, Geneva, Switzerland (Retired)

**Abstract:** Diarrhea morbidity and mortality remain important child health problems in low- and middle-income countries. The treatment of diarrhea and accurate measurement of treatment coverage are critical if child mortality is going to continue to decline. In this review, we examine diarrhea treatment coverage indicators collected in two large-scale community-based household surveys—the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS). Current surveys do not distinguish between children with mild diarrhea episodes and those at risk for dehydration. Additional disease severity questions may improve the identification of cases of severe diarrhea but research is needed to identify indicators with the highest sensitivity and specificity. We also review the current treatment indicators in these surveys and highlight three areas for improvement and research. First, specific questions on fluids other than oral rehydration salts (ORS) should be eliminated to refocus the treatment of dehydration on ORS and to prevent confusion between prevention and treatment of dehydration. Second, consistency across surveys and throughout translations is needed for questions about the caregiver behavior of "offering" the sick child fluid and food. Third, breastfeeding should be separated from other fluid and food questions to capture the frequency and duration of nursing sessions offered during the illness. Research is also needed to assess the accuracy of the current zinc indicator to determine if caregivers are correctly recalling zinc treatment for current and recent diarrhea episodes.

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection*

## Introduction

Child diarrhea mortality rates have declined dramatically over the past 30 years, yet diarrhea morbidity has remained relatively constant among children under 5 years of age in low- and middle-income countries [1,2]. Current estimates suggest that, despite improvements in water, sanitation, and hygiene, children under 5 years old have 2.9 episodes of diarrhea every year with the highest rates among children aged 6–11 months [2]. Rotavirus is widely accepted as the leading cause of hospitalizations among children under 5 years but other viruses, bacteria, and parasites also cause serious diarrheal morbidity and mortality [3,4].

Oral rehydration salts (ORS) have been the cornerstone of diarrhea treatment since the 1980s [5]. Combined with continued feeding and the provision of home-based sugar-salt solution and other fluids, diarrhea treatment in the home should now be easier than ever for most community-acquired acute diarrhea episodes and experts had high hopes for accelerated uptake and widespread use of ORS within the community [6]. Unfortunately, although knowledge of ORS has remained high, more than two-thirds of low- and middle-income countries have reported declines in ORS use rates in the years following the initial campaigns and promotional efforts [7]. More positively, with the introduction of zinc supplementation for 10–14 days as an adjunct treatment for all episodes of childhood diarrhea [8], diarrhea treatment is now more effective than ever, and remains simple, inexpensive, and appropriate for community-based care.

A comprehensive understanding of which children with diarrhea are getting treatment and how this has or will change over time is critical for targeting child health programs. Our current understanding of the coverage of ORS and now zinc treatment comes primarily from Demographic and Health Surveys (DHS) and UNICEF Multiple Indicator Cluster Surveys (MICS). In this paper, which is part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, we review the current methodology for assessing coverage of ORS, additional fluids, continued feeding, and zinc treatment. We identify problems with the currently accepted indicators and propose opportunities for improving these indicators. Better coverage indicators will increase our understanding of current trends in the treatment of diarrhea

**Citation:** Fischer Walker CL, Fontaine O, Black RE (2013) Measuring Coverage in MNCH: Current Indicators for Measuring Coverage of Diarrhea Treatment Interventions and Opportunities for Improvement. PLoS Med 10(5): e1001385. doi:10.1371/journal.pmed.1001385

**Academic Editor:** David Osrin, Wellcome Trust Senior Research Fellow in Clinical Science, UCL Reader in International Child Health, Honorary Consultant, Great Ormond Street Hospital for Children, United Kingdom

**Published** May 7, 2013



**Funding:** This work was conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF, with financial support from the Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** DHS, Demographic and Health Survey/s; MICS, Multiple Indicator Cluster Survey/s; MNCH, maternal, newborn, and child health; NPV, negative predictive value; ORS, oral rehydration salts; PPV, positive predictive value

* E-mail: cfischer@jhsph.edu

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

and enhance our ability to target and improve these interventions among young children.

## Overview of Current Methods for Assessing Coverage

DHS and MICS surveys have been measuring the coverage of key child survival interventions since 1984 and 1995, respectively [9–11]. These large, representative, cross-sectional household surveys are designed to track demographic and health indicators and to measure the coverage of key interventions in low- and middle-income countries. Details of the DHS and MICS survey methodologies are provided elsewhere in this Collection [11].

Except for the DHS in Peru and Senegal, which now employ continuous rolling surveys, DHS or MICS surveys are not executed in the field over a full calendar year. Because diarrhea disease rates are highly dependent on season [4], both surveys suffer from seasonality bias and cannot be used to accurately measure diarrhea incidence among children under 5 years of age [12,13]. Cross-sectional surveys can, however, determine the standard denominator for measuring the coverage of key diarrhea treatment interventions—2-week diarrhea point prevalence. Thus, in DHS and MICS surveys, the caregivers of children under the age of 5 are asked if the child has had diarrhea at any time in the past 2 weeks. The questionnaire is designed to capture all episodes of diarrhea of varying degrees of severity, with or without blood in the stool, using local terminology for diarrhea rather than clinical definitions of disease duration or severity. Both DHS and MICS survey guidelines state that local terms should be included to capture the full spectrum of diarrhea as perceived in the local community.

DHS and MICS surveys also capture information on the coverage of key interventions for the treatment of diarrhea, namely ORS for the prevention and treatment of dehydration, zinc supplementation, continued feeding, and the provision of additional fluids during the episode. These treatment indicators are captured for all children with a diarrhea episode in the past 2 weeks. The caregiver is also asked about the quantity of fluids and foods that have been given during the current diarrhea episode in relation to what the child is normally given. The definitions of treatment coverage indicators most commonly used for program purposes are defined in Box 1.

## The Definition of Diarrhea Episodes in DHS and MICS Surveys

The current denominator for treatment coverage obtained from DHS and MICS surveys includes all children with a diarrhea episode in the last 2 weeks as defined by the caregiver. Unfortunately, although considerable formative research was done in the early 1980s to understand local terms and beliefs about diarrhea [14], this information is often overlooked in the design of questionnaires and simple mistakes in the translation of survey instruments or the omission of key local terminology can result in missed diarrhea episodes. Moreover, current surveys assume that all diarrhea episodes are in need of the same level of treatment. Older surveys asked caregivers about the duration of the diarrhea episode, but this is no longer common. Currently, caregivers are asked about the presence of blood in the stool, but are not asked questions designed to classify disease severity, because this is thought to be difficult in household surveys with 2-week recall. Because the denominator broadly captures all diarrhea episodes, it is impossible to determine whether the children receiving appropriate treatment are the children most in need or simply a random sample of all children with diarrhea.

### Box 1. Treatment Coverage Indicators Included in DHS and MICS Surveys

*Indicators from direct responses to DHS/MICS questionnaires*

1. Proportion of children with diarrhea in last 2 weeks who were given ORS
2. Proportion of children with diarrhea in last 2 weeks who were given zinc
3. Proportion of children with diarrhea in last 2 weeks who were given recommended home fluids
4. Proportion of children with diarrhea in last 2 weeks who were given the same or more to drink
5. Proportion of children with diarrhea in last 2 weeks who were given the same or more to eat

*Indicators calculated from multiple DHS/MICS questions*

1. Proportion of children who received oral rehydration therapy (defined as ORS or recommended home fluids)
2. Proportion of children who received continued feeding and oral rehydration therapy or increased fluids

## How Can the Definition of Diarrhea Episodes Be Improved?

It is clear from the above discussion that large surveys need to include additional questions on diarrhea duration and severity to enable accurate measurement of treatment coverage. Such additional questions could use easily recognizable signs and symptoms, such as fever or abnormal thirst and vomiting to measure the prevalence of dehydration in the current or recalled episodes.

We identified two studies that showed various combinations of these signs as being predictive of dehydration [15,16]. In a case control study in Brazil, researchers enrolled children who were hospitalized for dehydration as cases and age-matched children who also had diarrhea but were not hospitalized as controls. Mothers were asked to recount signs and symptoms on the first day of illness. The study concluded that using vomiting or fever as an indicator would identify 75% of diarrhea episodes with dehydration [15]. In a similar study in Mozambique, vomiting or fever had a sensitivity of 68.3% (sensitivity measures how well an indicator is able to identify true positive cases, i.e., diarrhea with dehydration). By adding "drinking more than usual" to this combination, the sensitivity of diarrhea prediction increased to 87.8% with a specificity of 34.1% (specificity measures how well an indicator is able to identify true negative cases, i.e., diarrhea without dehydration) [16]. We also found several studies that published risk factors for severe disease including socioeconomic factors, child characteristics, and some clinical signs and symptoms [17–19]. Although additional research is clearly needed to refine and retest these signs and symptoms before introducing new questions in large-scale surveys across low- and middle-income countries, these data suggest that qualifying diarrhea severity at the community level is possible.

With the objective of demonstrating that the addition of simple questions to large-scale surveys might provide valuable insights into diarrhea severity and into treatment by diarrhea severity, we used the sensitivities and specificities provided by Victora et al. [15] for combinations of reported signs and symptoms and $2\times2$ tables to calculate the positive and negative predictive values of these symptoms as predictors of dehydration during diarrhea. The

positive predictive value (PPV) of a test indicates the proportion of individuals with a positive test result (here, diarrhea with dehydration) who actually have the disease being tested for. The negative predictive value (NPV) indicates the proportion of individuals with a negative test result who do not have the disease being tested for.

Victora et al. reported that the combination of "vomiting or fever or abnormal thirst" had the highest sensitivity (90%) and the lowest specificity (38%) (Table 1) [15]. The combination of "fever or vomiting" had the lowest reported sensitivity of 75% and the highest specificity of 66%. Typically, with any test, as sensitivity increases, specificity decreases and vice versa. Thus, as the signs or symptoms used to define diarrhea with dehydration become broader, more children meet the criteria of a "case" and the definition will capture a higher percentage of true cases. However, many more children who are not truly cases will also meet the case definition, which will increase false positivity and lower specificity.

Unlike sensitivity and specificity, PPV increases as the disease prevalence increases. We therefore tested scenarios with a 5% and a 10% prevalence of dehydrating diarrhea. At these relatively low prevalences, the PPV for the different combinations of signs and symptoms did not vary widely (Table 1). Thus, for this set of specificities and sensitivities, additional questions about the presence of vomiting, fever, and abnormal thirst could correctly identify 75%–90% (i.e., sensitivity can be high) of diarrhea cases most in need of ORS (i.e., diarrhea with dehydration), but of those individuals that appear to have severe diarrhea using these signs and symptoms as a set of indicators, less than 20% will truly have an episode of diarrhea with dehydration (i.e., PPV is low).

Additional validation studies are needed to test these and other possible indicators in several settings before any disease severity questions are universally added to surveys. The specific wording of questions will also need to be studied across several locations and in several diverse cultures. Although other risk factors for diarrhea or severe diarrhea have been identified [17–19], we suggest that the focus in DHS and MICS questionnaires should remain on signs and symptoms of the episode that are simple to identify and recall. In addition to those reported by Victora et al. [15], questions on total days with diarrhea for completed episodes, number of days of illness for current episodes, and stools per day may further define the severity of the diarrhea episodes and improve our understanding of differences, if any, with regard to treatment or care seeking. These questions should be evaluated to determine if adding them to the survey will increase specificity and sensitivity of identifying cases of diarrhea at risk of progressing to dehydration.

## Two-Week Point Prevalence of Diarrhea or Less?

Surveys are currently designed to capture a 2-week point prevalence of diarrhea, which assumes that recall of up to 2 weeks accurately captures both current and past episodes. Research has shown that longer recall periods actually underestimate milder diarrhea cases by approximately 40% and more severe cases by approximately 20% [20]. A shorter recall period may therefore be critical to more accurately describe diarrhea severity and duration even though a 1-week versus a 2-week recall would reduce the number of cases available for calculating treatment and care-seeking behaviors [21]. As with the inclusion of additional indicators to define severity, shifting the recall period in surveys from 2 weeks to 1 week needs to be tested before widespread changes are made. Making such changes may be logistically challenging given that DHS/MICS do not base sample size on diarrhea prevalence. Nevertheless, understanding the ideal recall period may still provide valuable information with regard to the coverage of diarrhea treatment interventions.

## Improving the Measurement of Coverage of ORS and Additional Fluids

Our review of the treatment indicators included in current DHS and MICS surveys revealed three major problems with the measurement of coverage of treatment with ORS and other fluids. Here, we describe these problems and recommend how they can be avoided in future surveys.

### Recommendation 1. Future Surveys Should Refocus on ORS

ORS is currently recommended for the *treatment* of all episodes of diarrhea. For cases of acute diarrhea with no signs of dehydration, fluids other than ORS may be used for the *prevention* of dehydration. The provision of additional fluids (except ORS) during a diarrhea episode was never intended as treatment for dehydration but, over time, the appropriate use of non-specific fluids in the management of dehydration has led to some confusion, and has also been poorly studied in general. For example, after the discovery of ORS, researchers tried to recreate a version of this life-saving intervention that could be made in the home if pre-packaged ORS was not available. Sugar-salt solutions were tested in hospitals against the packaged ORS formula and proved to be beneficial in clinical settings [22]. By contrast, evaluations of sugar-salt solutions in community settings were typically undertaken without control groups or comparison areas [23,24]. Many other types of fluids have

**Table 1.** Range of diagnostic values for dehydration from diarrhea based on selected severity indicators.

| Indicators | Sensitivity[a] | Specificity[a] | 5% Prevalence of Dehydration | | 10% Prevalence of Dehydration | |
|---|---|---|---|---|---|---|
| | | | PPV | NPV | PPV | NPV |
| Thirst or fever or vomiting | 90% | 38% | 7% | 99% | 14% | 97% |
| Thirst or fever | 89% | 44% | 8% | 99% | 15% | 97% |
| Thirst or vomiting | 89% | 40% | 7% | 99% | 14% | 97% |
| Fever or vomiting | 75% | 66% | 10% | 98% | 20% | 96% |

[a]Assumed sensitivity and specificity values are taken from Victora et al. (15).

doi:10.1371/journal.pmed.1001385.t001

also been introduced into the category of home fluids without evidence to suggest a benefit [22].

Current surveys ask about ORS, "recommended home fluids," and the quantity of additional fluids, which includes plain water. However, the adaptation and translation of survey instruments for local use rarely inserts a description of locally recommended fluids containing a salt and starch into questions about "recommended home fluids." Instead, many surveys are implemented in the community with a direct translation of the phrase "recommended home fluids," which may generate meaningless responses. Furthermore, asking the caregiver about giving plain water and other fluids in separate questions does not provide information about the management of dehydration in the child.

In particular, the inclusion of the broad category of oral rehydration therapy, which includes feeding practices and the provision of nearly any liquid, as a valid indicator in surveys for the appropriate coverage of diarrhea treatment is questionable. The definition of this indicator has changed over time so looking at "oral rehydration therapy" in isolation of the definition describing the indicator does not allow accurate time trends to be described [25]. Moreover, with limited evidence suggesting a measured benefit of fluids other than ORS on diarrhea mortality, putting weight on this and other nebulous categories (for example, recommended home fluids) does not further our understanding of correct diarrhea treatment and will not do so until improvements in the denominator can help us differentiate which episodes of diarrhea can be managed with simple fluids and which episodes are in need of ORS for dehydration prevention and treatment.

We recommend, therefore, that questions about specific fluids other than ORS should be eliminated from large-scale household surveys. This will help to refocus attention on the importance of ORS and prevent confusion between appropriate treatments for preventing and treating dehydration.

### Recommendation 2. Future Surveys Should Focus on on the Behavior of "Offering" the Child Fluid and Food

In large surveys the caregiver is asked if the child was offered or given ORS, additional fluids, or food. It is often difficult to distinguish which question—was the child "given" (i.e., consumed) versus was the child "offered"—has been asked in the translated version(s) of the questionnaire, even though back translation is routinely undertaken to check the accuracy of translations. Indeed, in some languages, there might not even be words that adequately distinguish between these two concepts. Moreover, reports of surveys may not accurately recount what was asked of the caregiver or, in some instances, may use both terms [26]. We suggest that the objective of the question in large surveys should be whether or not the child was offered ORS, additional fluids, or food. This is the behavior that is in the control of the caregiver—sick children may not take all that is offered. Smaller-scale surveys may be able to ask about the intake of the child to further our understanding of caregiver behaviors that may influence the child's response to food and drink being offered during the illness, but large DHS and MICS surveys will not be able to include both sets of questions.

In many countries, it is still the cultural norm to restrict feeding during the diarrhea episode or to change the food offered to those that are perceived as easy to digest. It is often thought that "less in" will mean "less out." Current recommendations emphasize the importance of continued feeding during the diarrhea episode [8] to ensure the child receives adequate nutrition. Secondary questions that ask caregivers if the child was offered/given the same, more, or less fluids and food than usual are already included in DHS and MICS surveys but again, it is important to ensure that these questions are consistently translated to ensure that the data gathered accurately reflects the concept of food and fluid offered and does not get confused with quantities eaten or drank. Quantifying exact quantities of food or fluid ingested is difficult to ascertain correctly from a household survey; therefore the concept of offered fluids and foods remains the best indicator of progress for programmatic consideration [27,28]. Furthermore, although a better understanding of changes in feeding practices with regard to what foods a child is offered during diarrhea is needed to ensure that appropriate nutrition is maintained during the episode, particularly in children who have longer episodes or multiple episodes, we believe it is inappropriate to include any other questions related to feeding practices in large surveys; instead smaller research studies should study feeding practices during diarrhea episodes to identify populations most at risk of malnutrition because of diarrhea.

### Recommendation 3. Future Surveys Should Separate Breastfeeding from Other Fluids and Foods

Current questionnaires include breast milk as part of the question dealing with the quantity of fluids offered. However, the quantity of breast milk cannot be measured if the child is fed at the breast. A mother can only report frequency of nursing opportunities or time spent nursing and changes in nursing patterns. We suggest that a separate question that asks if the child was given opportunities to nurse the same, more, or less frequently and/or for more, the same, or less time during the illness would allow the mother to more accurately describe the child's nursing pattern in lieu of providing information on the quantity of breast milk ingested.

## Coverage of Zinc Supplementation in DHS and MICS Surveys

Although zinc supplementation was added to the UNICEF/WHO recommendation in 2004, it has still not been incorporated widely into diarrhea treatment programs [6,8]. Indicators of zinc supplementation have been added to the routine DHS and MICS surveys but little testing has been done to determine whether caregivers can differentiate zinc from more commonly prescribed antibiotics and antidiarrheals in communities where zinc is a new treatment, particularly when recalling an episode in the past 2 weeks that has resolved and for which treatment is no longer being given. Caregivers may be able to more accurately recall treatments given if the recall period is shortened. A shorter recall period would result in a greater number of children currently receiving treatment and thus enable the surveyor to ask to see the packaging of the treatment being given to validate the caregiver's response. On the downside, a shorter recall period would limit the sample size and prevent opportunities for stratified analyses. As mentioned earlier in the context of reducing the recall period to improve reporting of diarrhea severity, research is needed to test whether shorter recall periods improve the accuracy of caregiver reports of zinc supplementation. Notably, few studies to date have examined the appropriate recall time for treatments given for childhood diseases. Additional research is also needed to better understand other aspects of the ability of caregivers to correctly identify and recount giving zinc for the diarrhea treatment, especially during the early years of introducing zinc into routine clinical practice.

## Conclusions

Current diarrhea coverage indicators seek to capture the coverage of zinc treatment and fluid replacement for the

## Box 2. Recommended Changes to Current Diarrhea Coverage Indicators

*Recommended changes requiring no additional research*

1. Eliminate questions with regard to recommended home fluids and renew focus on ORS.
2. Ensure questions in DHS and MICS that focus on "offering" a child fluids/food are clearly differentiated from questions that concern fluid/food intake.
3. Separate breastfeeding from other fluids and foods in all questions.

*Recommended changes requiring field-testing or additional research*

1. Develop and test the validity of including selected diarrhea severity questions as part of large household surveys.
2. Determine if 1-week recall can accurately collect information with regard to diarrhea severity in addition to basic coverage data.
3. Determine if caregiver recall of zinc supplementation in both tablet and syrup forms is accurate in a 2-week recall questionnaire.

## Key Points

- Diarrhea morbidity and mortality remain important child health problems in low- and middle-income countries.
- Additional research is needed to determine the most appropriate signs and symptoms to include in household surveys to enable the distinction between mild and more severe diarrhea to be made.
- Simplification of current surveys to renew/increase the focus on oral rehydration salts, on the behavior of offering a child fluids and food throughout their illness, and on breastfeeding should improve the measurement of treatment coverage.
- Zinc for the treatment of diarrhea is new and research is needed to understand the ability of caregivers to recognize and accurately recall zinc treatment for diarrhea
- By improving coverage indicators, we will increase our understanding of trends in diarrhea morbidity and mortality and be able to target interventions better.

prevention and treatment of dehydration. Because many of the current indicators have not changed over time, tracking coverage over time is possible. Nevertheless, in this paper, we present several opportunities for improving upon currently accepted indictors. We recognize that the scientific evidence supporting the methods for improvements we propose is lacking in most cases (Box 2). For this reason, this Review should be viewed as a call to action rather than as a guide to changes that should be made immediately.

Our motivation for suggesting changes to coverage indicators is driven by a need to better understand who and what types of diarrhea are being treated appropriately and to identify the episodes of diarrhea most likely to progress to dehydration. Diarrhea remains an important cause of morbidity and mortality for children in low- and middle-income countries around the world, and diarrhea treatment rates do not appear to be improving. However, it is possible that the children who need ORS most are now receiving adequate treatment and that this has contributed in part to the recent decline in diarrhea mortality [1].

Enhanced coverage indicators should enable us to better track changes in coverage over time. Importantly, however, because DHS/MICS surveys are already long, additional questions should not be added without considerable reflection on the cost-benefit of the new data collected. The ability of the changes proposed here to improve the measurement of treatment coverage for diarrhea must also be carefully studied before inclusion in routine DHS/MICS surveys to ensure that these surveys remain valuable tools without becoming too burdensome for practical use.

Finally, we acknowledge that making the changes we propose will inevitably mean that new indicators will not be comparable to past indicators. We suggest that this limitation may be a reasonable price to pay for improvements in our understanding of how well programs are targeting the children and the diarrhea episodes most in need of appropriate treatment.

## Acknowledgments

The authors would like to thank the Child Health Epidemiology Reference Group (CHERG)/Coverage Indicator Working Group for comments and feedback during the manuscript preparation and finalization.

## Author Contributions

Wrote the first draft of the manuscript: CLFW. Contributed to the writing of the manuscript: REB OF. ICMJE criteria for authorship read and met: CLFW REB OF. Agree with manuscript results and conclusions: CLFW REB OF.

## References

1. Liu L, Johnson HL, Cousens S, Scott S, Lawn JE, et al. (2012) Global, regional, and national causes of child mortality in 2000–2010: an updated systematic analysis. Lancet. doi:10.1016/S0140-6736(12)60560-1.
2. Fischer Walker CL, Perin J, Aryee MJ, Boschi-Pinto C, Black RE (2012) Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. BMC Public Health 12: 220.
3. Patel MM, Glass R, Desai R, Tate JE, Parashar UD (2012) Fulfilling the promise of rotavirus vaccines: how far have we come since licensure? Lancet Infect Dis 12: 561–570.
4. Black RE, Fischer Walker CL, Lanata CF (2012) Diarrheal diseases. Nelson KE, Williams C, editors. Infectious disease epidemiology: theory and practice. Sudbury (Massachusetts): Jones & Bartlett Learning.
5. Taylor CE, Greenough WB, 3rd (1989) Control of diarrheal diseases. Ann Rev Public Health 10: 221–244.
6. Santosham M, Chandran A, Fitzwater S, Fischer-Walker C, Baqui AH, et al. (2010) Progress and barriers for the control of diarrhoeal disease. Lancet 376: 63–67.
7. Ryland S, Raggers H (1998) Childhood morbidity and treatment patterns. DHS comparative studies No 27. Calverton (Maryland): Macro International.
8. WHO/UNICEF (2004) Joint Statement: Clinical Management of Acute Diarrhoea (WHO/FCH/CAH/04.07). Geneva and New York: World Health Organization, Department of Child and Adolescent Health and Development, and United Nations Children's Fund, Programme Division.
9. Rutstein SGR (2006) Guide to DHS statistics. Calverton (Maryland): ORC Macro.
10. Ahmed S, Ali D, Bisharat L, Hill A, LaFond A, et al. (2009) Evaluation of UNICEF Multiple Indicator Cluster Surveys Round 3 (MICS3) Final Report. Boston: JSI.
11. Hancioglu A, Arnold F (2013) Measuring coverage in MNCH: Tracking progress in health for women and children using DHS and MICS household surveys. PLoS Med 10: e1001391. doi:10.1371/journal.pmed.1001391.
12. Becker S, Pullman T (2007) External evaluation of the Peru continuous survey experiment (Report No. 07-001-48). Washington (D.C.): USAID.
13. US Global Health Initiative (2011) The United States Global Health Strategy: Senegal Global Health Initiative Strategy. Washington (D.C.): US Global Health Initiative.
14. Kendall C, Foote D, Martorell R (1984) Ethnomedicine and oral rehydration therapy: a case study of ethnomedical investigation and program planning. Soc Sci Med 19: 253–260.

15. Victora CG, Kirkwood BR, Fuchs SC, Lombardi C, Barros FC (1990) Is it possible to predict which diarrhoea episodes will lead to life-threatening dehydration? Int J Epidemiol 19: 736–742.
16. Herman E, Cutts F, Soares A (1991) Assessing diarrhoea-induced dehydration at the community level. J Trop Med Hyg 94: 224–230.
17. Zodpey SP, Deshpande SG, Ughade SN, Kulkarni SW, Shrikhande SN, et al. (1999) A prediction model for moderate or severe dehydration in children with diarrhoea. J Diarrhoeal Dis Res 17: 10–16.
18. Fuchs SC, Victora CG, Martines J (1996) Case-control study of risk of dehydrating diarrhoea in infants in vulnerable period after full weaning. BMJ 313: 391–394.
19. Patel AB, Ovung R, Badhoniya NB, Dibley MJ (2012) Risk factors for predicting diarrheal duration and morbidity in children with acute diarrhea. Indian J Ped 79: 472–477.
20. Alam N, Henry FJ, Rahaman MM (1989) Reporting errors in one-week diarrhoea recall surveys: experience from a prospective study in rural Bangladesh. Int J Epidemiol 18: 697–700.
21. Schmidt WP, Arnold BF, Boisson S, Genser B, Luby SP et al. (2011) Epidemiological methods in diarrhoea studies – an update. Int J Epidemiol 40: 1678–1692.
22. Munos MK, Fischer Walker CL, Black RE (2010) The effect of oral rehydration solution and recommended home fluids on diarrhoea mortality. Int J Epi 39: i75–i87.
23. Kielmann AA, McCord C (1977) Home treatment of childhood diarrhea in Punjab villages. J Trop Ped Env Ch H 23: 197–201.
24. Ismail R, Pardede N, Darwin S, Nazir M, Mukti S (1986) Home-made rice water salt solution for oral rehydration therapy–a field trial. J Diarrhoeal Dis Res 4: 20–25.
25. Victora CG, Bryce J, Fontaine O, Monasch R (2000) Reducing deaths from diarrhoea through oral rehydration therapy. Bull World Health Organ 78: 1246–1255.
26. Zimbabwe National Statistics Agency (2012) Zimbabwe: DHS, 2010–11 - final report (English). Calverton (Maryland): Zimbabwe and ICF International Inc.
27. Stanton BF, Rowland MG, Clemens JD (1987) Oral rehydration solution–too little or too much? Lancet 1: 33–34.
28. Zaman K, Sack DA, Chakraborty J, Yunus M, Baqui AH, et al. (2002) Children's fluid intake during diarrhoea: a comparison of questionnaire responses with data from observations. Acta Paediatr 91: 376–382.

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: Tracking Progress in Health for Women and Children Using DHS and MICS Household Surveys

**Attila Hancioglu[1]*, Fred Arnold[2]**

**1** Statistics and Monitoring Section, Division of Planning and Strategy, UNICEF, New York, New York, United States of America, **2** MEASURE DHS, ICF International, Calverton, Maryland, United States of America

**Abstract:** Household surveys are the primary data source of coverage indicators for children and women for most developing countries. Most of this information is generated by two global household survey programmes—the USAID-supported Demographic and Health Surveys (DHS) and the UNICEF-supported Multiple Indicator Cluster Surveys (MICS). In this review, we provide an overview of these two programmes, which cover a wide range of child and maternal health topics and provide estimates of many Millennium Development Goal indicators, as well as estimates of the indicators for the Countdown to 2015 initiative and the Commission on Information and Accountability for Women's and Children's Health. MICS and DHS collaborate closely and work through interagency processes to ensure that survey tools are harmonized and comparable as far as possible, but we highlight differences between DHS and MICS in the population covered and the reference periods used to measure coverage. These differences need to be considered when comparing estimates of reproductive, maternal, newborn, and child health indicators across countries and over time and we discuss the implications of these differences for coverage measurement. Finally, we discuss the need for survey planners and consumers of survey results to understand the strengths, limitations, and constraints of coverage measurements generated through household surveys, and address some technical issues surrounding sampling and quality control. We conclude that, although much effort has been made to improve coverage measurement in household surveys, continuing efforts are needed, including further research to improve and refine survey methods and analytical techniques.

*This paper is part of the* PLOS Medicine "*Measuring Coverage in MNCH" Collection.*

## Introduction

Considerable progress has been made since the mid-1990s in reducing maternal and child mortality [1]. However, there are still many unnecessary deaths among women and children (about 287,000 maternal deaths and 7.6 million child deaths in 2010 [2]), even though effective and affordable interventions are available. Tracking coverage indicators for women and children is necessary to guide global, regional, and country efforts to improve health so that scarce resources are directed to where they are most needed and will be most effective in saving lives.

Coverage measurements reflect the proportion of individuals needing an intervention, and must therefore be representative of the reference population. Two recent reviews assessed potential sources of coverage data for reproductive, maternal, newborn, and child health (RMNCH) interventions in the 75 countries that account for over 95% of maternal and child deaths [2,3]. These reviews concluded that although routine data from health management information systems are the preferred source of coverage data because they provide information on a continuous basis at lower administrative levels such as districts, these systems are currently too weak in these countries to provide data of adequate quality for assessing and guiding health programmes. Demographic surveillance systems often produce higher quality data but for limited geographic areas that become progressively less representative of national populations over time if health intervention trials are conducted in these areas. Both reviews identified high-quality, nationally representative household surveys as the method of choice for measuring RMNCH coverage for the foreseeable future in most low- and middle-income countries. Importantly, even after health management information systems become reasonably complete and accurate, national household surveys will need to continue as a complementary data source since these surveys are representative of the general population and provide vital information on background characteristics and the determinants of population and health conditions. Household surveys are also needed for measuring inequalities in coverage [4].

A large majority of household surveys that have produced coverage estimates in low-income countries have been conducted

**Citation:** Hancioglu A, Arnold F (2013) Measuring Coverage in MNCH: Tracking Progress in Health for Women and Children Using DHS and MICS Household Surveys. PLoS Med 10(5): e1001391. doi:10.1371/journal.pmed.1001391

**Academic Editor:** Nyovani Madise, Professor of Demography and Social Statistics, University of Southampton, United Kingdom

**Published** May 7, 2013



**Funding:** This work was conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF, with financial support from The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** AIS, AIDS Indicator Surveys; DHS, Demographic and Health Surveys; MICS, Multiple Indicator Cluster Surveys; MIS, Malaria Indicator Surveys; RMNCH, reproductive, maternal, newborn, and child health

* E-mail: ahancioglu@unicef.org

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

under the USAID-supported Demographic and Health Surveys (DHS) [5] and the UNICEF-supported Multiple Indicator Cluster Surveys (MICS) programmes [6]. For example, in the global databases compiled by UNICEF, information on the use of oral rehydration therapy for children with diarrhoea comes from DHS and MICS surveys for 98% of the countries with available data. For care seeking for pneumonia, the comparable figure is 93%. DHS and MICS data have also made a major contribution to the scientific literature. According to a recent study and data from the DHS website, more than 1,100 articles based entirely or primarily on DHS data have been published in 346 peer-reviewed journals [7]. Both programmes provide free public access to survey reports and datasets.

Other standardized household survey programmes that have generated coverage data on selected indicators include the CDC-supported Reproductive Health Surveys conducted mostly in Latin America, Eastern Europe, and Central Asia between 1975 and 2009 [8], the Pan-Arab Population and Family Health Surveys (PAPFAM), supported by the Arab League and conducted in the Arab region [9], and the WHO-supported World Health Survey conducted from 2002–2004 [10]. There are also standard household surveys focused on individual diseases or intervention programmes such as the Malaria Indicator Surveys [11], the AIDS Indicator Surveys [12], and the Standardized Monitoring and Assessment of Relief and Transitions (SMART) surveys conducted in many sub-Saharan African countries in the early 2000s to provide information related to child nutrition, some of which are still being conducted on an annual or semi-annual basis [13]. Also, a few countries have mounted their own nationally representative surveys, usually based on adapted versions of the DHS and MICS protocols. Over time and as a result of arduous consultations at the global level, a set of standard indicators and "gold-standard" methodologies have emerged that are incorporated in DHS and MICS surveys, as well as in some more specialized surveys.

The methodological challenges of measuring RMNCH coverage received relatively little attention in the literature until recently, in contrast to the measurement of mortality [14]. Much of the testing of alternative coverage indicators, questions, and analytical techniques is available only in internal reports of work conducted by DHS or in the heads of the technical experts who have conducted the surveys. In this review, which is part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, we draw on the DHS and MICS experience to highlight key methodological principles and challenges in using household survey data to measure RMNCH coverage. For more details, we direct readers to resource documents on survey design including sampling, questionnaires, data cleaning, and analysis as well as reports [5,6,15]. Other reviews and research articles in this collection will focus on measurement challenges in tracking trends in coverage for interventions targeting specific health conditions [16–21] and cross-cutting methodological issues [22]. Here, we complement this content by providing insights to help improve survey measurements of RMNCH coverage and to promote the informed use of coverage results to improve programmes.

## Overview of the DHS and MICS Survey Programmes

The DHS programme has been operating since 1984 with core funding from USAID and substantial contributions from other donors and participating countries. The programme is coordinated by ICF International. Its aim is to provide high-quality nationally representative data on health and population trends, with emphasis on fertility, family planning, mortality, reproductive health, child health, gender-related issues such as domestic violence, HIV/AIDS, malaria, and nutrition.

The MICS programme was developed by UNICEF in 1995 in response to the World Summit for Children and has expanded over time to measure progress towards the Millennium Development Goals and other international targets for women and children. MICS surveys provide key information on mortality, health, nutrition, education, HIV/AIDS, and child protection for use in programme decision making, advocacy, and national and global reporting.

Figure 1 shows the number of DHS and MICS surveys conducted annually since 1984. About ten to 15 DHS surveys have been conducted annually since 1995; consultative processes have led to major revisions in the core questionnaires every 5 years. MICS surveys were conducted in "rounds," every 5 years until 2007 and every 3 years thereafter, with about 60 surveys in each round.

The core DHS [23] and MICS [24] questionnaires have expanded over time, in content and complexity, to respond to country and global needs and the growing number of effective RMNCH interventions. Both surveys now include an increasing number of optional modules and complementary data collection tools for use in individual countries. Table 1 summarizes the characteristics of both survey programmes, including questionnaire content, which is typically decided through collaboration among government agencies, donors, key stakeholders, and DHS/MICS. Table 2 lists the major differences between DHS and MICS surveys and Box 1 highlights a particularly important difference between DHS and MICS surveys—the way in which they handle information on orphans and foster children.

Since their inception, DHS and MICS surveys have played an important role in shaping the global agenda on tracking coverage and in populating global databases. They have also influenced policies and intervention strategies. For example, DHS/MICS data are often used to establish targets in national economic and social development plans, to provide advocacy for programmes to improve women's and children's health, and to assist programmes in identifying target groups in most need of interventions. The role that these data play at the national and international level make it imperative that data quality is the foremost consideration when designing surveys and providing estimates of key indicators. In the following sections, therefore, we draw on DHS and MICS experience to highlight the challenges associated with measuring coverage through household surveys.

## Basic Principles and Survey Design

Valid measurement of coverage requires, first and foremost, representative population samples based on scientific probability sampling. We will discuss this essential aspect of coverage measurement in the next section of our review but first we will present three other key considerations that need to be taken into account when planning and conducting household surveys and when using their results.

First, some information is simply not amenable to collection through a household survey, because respondents do not know or cannot recall the required information. For example, a parent cannot know a child's birthweight if the child was not weighed at birth and is unlikely to remember the exact timing of tetanus toxoid vaccinations received over a lifetime. Decisions about what questions respondents can answer with acceptable accuracy are generally made by survey designers in consultation with technical experts; the research papers in this collection provide some of the

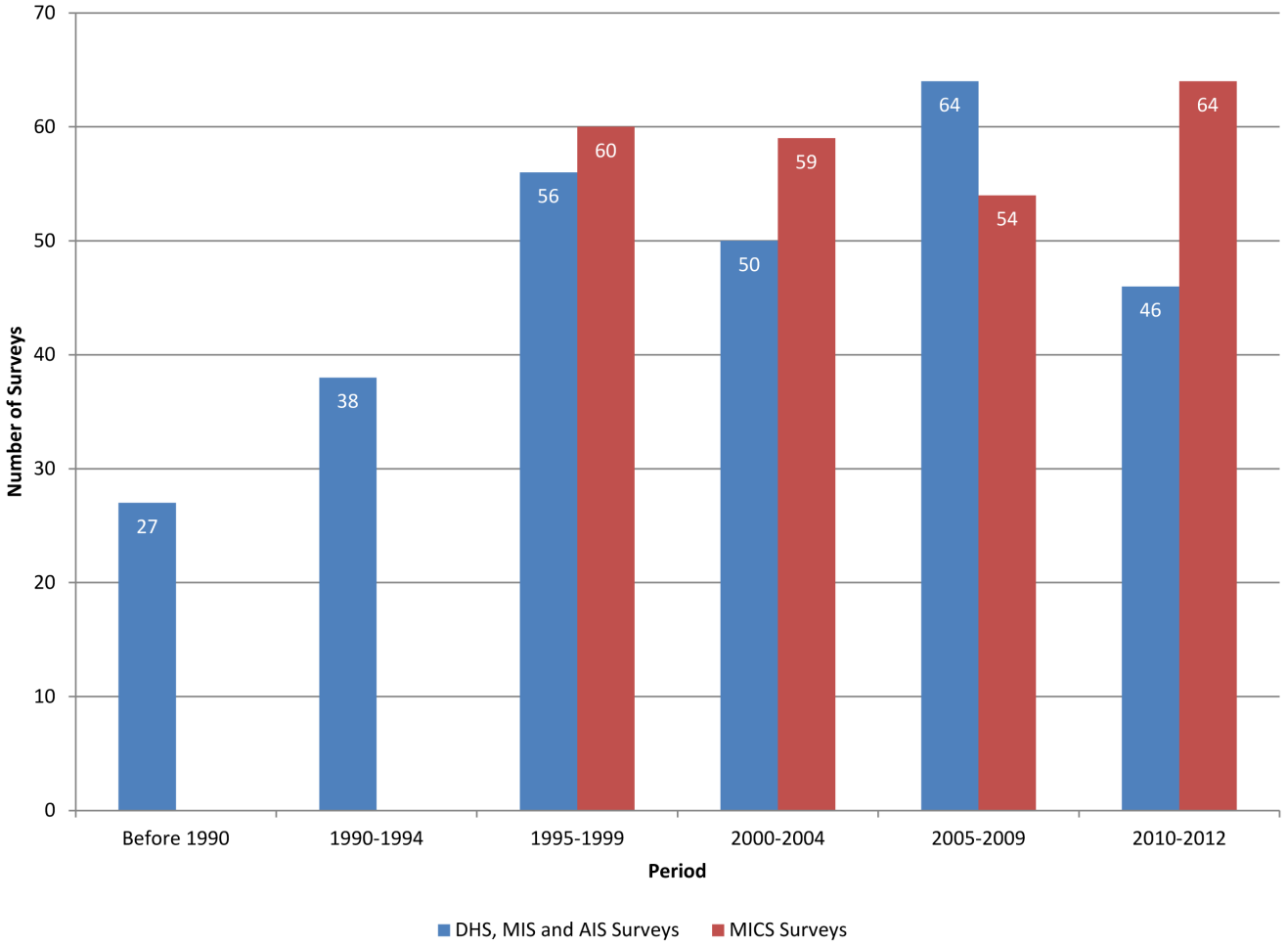


**Figure 1. Number of DHS and MICS surveys by year.**
doi:10.1371/journal.pmed.1001391.g001

first rigorous research assessing the validity of self-reported exposure to RMNCH interventions [17,18,25–28]. For coverage questions that cannot be answered through household surveys, alternative methods should be explored.

Second, survey interview length needs to be taken into account. In principle, a survey interview should be long enough to engage the respondent and obtain complete answers to the survey questions without rushing, but not so long that respondents become bored or frustrated. Completion of all questionnaires in a household averages around 2 hours for DHS and MICS surveys. Technical staff at DHS and MICS are concerned that the current protocols—especially when optional modules, biomarkers, and country-specific questions are included—are approaching lengths that adversely affect data quality. The need to keep survey interviews to a reasonable length is an important source of tension when discussing the addition of new questions with stakeholders.

Third, response rates are a concern in household surveys as low response rates can adversely affect the representativeness of the interviewed sample. Efforts need to be made to ensure that non-response is as low as possible. Non-response is well below 10% in DHS and MICS surveys for most countries. Even though DHS and MICS protocols specifically prohibit the substitution of households, response rates remain high thanks to training protocols emphasizing multiple revisits to households and close monitoring of response rates by field staff.

## Survey Sampling

The sample design determines the representativeness of household survey results, which is required to produce coverage estimates for the general population. Textbooks have been written about survey sampling (e.g., [29,30]) and we will not attempt to reduce them to a few paragraphs here. We recommend strongly that all survey planning teams should include an experienced technical sampling expert. Moreover, it is important not to underestimate the time involved in doing sampling properly. DHS [31] and MICS [32] surveys adhere to the fundamentals of scientific sampling, including complete coverage of the target population, use of suitable sample sizes, the need to conduct a new household listing and pre-selection of sample households, and preparation of appropriate sample documentation. However, deviations from standard procedures are sometimes required owing to cost limitations and practical considerations, including security concerns. Importantly, coverage levels in household surveys that use uncontrolled non-probability sampling (such as quota sampling or purposive sampling) do not provide valid estimates of population coverage.

DHS and MICS sampling frames are limited to the population residing in fixed households, and exclude populations living in group quarters (such as military barracks, hospitals, and hotels) and persons living on the street. This is a potential source of bias,

## Box 1. Orphans and Foster Children

An important difference between MICS and DHS surveys is in the collection of information on children under 5. In MICS surveys, information on children under 5 is collected from mothers or primary caregivers of children under 5 in the household, making it possible to collect information on all children under 5 (including orphans and foster children), regardless of whether their biological mothers are in the same household. In DHS surveys, the bulk of information on children under 5 is collected from their biological mothers in the Woman's Questionnaire. Therefore, information on some coverage indicators is not collected for children under 5 who are orphaned or not living with their biological mothers. An exception is the collection of anthropometric measurements and biomarkers for all children under 5 in the household, regardless of the survival status or whereabouts of their biological mothers.

In a recent analysis of 12 DHS and MICS surveys in 12 countries (unpublished), we found that the inclusion or exclusion of orphaned and fostered children does not have a substantial influence on national estimates of undernutrition. In three of the surveys, there was more than a 5 percentage point difference in the prevalence of underweight between children living with their biological mothers and other children. However, in seven of the surveys, the difference was less than 2 percentage points. In all 12 surveys, the difference between the *national* estimate of the prevalence of underweight for all children under 5 and for children whose mothers were interviewed was negligible. The differentials may vary for other indicators, so additional research would be desirable.

because those populations are likely to have different characteristics and health conditions from those living in households.

One very practical lesson learned through the DHS and MICS experience is that sample households should be selected in the central office, not in the field, if possible. Only preselected households should be eligible for interviewing, with no substitutions allowed during the fieldwork. This prevents interviewing teams from reducing their workload by avoiding listing or interviewing large or more remote households.

A key sampling question is how many households/individuals should be interviewed per sample cluster. DHS recommends a sample "take" of about 30–40 women per rural cluster and 20–25 women per urban cluster. MICS recommends 15–30 households per cluster, although detailed analyses of available budgets, logistical limitations, survey content, and information on sampling errors from previous MICS and DHS surveys are also undertaken on a case-by-case basis.

Currently, DHS and MICS surveys include around 15,000 and 10,000 households respectively, which is a sufficient sample size to produce statistically reliable estimates of most indicators at the national, urban–rural, and regional levels, but not at lower administrative levels, such as districts, slums, and small population groups. However, both survey programmes can oversample population groups or geographical areas to produce statistically valid estimates, when needed. Some DHS and MICS surveys have large enough sample sizes to produce estimates at lower administrative levels, but the added value of producing coverage estimates at these administrative levels with larger samples needs to be weighed against the added logistical and management challenges.

Unfortunately, the use of sampling weights is often misunderstood, and consequently misused or ignored. Some strata (such as urban areas or selected regions or provinces) are often oversampled to ensure that the final dataset includes enough cases to produce reliable results. In these cases, sampling weights need to be applied to account for varying design weights and non-response levels. Generally, analyses of survey data should use sampling weights calculated for each interviewed household or individual respondent. There is a lack of consensus, however, about whether weights should be used in multivariate analyses, and the decision is often based on the purpose of the analysis [33,34].

## Comparability of Measurement over Time and between DHS and MICS

Tracking changes in RMNCH coverage over time is a valuable way to assess progress and is one of the most widespread applications of DHS/MICS data. But there are many pitfalls, and often coverage estimates are taken from survey reports and plotted to show changes without sufficient attention being paid to comparability. Notably, RMNCH coverage indicators change over time in response to modifications in policies and programmes and as lessons are learned about measurement and interpretation. Other papers in this collection report on multiple changes in indicators of diarrhoea case management since 1990 [19], more recent changes in treatment of childhood malaria [18], and postnatal care for mothers and infants [35]. Rather than abstracting coverage data from published survey reports, users should access the UNICEF childinfo.org database, in which all standard global indicators have been checked for comparability and recalculated where necessary. If users obtain DHS and MICS data files themselves to recalculate coverage indicators over time, care must be taken to use standard indicator definitions and appropriate sampling weights.

MICS and DHS collaborate closely and work through interagency processes to ensure that their survey tools are harmonized and comparable and their data can, therefore, be combined in global databases covering a large majority of developing countries. The article in this collection by Requejo, Newby, and Bryce highlights the importance of using standard methods to produce comparable coverage data across countries [36]. Differences between DHS and MICS surveys that may affect RMNCH coverage estimates are presented in Table 2, which also summarizes the evidence on the possible magnitude of the effects of these differences. Finally, as mentioned earlier, Box 1 highlights a particularly important difference between DHS and MICS surveys—their inclusion or exclusion of information on orphans and foster children.

## Challenges in Survey Implementation

Accumulated experience from the two survey programmes underscores the importance of incorporating quality control mechanisms at every step in the survey process. Some data quality steps are described elsewhere [36]; here we focus on several common field problems that could affect the quality of RMNCH coverage estimates.

One of the major field problem concerns is related to interviewer training and supervision. Table 1 presents information about how certain aspects of survey implementation are dealt with in MICS and DHS surveys. Both programmes have minimum requirements for selecting interviewers (at least a high school diploma). Moreover, interviewers are not directly involved in the management/provision of health services to avoid potential conflicts of interest.

**Table 1.** Characteristics of the DHS and MICS survey programmes.

| Characteristics | DHS | MICS |
|---|---|---|
| **Content of ''core'' questionnaires and modules (2012)** | | |
| *Both surveys* | Fertility and family planning; infant and child mortality; maternal mortality; antenatal care (number of visits, provider, components of antenatal care, intermittent preventive treatment for malaria during pregnancy); delivery care (place of birth, delivery assistance, cesarean section, birth weight, birth size); postnatal care (postnatal care visits, timing of visits, type of provider); child protection (birth registration, child marriage); child feeding practices (prelacteal feed, breastfeeding, diet); child immunisation coverage; childhood fever, acute respiratory infections, diarrhoea (prevalence, care-seeking behaviour, place and type of treatment); children's living arrangements; malaria (ownership and use of mosquito nets, treatment of fever, indoor residual spraying against mosquitoes, malaria diagnosis); HIV (knowledge of transmission and prevention, prior testing, stigma, and discrimination); sexual behaviour; female genital cutting; environmental health (water, sanitation, handwashing, disposal of children's stools, cooking fuel); biomarkers (height, weight) | See DHS |
| *Primarily one survey* | Vitamin A supplementation, iron supplementation, sexually transmitted infections other than HIV (self reports, symptoms), exposure to second-hand smoke, biomarkers including tests for anaemia, HIV, and malaria, timing of antenatal care visits, domestic violence, fistula, women's empowerment | Child labour, child discipline, early child development, knowledge of danger signs for child illness |
| *Complementary protocols (2012)* | MIS, AIS, SPA surveys, KIS | |
| **Guidelines for survey implementation** | | |
| Length of interviewer training, including field practice | 4 weeks | 3 weeks |
| Composition of field teams | Supervisor, field editor, and four interviewers who are the same sex as the respondents | See DHS |
| | Health technician(s) for biomarker testing | Separate measurer for anthropometry |
| Software package used for primary data processing | CSPro | CSPro |
| Imputation and data analysis | CSPro | CSPro→SPSS |
| Preparation of report | In-country report writing workshop | Regional workshops, in-country support |
| Technical assistance | Technical assistance visits by ICF International | Regional workshops, in-country support, regional coordinators |
| **Characteristics of nationally representative surveys conducted in 2011** | | |
| Typical duration of fieldwork | 3–6 months | 2–4 months |
| Mean number of households | Around 15,000 households | Around 10,000 households |
| Average time between completion of data collection and release of the report | 3 months for Preliminary Report, 10–12 months for Final Report | 12–13 months for Final Report |
| **Other characteristics** | | |
| Free public access to datasets | www.measuredhs.com | www.childinfo.org |
| Easy access to survey results | STATcompiler | MICS Compiler |

AIS, AIDS Indicator Surveys; KIS, Key Indicator Surveys; MIS, Malaria Indicator Surveys; SPA, Service Provision Assessment Surveys.
doi:10.1371/journal.pmed.1001391.t001

Importantly, the MICS programme was originally intended to be a relatively quick and light exercise at the country level, with limited in-country technical support. The MICS programme was designed to produce the limited data needed to monitor progress towards the World Summit for Children Goals. Over time, however, concerns about quality issues associated with the expanded questionnaire length and increased sample sizes have led MICS to develop a technical support system that emphasizes the duration and content of interviewer training and field supervision, but also includes rigorous controls on data entry and checking. These changes bring the quality of MICS data more in line with that of DHS (which has always included substantial centralized technical support).

Another important quality assurance measure for both survey programmes concerns the entry of data for paper-based questionnaires simultaneously with the fieldwork. For both paper-based questionnaires and surveys conducted with personal data assistants (PDAs) or tablets, field check tables are produced while the fieldwork teams are still in the field, thereby making it possible to spot systematic errors in data collection and take measures to improve data quality.

However, despite meticulous attention to quality control, all household surveys are prone to quality problems, the most important of which we will now highlight. One of the most challenging aspects of implementing DHS and MICS protocols is how to ensure that these protocols are appropriately adapted to the country context. Despite the availability of technical guidelines and assistance, our experience suggests that problems frequently occur in this area. One clear example is the adaptation of response categories for diarrhoea treatment at home. The standard response category for ''government-recommended home fluids'' is often not customized to reflect country-specific recommendations, in many cases because of the lack of clear national policies or the long list of recommended fluids [19]. Another example is

**Table 2.** Differences between standard DHS and MICS protocols and their potential implications for coverage measurement.

| Characteristics | DHS | MICS | Potential Implications for Coverage Measurement |
|---|---|---|---|
| **Sampling and survey design** | | | |
| Sample size per cluster | Rural: 30–40 women; Urban: 20–25 women | 15–30 households | — |
| Construction of household rosters | All usual members of the household *plus* visitors who spent the previous night in the household. DHS tables on coverage measurement are based on de facto persons in the household (that is, persons who stayed in the household the previous night). | All usual members of the household (de jure household members) included. | De facto approach gives better representation of mobile populations. De jure approach is more consistent with selection probabilities based on censuses. Unlikely to lead to any bias, since response rates remain very high in both approaches. |
| Respondents for information about children less than 5 years of age | Biological mothers only except for anthropometric indicators and anaemia, which are collected for all children. | Mothers or primary caregivers of children under 5 living in the household. | Inclusion of caregivers means orphans and foster children are included in the samples for MNCH coverage estimates for MICS, and not for DHS. See Box 1 for implications for coverage measurement. |
| **Reference periods for selected MNCH coverage indicators** | | | |
| Skilled attendance at delivery | All births during the past 5 years | Last birth during the past 2 years | The advantage of a shorter reference period is that the coverage estimates refer to a more recent date; on the other hand, the sample size is reduced when the reference period is shorter, which increases the confidence intervals. |
| Antenatal care | Last birth during the past 5 years | Last birth during the past 2 years | See above |
| Tetanus toxoid | Last birth during the past 5 years | Last birth during the past 2 years | See above |
| Initial Breastfeeding | Last birth in the past 5 years | Last birth during the past 2 years | See above |
| Exclusive breastfeeding | Youngest child age 0–4 years living with the mother | All living children age 0–4 years | See above |
| Postnatal care | Last birth during the past 5 years | Last birth during the past 2 years | See above |
| Birth weight | All births in the past 5 years | Last birth during the past 2 years | See above |

MNCH, maternal, newborn, and child health.
doi:10.1371/journal.pmed.1001391.t002

described by Hazel and colleagues in this collection [21] who show that coverage gains attributable to recently adopted strategies for treating childhood illness at the community level cannot be assessed unless the response options for care seeking are adapted to reflect the country-specific providers of care at the community level. Ensuring that standard survey protocols are adapted to specific country contexts is a continuing challenge, especially given the numerous topics included in each survey and the time and resources needed to adapt questionnaires in each country.

Another implementation challenge relates to what demographers working on child mortality refer to as the tendency for reports of child deaths to "heap" at certain ages, notably 12 months [37]. If this tendency is also true for reports of service utilisation and receipt of interventions, it could affect the accuracy of coverage estimates with time-bounded effectiveness, such as the timing of antenatal or postnatal care. We are not aware of any research that investigates the effects of age heaping in coverage measurement, and believe that this warrants further attention.

Finally, a related problem is the tendency for interviewers to "transfer" children to age groups (especially over 5 years of age) that exclude them from lengthy portions of the interview and therefore reduce interviewer workload [37]. Even if only some interviewers modify children's ages in this way occasionally, it can affect coverage estimates, especially if the children whose ages are changed systematically have different characteristics from other children. This is a continuing problem in some surveys despite serious attempts to minimize abuses. Age displacement may have more of an effect on fertility and mortality estimates than RMNCH estimates, but displacement remains a matter of concern overall.

## Conclusions

Large-scale, nationally representative household surveys are the primary source of data on RMNCH coverage. Despite efforts to improve routine information systems, surveys are likely to remain the primary source of data for many years to come. As we discuss in this review, it is essential that survey planners and consumers of survey results understand the strengths, limitations, and constraints of coverage measurements generated through household surveys

## Key Points

- DHS and MICS surveys are the principal source of national-level data on maternal, newborn, and child health indicators in low- and middle-income countries.
- Despite efforts to improve routine information systems, household surveys are likely to remain the primary source of population data for the foreseeable future.
- To analyse coverage estimates over time and across countries from household survey data, it is essential that similar survey methods and questions be employed to ensure comparability.
- The strengths and weaknesses of all data collection efforts need to be transparent and well understood by data users.
- DHS and MICS will continue to benefit from research findings to make further improvements in the collection of reproductive, maternal and child health coverage data.

and also appreciate the technical issues involved in sampling and quality control. Moreover, as our review and the other articles in this collection highlight, conducting household surveys that generate valid and reliable information on coverage is a complex exercise. We believe that the findings of current and future experimental studies will help to inform continuing efforts to improve coverage measurement in household surveys, particularly in the areas of improved question wording and interviewer training. Finally, we stress that calls for more and better data on coverage must be accompanied by sufficient resources and by an ongoing research programme to continue to improve and refine methods and analytical techniques.


## Acknowledgments

We appreciate the support, valuable guidance, and incisive comments we received from Jennifer Bryce at every stage in the preparation of this paper. We are very grateful to Ann Blanc and Jennifer Requejo for their reviews and comments.


## Author Contributions

Analyzed the data: AH FA. Wrote the first draft of the manuscript: AH FA. Contributed to the writing of the manuscript: AH FA. ICMJE criteria for authorship read and met: AH FA. Agree with manuscript results and conclusions: AH FA.

## References


1. Requejo J, Bryce J, Victora C, the Countdown to 2015 Report Writing Team (2012) Building a future for women and children: the 2012 Report. Washington (D.C.): World Health Organization and UNICEF. Available: http://www.countdown2015mnch.org/documents/2012Report/2012-complete-no-profiles.pdf. Accessed 8 December 2012.
2. Commission on Information and Accountability for Women's and Children's Health (2011) Keeping promises, measuring results. Geneva: World Health Organization.
3. Countdown to 2015, HMN, UNICEF, WHO (2011) Monitoring maternal, newborn and child health: understanding key progress indicators. Geneva: World Health Organization.
4. Barros AJD, Victora CG (2013) Measuring coverage in MNCH: Determining and interpreting inequalities in coverage of maternal, newborn, and child health interventions. PLoS Med 10: e1001390. doi:10.1371/journal.pmed.1001390
5. MEASURE DHS Demographic and Health Surveys. Available: http://www.measuredhs.com/. Accessed 8 December 2012.
6. UNICEF. Statistics and Monitoring: Multiple Indicator Cluster Survey. Available: http://www.unicef.org/statistics/index_24302.html Accessed 8 December 2012.
7. Fabic MS, Choi YJ, Bird S (2012) A systematic review of Demographic and Health Surveys: data availability and utilization for research. Bull World Health Organ 90: 604–612.
8. CDC (2012) Information on the US Centers for Disease Control and Prevention Reproductive Health Surveys. Available: www.cdc.gov/reproductivehealth/Global/index.htm. Accessed 8 December 2012.
9. AidData (2011) Information on PAPFAM surveys. Available: http://www.aiddata.org/project/show/2433396. Accessed 8 December 2012.
10. WHO (2012) Information on the World Health Surveys. Available: http://www.who.int/healthinfo/survey/whscurrent/en/index.html. Accessed 8 December 2012.
11. Malariasurveys.org (2011) Information on Malaria Indicator Surveys. Available: www.malariasurveys.org. Accessed 8 December 2012.
12. MEASURE DHS (2012) Information about AIDS Indicator Surveys. Available: http://measuredhs.com/What-We-Do/Survey-Types/AIS.cfm. Accessed 8 December 2012.
13. SMART (2006) SMART. Measuring mortality, nutritional status and food security in crisis situations: SMART methodology. Version 1, April 2006. Available at http://smartindicators.org. Accessed 8 December 2012.
14. Mathers C, Boerma T (2010) Mortality measurement matters: improving data collection and estimation methods for child and adult mortality. PLoS Med 7: e1000265. doi:10.1371/journal.pmed.1000265.
15. United Nations Statistics Division (2005) Household sample surveys in developing and transition countries, studies in methods, series F number 96, United Nations publication, sales number 05.XVII.6. Available http://unstats.un.org/unsd/publication/seriesf/Seriesf_96e.pdf. Accessed 8 December 2012.
16. Campbell H, el Arifeen S, Hazir H, O'Kelly J, Bryce J, et al. (2013) Measuring coverage in MNCH: Challenges in monitoring the proportion of young children with pneumonia who receive antibiotic treatment. PLoS Med 10: e1001421. doi:10.1371/journal.pmed.1001421
17. Hazir T, Bequem K, el Arifeen S, Khan AM, Huque MH, et al. (2013) Measuring coverage in MNCH: A prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia. PLoS Med 10: e1001422. doi:10.1371/journal.pmed.1001422
18. Eisele T, Silumbe K, Yukich J, Hamainza B, Keating J, et al. (2013) Measuring coverage in MNCH: Accuracy of measuring diagnosis and treatment of childhood malaria from household surveys in Zambia. PLoS Med 10: e1001417. doi:10.1371/journal.pmed.1001417
19. Fischer Walker CL, Fontaine O, Black RE (2013) Measuring coverage in MNCH: Current indicators for measuring coverage of diarrhea treatment interventions and opportunities for improvement. PLoS Med 10: e1001385. doi:10.1371/journal.pmed.1001385
20. Cutts FT, Izurieta H, Rhoda DA (2013) Measuring coverage in MNCH: Design, implementation, and interpretation challenges associated with tracking vaccination coverage using household surveys. PLoS Med 10: e1001404. doi:10.1371/journal.pmed.1001404
21. Hazel E, Requejo J, David J, Bryce J (2013) Measuring coverage in MNCH: Evaluation of community-based treatment of childhood illnesses through household surveys. PLoS Med 10: e1001384. doi:10.1371/journal.pmed.1001384
22. Eisele T, Rhoda DA, Cutts FT, Keating J, Ren R, et al. (2013) Measuring coverage in MNCH: Total survey error and the interpretation of intervention coverage estimates from household surveys. PLoS Med 10: e1001386. doi:10.1371/journal.pmed.1001386
23. MEASURE DHS (2012) DHS core questionnaire. Available: http://www.measuredhs.com/publications/publication-DHSQ6-DHS-Questionnaires-and-Manuals.cfm. Accessed 8 December 2012.
24. Childinfo (2012) MICS core questionnaire. Available: http://www.childinfo.org/mics4_questionnaire.html. Accessed 8 December 2012.
25. Stanton CK, Rawlins B, Drake M, dos Anjos M, Cantor D, et al. (2013) Measuring coverage in MNCH: Testing the validity of women's self-report of key maternal and newborn health interventions during the peripartum period in Mozambique. PLoS ONE 8: e60694. doi:10.1371/journal.pone.0060694
26. Tunçalp Ö, Stanton C, Castro A, Adanu RK, Heymann M, et al. (2013) Measuring coverage in MNCH: Validating women's self-report of emergency cesarean sections in Ghana and the Dominican Republic. PLoS ONE 8: e60761. doi:10.1371/journal.pone.0060761
27. Guo Y, Liu L, Li M, Yang L, Ju L, et al. (2013) Measuring coverage in MNCH: A validation study linking population survey derived coverage to maternal, newborn, and child health care records in rural China. PLoS ONE 8: e60762. doi:10.1371/journal.pone.0060762
28. Stringer J, Stinson K, Tih PM, Giganti M, Ekouevi D, et al. (2013) Measuring coverage in MNCH: Population HIV-free survival among children under two years of age in four African countries. PLoS Med 10: e1001424. doi:10.1371/journal.pmed.1001424
29. Levy PS, Lemeshow S (2008) Sampling of populations: methods and applications. New York: John Wiley & Sons, Inc. 616 p.
30. Kish L (1965) Survey sampling. New York: John Wiley & Sons, Inc. 643 p.
31. ICF Macro (2012) Sampling manual for use with Demographic and Health Surveys. Calverton (Maryland): ICF Macro. Available: http://www.measuredhs.

com/publications/publication-DHSM4-DHS-Questionnaires-and-Manuals.cfm. Accessed 8 December 2012.
32. Childinfo (2012) http://www.childinfo.org/files/MICS3_Chapter_4_-_Designing_and_Selecting_the_Sample_060219.pdf.Accessed 8 December 2012.
33. Deaton A (1997) The analysis of household surveys: a microeconometric approach to development policy. Baltimore (Maryland): Johns Hopkins University Press. 479 p.
34. Woolridge JM (2002) Econometric analysis of cross section and panel data. Cambridge (Massachusetts): MIT Press. 752 p.
35. Moran A, Kerber K, Sitrin D, Guenther T, Morrissey CS, et al. (2013) Measuring coverage in MNCH: Indicators for global tracking of newborn care. PLoS Med 10: e1001415. doi:10.1371/journal.pmed.1001415
36. Requejo JH, Newby H, Bryce J (2013) Measuring coverage in MCNH: Challenges and opportunities in the selection of coverage indicators for global monitoring. PLoS Med 10: e1001416. doi:10.1371/journal.pmed.1001416
37. Pullum TW (2006) An assessment of age and date reporting in the DHS surveys, 1985–2003. DHS Methodological Reports number 5. Calverton (Maryland): Macro International Inc.

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: Determining and Interpreting Inequalities in Coverage of Maternal, Newborn, and Child Health Interventions

**Aluísio J. D. Barros*, Cesar G. Victora**

Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil

**Abstract:** To monitor progress towards the Millennium Development Goals, it is essential to monitor the coverage of health interventions in subgroups of the population, because national averages can hide important inequalities. In this review, we provide a practical guide to measuring and interpreting inequalities based on surveys carried out in low- and middle-income countries, with a focus on the health of mothers and children. Relevant stratification variables include urban/rural residence, geographic region, and educational level, but breakdowns by wealth status are increasingly popular. For the latter, a classification based on an asset index is the most appropriate for national surveys. The measurement of intervention coverage can be made by single indicators, but the use of combined measures has important advantages, and we advocate two summary measures (the composite coverage index and the co-coverage indicator) for the study of time trends and for cross-country comparisons. We highlight the need for inequality measures that take the whole socioeconomic distribution into account, such as the relative concentration index and the slope index of inequality, although simpler measures such as the ratio and difference between the richest and poorest groups may also be presented for non-technical audiences. Finally, we present a framework for the analysis of time trends in inequalities, arguing that it is essential to study both absolute and relative indicators, and we provide guidance to the joint interpretation of these results.

*This paper is part of the* PLOS Medicine *"Measuring Coverage in MNCH" Collection*

## Introduction

Equity in health has been part of the public health agenda for quite some time in the US, Europe, and Latin America [1–3], but interest in health inequities has boomed since the 1990s, with a large number of publications considering definitions [3], measurement [4–6], and controversies about health inequalities [7,8] (throughout this review we will refer to equity when we are considering the concept of fairness/justice and inequality when we are considering the measurement of differences in coverage, which are used to make judgments about equity/inequity). Interest in health equity has also started to increase in low- and middle-income countries, where few analyses of inequalities were available prior to 2000 [9–11]. Although reducing inequalities was not a key element in the health-related Millennium Development Goals, it is an important focus of the post-2015 agenda, which involves studying how inequalities change, how they relate to policies and health systems, and how they relate to global processes, such as conflict or economic growth or recession [12]. The need to make a clear link between broad social and economic inequalities and disparities in the coverage of health interventions has also been championed by the Social Determinants of Health movement [13,14].

In spite of recent developments, descriptive cross-sectional studies of health inequalities are still the most common and useful type of study for the design and implementation of public health policies aimed at improving equity. Such studies require the measurement, presentation, comparison, and interpretation of inequalities in health. In this article, which is part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, we do not intend to provide a broad review of inequalities in health and their measurement, but rather we rely on our own recent experience in monitoring inequalities [15] to provide practical advice to researchers and policymakers in low- and middle-income countries on how to carry out and interpret such analyses. We discuss methodological issues relevant to these objectives, including the assessment of socioeconomic position, choice of outcome measures, measures of the degree of inequality, and assessment of changes in inequalities over time. The examples we include are derived from data analyses carried out for the World Health Organization Global Health Observatory [16] and for the Countdown to 2015: Maternal, Newborn and Child Survival initiative [17]. Our examples use primary data mainly collected by the Demographic and Health Surveys (DHS) [18] and Multiple

**Citation:** Barros AJD, Victora CG (2013) Measuring Coverage in MNCH: Determining and Interpreting Inequalities in Coverage of Maternal, Newborn, and Child Health Interventions. PLoS Med 10(5): e1001390. doi:10.1371/journal.pmed.1001390

**Academic Editor:** Nyovani Madise, Professor of Demography and Social Statistics, University of Southampton, United Kingdom

**Published** May 7, 2013



**Funding:** This work was conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF, with financial support from The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** CCI, composite coverage index; CIX, concentration index; DHS, Demographic and Health Survey/s; MICS, Multiple Indicator Cluster Survey/s; MNCH, maternal, newborn, and child health; RII, relative index of inequality; SII, slope index of inequality

* E-mail: abarros@equidade.org

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

Indicator Cluster Surveys (MICS) [19], large, population-based household surveys that have been carried out repeatedly in low- and middle-income countries since the 1990s. Comprehensive overviews of health inequalities based on data from these surveys are available elsewhere [20–22].

## Measuring Socioeconomic Position

There are multiple dimensions to health equity according to gender, wealth, education, place of residence, ethnicity, and sexual orientation, among other factors. In this article we focus on "socioeconomic position," a term that is now preferred over "socioeconomic status" in the equity literature [23].

Socioeconomic position can be ascertained using different types of indicators, each reflecting slightly—or sometimes markedly—different underlying constructs. From the standpoint of statistical analyses, an indicator should be easy to measure in a valid way during surveys, should not change rapidly over time, should allow breakdown into several categories (preferably of similar size), and should be comparable over time and place. No single measure of socioeconomic position fulfills all these criteria in a satisfactory way. Howe et al. have recently reviewed the advantages and limitations of socioeconomic position indicators in low- and middle-income countries [24], and we provide a brief discussion of four ways to measure socioeconomic position—education, income, household consumption, and occupation—in Box 1.

In light of the problems with the above socioeconomic position indicators, an alternative was proposed by Filmer and Pritchett in 1998 [25]—the asset index. This index is based on a relatively short list of household possessions (radio, television, refrigerator, etc.) and characteristics of the house (building materials, toilet, electricity, etc.), and may include the educational attainment of household members. These variables, which are collected in DHS and MICS surveys, are subjected to principal component analysis, a data reduction technique that produces a single continuous composite score from all the variables that retains as much variance as possible [26]. Each household is then assigned an asset score, and samples can be broken down into quintiles or other equal-sized groups of households based on this asset score.

Like all other socioeconomic position indicators, asset indices have limitations [27]. First, different choices of assets can change the classification of families [28,29]. Second, families in the wealthiest quintile in most low- and middle-income countries are mainly urban [30], so that wealth inequalities are closely associated with urban/rural disparities [31]. A third limitation is that quintiles assess only relative socioeconomic position, rather than absolute socioeconomic position [32]. For example, the poorest quintile in a middle-income country may be richer than one of the wealthier quintiles in a low-income country. A similar problem can arise in time-trend analysis for a country that is getting richer. Moreover, because fertility tends to be higher among the poor, there tend to be more than 20% of the mothers and children in the lowest quintile of household wealth, and fewer than 20% in the richest quintile. This effect is even more marked for disease episodes: data on oral rehydration therapy, for example, are often based on a much larger sample (which is dependent on the number of diarrhea episodes) in the poorest than in the wealthiest quintile.

These limitations do not, however, preclude the widespread and valid use of asset indices for documenting the wide gaps between rich and poor that are present in most low- and middle-income countries, as is evident by the consistent associations between asset indices and more complex measures of socioeconomic position [31] and by the marked inverse associations between asset indices and child mortality and undernutrition [20,21]. Compared to the other methods discussed above, the asset index has clear advantages, and this has resulted in its widespread international adoption as the preferred way of measuring socioeconomic position in low- and middle-income countries [20].

### Box 1. Common Measures of Socioeconomic Position

#### Education of the Mother

- Easy to measure but can also have a direct effect on health [12].
- Often results in unbalanced groups. In poor countries, a large proportion of women may have no education, whereas in wealthier countries most will have completed secondary school.
- Size of the categories will vary over time, as more women are educated, which affects the comparison of time trends.
- May be difficult to use in country comparisons because of different schooling structure, level names, and content.

#### Income

- Requires several questions to be asked about different sources of income.
- Misreporting is frequent, and monthly variability may be important in low-income societies where casual labor and agricultural production are common.
- More positively, income is a continuous variable that can be broken down in groups of uniform size, which allows comparisons over time.

#### Consumption Expenditure

- Reflects what people spend rather than what they earn.
- Difficult to measure, requiring respondents to keep diaries and to answer long questionnaires, and requiring multiple visits by interviewers.
- Affected by misreporting, seasonality, in-kind exchanges, and domestic production of goods [24,44,45].
- If properly measured, consumption expenditure is a useful indicator, but its practical limitations have so far restricted its use in health research in low- and middle-income countries.

#### Occupation

- Commonly used in high-income country studies, this measure of socioeconomic position is problematic in low- and middle-income countries, where changes in occupation and multiple jobs are common and unemployment or informal jobs predominate.
- Long questionnaires and complex post-processing are required to capture all the subtleties of occupation in low- and middle-income countries, where large proportions of the population may fall into a single category. "Farmers," for example, may include anyone from a landless laborer to a plantation owner.
- Several classifications used in countries are not ordinal, making it impossible to rank groups.

## Measuring Inequalities in Intervention Coverage

There are two basic approaches for measuring inequalities in intervention coverage. The first is to carry out separate analyses for each relevant coverage indicator, such as contraceptive use, presence of skilled birth attendant, measles vaccine coverage, oral rehydration therapy, etc. The Countdown to 2015 initiative [17] provides inequality breakdowns of 18 such indicators for 75 countries. There are at least two caveats with this approach. First, several coverage indicators are based on small subsets of mothers or children. For example, vaccination status is assessed among children aged 12–23 months, and oral rehydration therapy is assessed among children who presented with diarrhea in the two weeks before the interview. When these indicators are broken down by quintiles, the number of mothers and children surveyed who (a) belong to a given quintile and (b) also belong to the subgroup that constitutes the denominator of the coverage indicator is often very small, even in large surveys, which leads to poor precision due to sampling variability. The second caveat is the difficulty in summarizing inequalities for a given country, because the magnitude of inequality may vary by indicator. On the other hand, these analyses may provide insights about which delivery channels are most equitable, and therefore contribute to better programming. For example, a recent Countdown to 2015 publication compared inequalities in coverage with skilled birth attendants—which requires access to a functioning health system, 24 hours a day and seven days a week—and inequalities in measles vaccine coverage—which is usually delivered on a single occasion in communities, often during national immunization days. Not surprisingly, inequalities were much greater for coverage with skilled birth attendants than for coverage with measles vaccine [22].

To avoid the problems of studying one coverage indicator at a time, two related measures have been proposed that combine the coverage of several interventions (Box 2). The co-coverage indicator is important since it shows what percentage of the population is receiving all, or most, of the main preventive interventions. If children or mothers receive just a few of a set of lifesaving interventions, efforts to improve the health of children or mothers may have little effect [33]. The composite coverage index (CCI), on the other hand, provides an overall estimate of coverage based on eight essential health interventions. By replacing a large number of coverage estimates, it makes multi-country and time-trend assessments easier to carry out and to understand. Both measures have been widely used in cross-country comparisons and in equity analyses; the most recent Countdown to 2015 report contains several examples [34].

## How to Express the Magnitude of Inequalities

There is no consensus on the ideal measure for expressing the magnitude of inequalities. In 1991, Wagstaff et al. [6] identified six such measures. In 1997 Mackenbach and Kunst [4] listed 12 measures, and in 2005 Harper and Lynch [35] compiled more than 15 of them. The main dichotomy in the expression of the magnitude of inequalities relates to whether the measure is *absolute* or *relative*. An example of an absolute measure of inequality is the difference between the extreme wealth quintiles—for example, measles immunization coverage is 10 percentage points higher in the top wealth quintile than in the bottom quintile. A relative measure of inequality is based on a ratio—for example, vaccine coverage is 20%, or 1.2 times, higher in the richest quintile than in the poorest. The distinction between percentage points and percentages is essential. If vaccine coverage in the richest and poorest groups is 70% and 50%, respectively, the absolute difference in coverage will be equal to 20 percentage points, while the relative ratio will be 1.4 (i.e., 70%/50%), or 40% (i.e., $[1.4-1]\times100\%$).

Despite their simplicity, these measures, which take into account only the top (Q5) and bottom (Q1) quintiles of the population under study, have important limitations. First, these measures are sensitive to changes in the number of individuals in each stratification category. For example, the rich/poor ratio for coverage with skilled birth attendants based on the 2008 Nigeria DHS survey [36] is 10.4 if we use deciles of wealth and 8.8 if we use quintiles. Another limitation is that in some cases the lowest and highest wealth groups will not have the lowest and highest

### Box 2. Indicators That Combine Coverage of Several Interventions

#### Co-Coverage

- Based on how many preventative interventions each mother/child pair received, out of a set of 8–9 essential interventions [33]: antenatal care (1+ visit with skilled provider); tetanus toxoid during pregnancy; skilled birth attendant; child received vitamin A supplementation, BCG (tuberculosis) vaccination, DTP3 (diphtheria–tetanus–pertussis) vaccination, and measles vaccination; improved drinking water source. Insecticide-treated bednets are also included in countries where relevant.
- Calculation of co-coverage requires reanalysis of original survey data, which is time-consuming, but because co-coverage is measured at the individual level, standard errors and confidence intervals can be calculated.
- Co-coverage is often reported as the percentage of children covered by at least three or six interventions but can also be presented through stacked bar graphs that show the percentage of children in the population covered by a given number of interventions, usually stratified by wealth quintiles (Figure 1).

#### Composite Coverage Index (CCI)

- Based on the weighted average of coverage of a set of eight preventative and curative interventions; the CCI gives equal weight to four stages in the continuum of care: family planning, maternal and newborn care, immunization, and case management of sick children [27].
- The weighted average for a group (e.g., a country or a wealth quintile) is calculated as

$$\frac{1}{4}\left(\mathrm{FPS}+\frac{\mathrm{SBA}+\mathrm{ANCS}}{2}+\frac{2\mathrm{DPT3}+\mathrm{MSL}+\mathrm{BCG}}{4}+\frac{\mathrm{ORT}+\mathrm{CPNM}}{2}\right)$$

  where FPS is family planning needs satisfied, SBA is skilled birth attendant, ANCS is antenatal care with skilled provider, DPT3 is three doses of diphtheria–pertussis–tetanus vaccine, MSL is measles vaccination, BCG is BCG (tuberculosis) vaccination, ORT is oral rehydration therapy for children with diarrhea, and CPNM is care seeking for pneumonia.
- Because the CCI is a group indicator, jackknife or similar resampling methods are required to estimate its standard error [46].

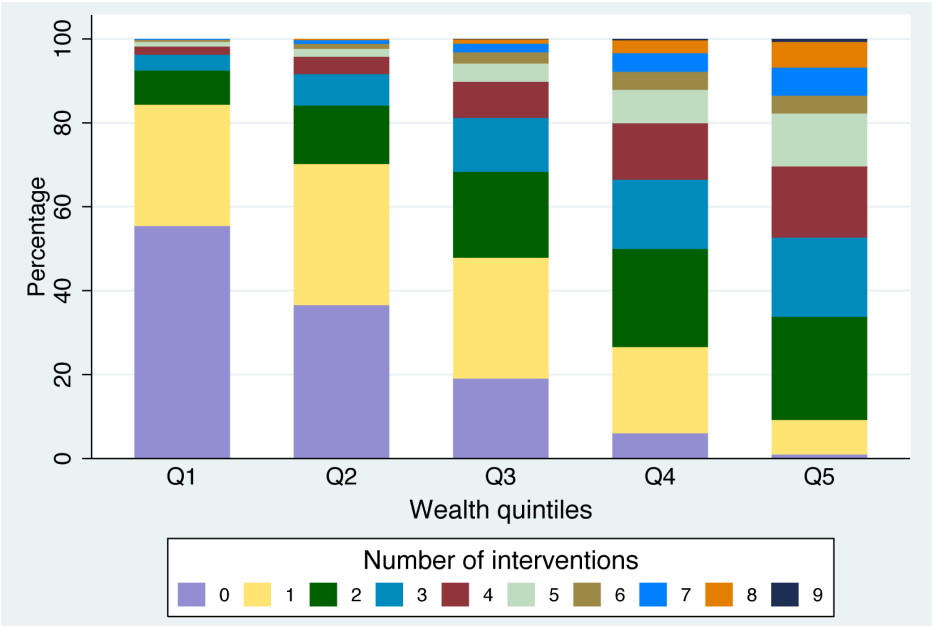


**Figure 1. Co-coverage of nine preventive interventions for Nigeria (DHS 2008), by wealth quintiles.** See Box 2 for more information on the interventions included.
doi:10.1371/journal.pmed.1001390.g001

coverage levels, particularly when overall coverage is high. This is the case for measles vaccine coverage in Bolivia (coverage of 75% in Q1 and 67% in Q5) and Tajikistan (coverage of 88% in Q4 and 84% in Q5) [22].

More importantly, the intermediate population groups (e.g., Q2 to Q4) will not be captured in these simple measures of inequality [4,6]. More sophisticated indicators can overcome this limitation by using information on the whole population. Harper and Lynch [35,37] recommended the use of the absolute concentration index or the slope index of inequality (SII) as indicators of absolute inequality, and use of the relative concentration index (CIX) or the relative index of inequality (RII) as indicators of relative inequality. In this article we will focus on the SII and the CIX, which are among the most used measures of inequality in the epidemiologic and economic literature, along with ratios and differences.

The CIX is related to the Gini coefficient [6], which is widely used to measure how much income is concentrated in the hands of the richest in a given country. The Gini coefficient can be

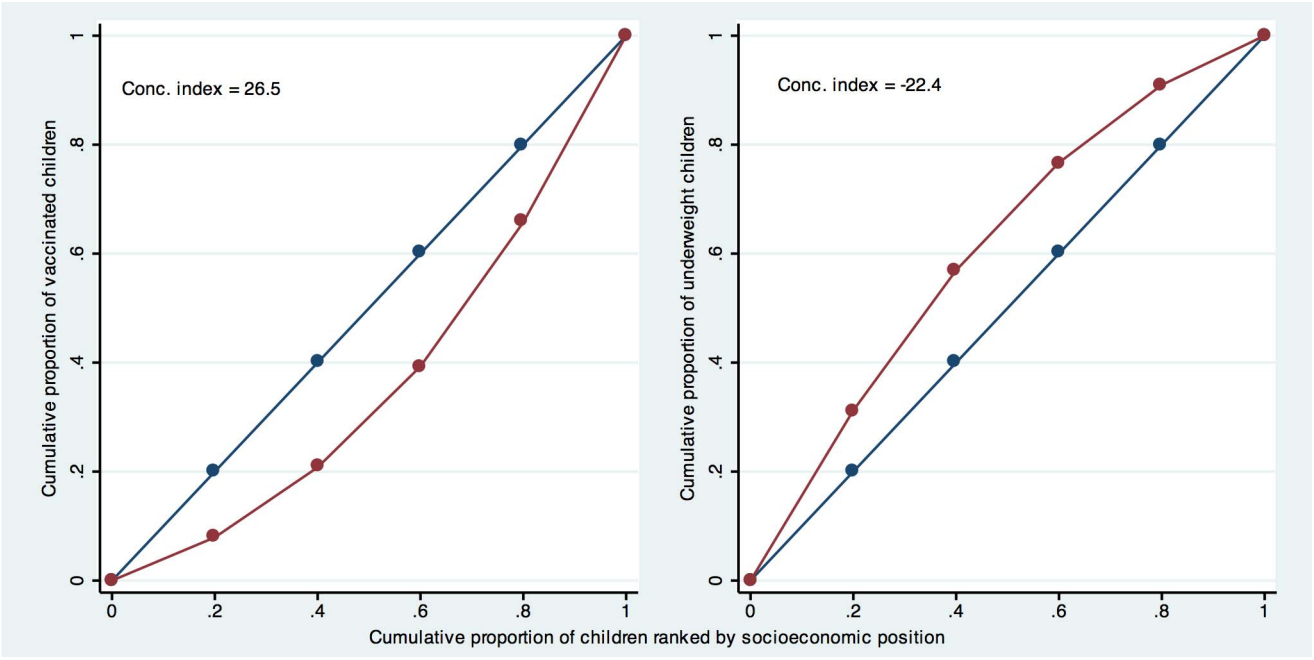


**Figure 2. Concentration curve for measles vaccination and underweight using data from the Nigeria 2008 DHS.** Conc. index, concentration index.
doi:10.1371/journal.pmed.1001390.g002

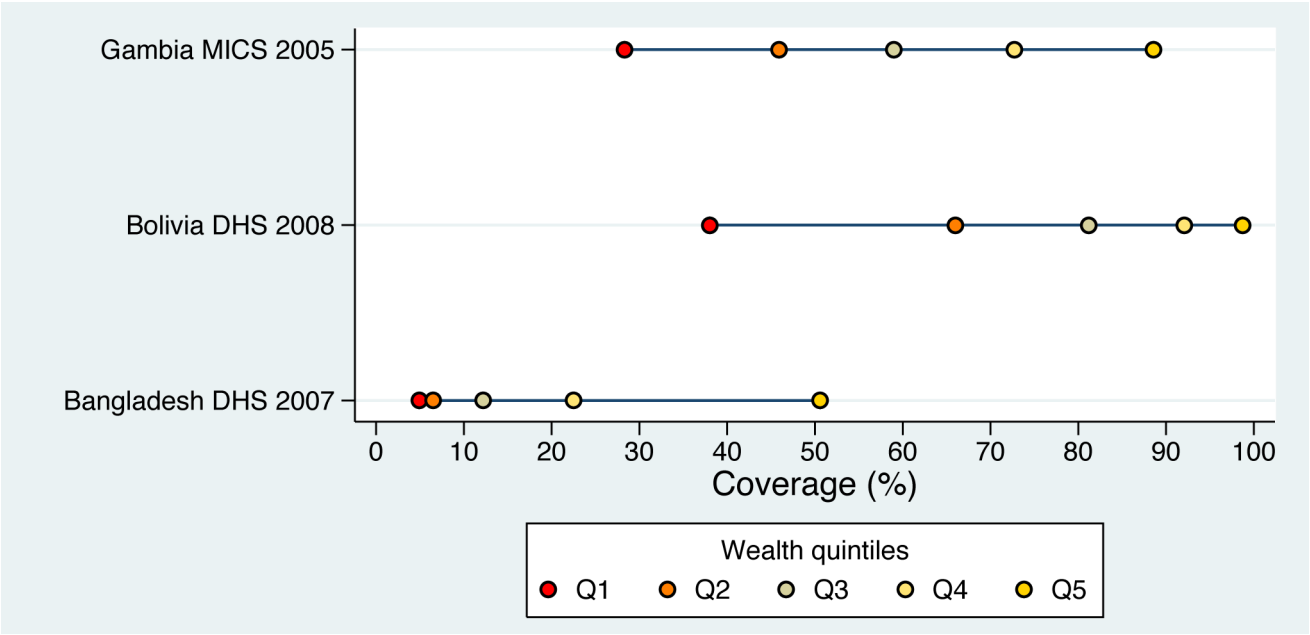


**Figure 3. Linear, bottom, and top patterns of inequality for skilled birth attendance in Gambia, Bolivia, and Bangladesh, respectively.** See Box 3 for further discussion of patterns of inequality.
doi:10.1371/journal.pmed.1001390.g003

## Box 3. Patterns of Inequality

- Inspecting the distance between groups in an inequality graph (such as the five-dot plots in Figure 3) can help in the design of more efficient approaches for improving coverage and reducing inequality.
- Three types of patterns of inequality have been described as "linear," "bottom" and "top" inequality patterns by Victora et al. [33] and as "marginal exclusion," "queuing," and "mass deprivation" by *The World Health Report 2005* [47]. In Figure 3, Gambia, Bolivia and Bangladesh are, respectively, typical examples of these three situations.
- Under usual conditions, low-coverage countries tend to show a top inequality pattern, with the richest quintile way ahead of the rest. As coverage increases they move to the linear pattern, where the distance between groups is similar. When higher levels of coverage are attained, a bottom inequality pattern usually appears, with the poorest lagging behind [33,47].
- Where there is a linear pattern of inequality (Gambia, Figure 3), increased coverage in all groups is still needed, but targeting the poor should also be considered to avoid a bottom inequality pattern evolving.
- Where there is a bottom inequality pattern (Bolivia, Figure 3), targeting the poor is recommended because most of the population has already achieved reasonable levels of coverage.
- Where there is a top inequality pattern (Bangladesh, Figure 3), it is important to disseminate interventions widely, because coverage is low even in the wealthiest group.

expressed in the form of a curve that shows the sample ranked by income on the *x*-axis, and the cumulative distribution of income on the *y*-axis. If everyone in the population has the same income, the curve lies exactly over the diagonal and the Gini index is equal to zero. The area between the diagonal and the observed curve is used to measure the degree of income concentration. The CIX uses an analogous approach by ranking individuals according to socioeconomic position on the *x*-axis and plotting, for example, cumulative intervention coverage on the *y*-axis. Thus, if every wealth quintile had 20% of all the vaccines distributed in a population, for example, the line would be exactly on the diagonal, and there would be no inequality [6,35].

Typically, however, health interventions are more concentrated towards the richer groups, and the CIX assumes a positive value, as the curve is below the diagonal. Figure 2 (left) shows the example of measles vaccination in Nigeria [36], where coverage levels in the five quintiles were 17%, 28%, 41%, 58%, and 75%, respectively, and the CIX is equal to 26.5. By contrast, in the case of ill health, where poorer groups are affected more than richer groups, the CIX is negative. So in Nigeria, where underweight prevalence for the wealth quintiles Q1 to Q5 was 36%, 29%, 22%, 16%, and 10%, respectively, the CIX is −22.4 (Figure 2, right).

The main downside of the CIX is the lack of direct interpretability of its values. Clearly, a value of 20 means more inequality than a value of eight, but these numbers lack a clear meaning, unlike Q5/Q1 ratios, which are easily interpretable.

Alternative formulations for CIX can be used to reflect absolute inequalities [35], but these are used less often than the formulation we have adopted here, which reflects relative inequalities. For measuring absolute inequalities, the SII is being increasingly used [35]. This index is typically derived through linear regression of the health outcome on the midpoints of the ranks obtained by ordering the sample by the explanatory variable when using grouped data. The ranks are scaled so that the values range from zero to one. When using ranks based on quintiles, each group includes approximately 20% of the sample, and the midpoints of

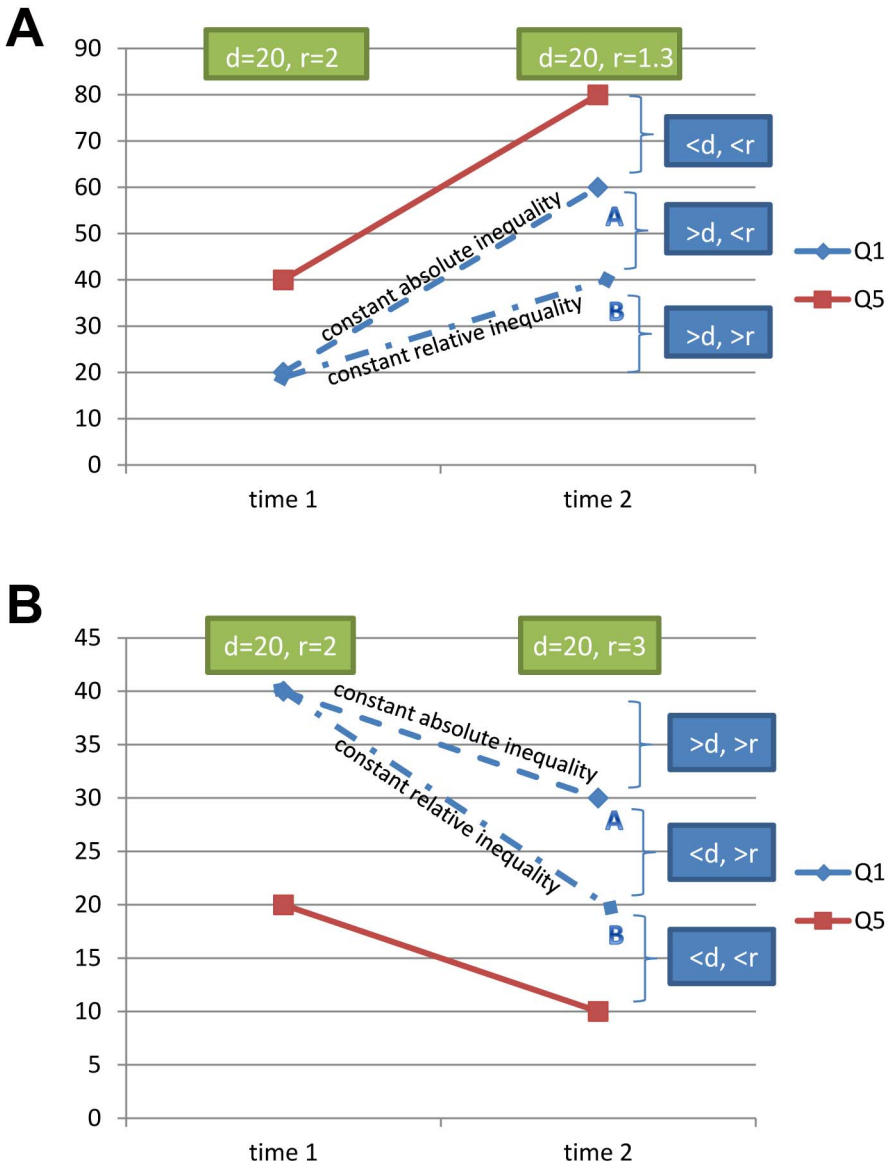


**Figure 4. Different situations in relation to the time trend of the health indicator studied, and how changes are related to increased or decreased measures of inequality.** (A) Situation 1—increasing rates of a health indicator, typical of a preventive intervention, such as immunization, or a desirable behavior such as exclusive breastfeeding. (B) Situation 2—declining rates of a health indicator, typical of an ill-health indicator, such as undernutrition or mortality, or a risk factor, such as high parity. "d" indicates the difference in coverage between the top and bottom quintiles; "r" indicates the ratio of the coverage in the top and bottom quintiles.
doi:10.1371/journal.pmed.1001390.g004

the ranks are close to 0.1, 0.3, 0.5, 0.7, and 0.9 for the five quintiles, respectively. The SII is the slope of the resulting regression line, and represents the absolute difference in the fitted value of the health indicator between the highest (score of 1) and the lowest (score of 0) values of the socioeconomic indicator rank. Using the same data used to calculate CIX in Figure 2 (measles vaccination in the 2008 Nigeria DHS survey) [36], we get a regression line that crosses the *y*-axis (where socioeconomic position equals zero) at 7.6% coverage, and crosses the right side of the chart (where socioeconomic position equals one) at 80%. The SII equals 72.4, which is the difference between these two coverage levels, and indicates that vaccine coverage at the top of the wealth scale is 72.4 percentage points higher than at the bottom.

There are two potential problems with a linear regression approach like this when used with an indicator, such as intervention coverage, that has a minimum of 0% and a maximum of 100%. The first is that it assumes a linear relationship between outcome and predictor, which is not always the case, particularly when a "top inequality" or "bottom inequality" pattern is present (see Box 3). Second, for rare or common outcomes, the model can fit values outside the 0%–100% interval. It can, for example, indicate negative coverage values among the poorest, which is clearly impossible. Using logistic regression instead of a linear model frequently solves both these problems. This approach allows the calculation of the difference between the estimated coverage at the top and bottom of the socioeconomic position scale and is best done using individual data rather than grouped data. In the

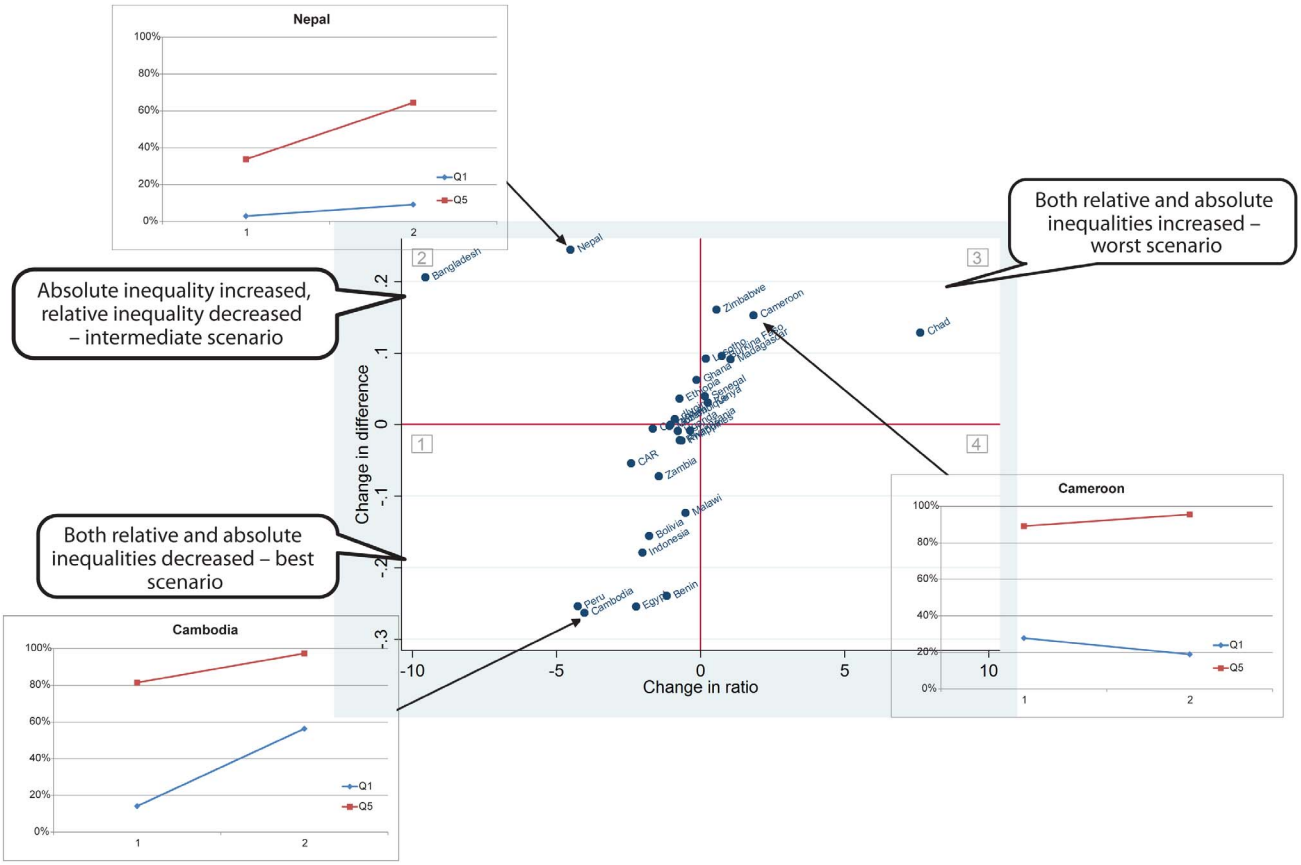


**Figure 5. Real example of a set of countries where skilled birth attendant coverage increased over time for the richest 20% of the population.** This example corresponds to situation 1 in Figure 4. CAR, Central African Republic.
doi:10.1371/journal.pmed.1001390.g005

example of measles vaccination in Nigeria, the logistic regression approach yields an SII equal to 66.8—smaller than but not greatly different from the SII obtained with the linear regression approach.

A measure that is closely related to the SII is the relative index of inequality, or RII. The curve-fitting procedure used to calculate the RII is the same as for the SII, but instead of calculating the difference between the fitted values for one and zero, the RII is the ratio between the two. In the Nigeria measles vaccination example, the RII equals 10.6 (80% divided by 7.6%) when the linear regression approach is used. The estimate reduces to 6.4 if logistic regression is used. Given the potential problems associated with linear regression, we strongly advise that logistic regression should always be used in the calculation of SII or RII for coverage indicators.

There is near consensus in the recent literature that no single measure of inequality reveals the full picture, and that authors should report both absolute and relative measures [35,38]. Relative measures—for example, coverage among the rich is twice as large as among the poor—give an idea of the degree of unfairness. Absolute measures—coverage among the rich is 60 percentage points higher than among the poor—give an idea of the actual effort that will be needed to close the gap. Value judgments, therefore, are implicit in the choice of measures [39], an issue that will be discussed in more detail in the next section. As an aside, there have also been discussions recently on the need to assess the pattern, or type, of inequalities, along with their magnitude, to fully understand the implications of these inequalities for health policy (Box 3).

## Trends in Inequalities

The debate on absolute versus relative measures of inequality alluded to above is particularly controversial when it relates to the issue of whether inequalities are increasing or declining over time [30,38–41]. In some cases, results have been deemed inconsistent because an absolute measure indicated increased inequalities and a relative measure indicated decreased inequalities or vice versa. We show here that absolute and relative measures of inequality are not inconsistent when assessing trends, but are actually complementary.

In Figure 4 we present two situations in which the time trend of a hypothetical health indicator is analyzed. In situation 1, the outcome indicator increases over time (e.g., the coverage of a preventive health intervention). In situation 2, the outcome decreases (e.g., mortality rate or nutritional deficit).

Let us assume that in situation 1 the richest quintile starts at 40% coverage, and the poorest at 20%. The baseline difference equals 20 percentage points, and the ratio equals two. Let us also assume that coverage among the richest increases to 80% at end point. We can then explore two alternatives for coverage among the poorest: coverage "A," where absolute inequality remains unchanged (the difference is the same as at baseline), and coverage "B," where relative inequality remains constant over time (same

## Key Points

- Recent international calls for increased accountability in measuring progress towards the Millennium Development Goals demand analyses of health indicators stratified by socioeconomic position and other equity-related variables.
- Socioeconomic position can be ascertained using many different indicators, but the use of a wealth classification based on assets is the best option for national surveys, being feasible and reliable.
- Intervention coverage can be assessed by individual indicators, but because combined measures are important for the study of time trends and for cross-country comparisons, we advocate the use of the composite coverage index and the co-coverage indicator.
- Inequality measures that take the whole socioeconomic distribution into account are essential, and at least one absolute (the slope index of inequality) and one relative (the relative concentration index) measure should always be presented.
- When analyzing inequality trends, absolute and relative inequality must be studied jointly because there is a clear interacting pattern of reduction or increase in inequality that sometimes produces apparently contradictory changes in absolute and relative inequality.

ratio). The worst scenario in terms of inequality is an end point for the bottom quintile where coverage is below B, because in this region of the chart, inequality will have increased both in terms of the difference and the ratio. The ideal scenario is an end point where coverage is above A. Here, both the difference and the ratio will have decreased. Finally, there are intermediate situations, where coverage is in between A and B. Here, results are apparently inconsistent. Compared to baseline, the difference between the extreme quintiles (absolute inequality) will have increased and the ratio (relative inequality) will have decreased.

In situation 2 where the outcome is declining, we have similar results for the worst- and best-case scenarios (coverage above A and below B, respectively) as in situation 1, with both the difference and ratio increasing or decreasing. The intermediate scenario, on the other hand, is different: the difference will have decreased but the ratio will have increased.

In other words, the apparent conflict between changes in absolute and relative inequalities reflects scenarios where inequalities have been reduced, but not so much that both absolute and relative measures have decreased. In Figure 5 we present a real data example for 29 countries included in the Countdown to 2015 initiative where skilled birth attendant coverage increased over time for the top quintile. Quadrant 1 (lower left) represents the best scenario, where both absolute and relative inequalities decreased, with the inlaid graph showing what happened in Cambodia, from 2000 to 2010. Quadrant 2 (top left) represents the intermediate scenario, where relative inequality decreased, but absolute inequality increased, and the inlay shows the trends for Nepal, from 1996 to 2006. Finally, quadrant 3 (top right) includes countries in the worst of the situations in terms of inequality—both absolute and relative inequality increased, with Cameroon (1998–2006) as an example.

Sample variability has been often overlooked in inequality analysis, irrespective of the measure used, which is problematic when looking at trends. The convenient regression approach to the estimation of the CIX presented by O'Donnell et al. [42] is very interesting, since it allows the computation of both the point and interval estimates. A similar approach can be used to obtain confidence intervals for the SII. O'Donnell's guide and the tools provided by the World Bank [43] are invaluable additional material for those interested in analyzing and interpreting health inequalities.

## Conclusions

In this article, we have provided practical guidance on assessing inequalities in coverage of health and nutrition interventions, with emphasis on survey data from low- and middle-income countries. From our own experience, we make several recommendations about how best to assess inequalities in health intervention coverage. First, we conclude that there is no single best measure of inequality, and recommend that at least one absolute and one relative measure should be presented when describing inequalities at a given point in time, as well as when reporting trends over time. Second, when comparing time points or countries, we emphasize how important it is to calculate measures that take the whole population into account, and advocate the use of the CIX and the SII. In addition, we strongly advise the use of logit-based SII for the measurement of absolute inequalities. Because the presentation of these indices is particularly appropriate for academic audiences, we also recommend calculation of the differences and ratios among extreme quintiles, because these are easy to convey to general audiences. Third, when assessing change in inequalities, we argue that it is essential not only to evaluate both absolute and relative changes, but also to report how they evolve jointly. Finally, in situations where conflicting results are provided by absolute and relative measures, we stress that it is essential that researchers spell out the different interpretations of these measures to public health experts, because these interpretations are affected by value judgments and are likely to affect the approaches taken to reduce inequalities in the coverage of health interventions.

## Acknowledgments

We wish to thank the members of Child Health Epidemiology Reference Group (CHERG) for their insightful comments on the draft version of this manuscript, and especially Jennifer Bryce for putting together this series, encouraging us in writing this piece, and making everything happen. We also thank the Pelotas International Center for Equity in Health (Federal University of Pelotas, Brazil), which produced all the equity data breakdowns used in our analyses and examples.

## Author Contributions

Analyzed the data: AJB CGV. Wrote the first draft of the manuscript: AJB CGV. Contributed to the writing of the manuscript: AJB CGV. ICMJE criteria for authorship read and met: AJB CGV. Agree with manuscript results and conclusions: AJB CGV.

## References

1. Gray AM (1982) Inequalities in health. The Black Report: a summary and comment. Int J Health Serv 12: 349–380.
2. Almeida-Filho N, Kawachi I, Pellegrini-Filho A, Dachs JN (2003) Research on health inequalities in Latin America and the Caribbean: bibliometric analysis (1971–2000) and descriptive content analysis (1971–1995). Am J Public Health 93: 2037–2043.
3. Whitehead M (1992) The concepts and principles of equity and health. Int J Health Serv 22: 429–445.

4. Mackenbach JP, Kunst AE (1997) Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. Soc Sci Med 44: 757–771.
5. Kakwani N, Wagstaff A, van Doorslaer E (1997) Socioeconomic inequalities in health: Measurement, computation, and statistical inference. J Econom 77: 87–103.
6. Wagstaff A, Paci P, van Doorslaer E (1991) On the measurement of inequalities in health. Soc Sci Med 33: 545–557.
7. Murray CJ, Gakidou EE, Frenk J (1999) Health inequalities and social group differences: what should we measure? Bull World Health Organ 77: 537–543.
8. Braveman P, Krieger N, Lynch J (2000) Health inequalities and social inequalities in health. Bull World Health Organ 78: 232–234; discussion 234–235.
9. Braveman P, Tarimo E (2002) Social inequalities in health within countries: not only an issue for affluent nations. Soc Sci Med 54: 1621–1635.
10. Victora CG, Wagstaff A, Schellenberg JA, Gwatkin D, Claeson M, et al. (2003) Applying an equity lens to child health and mortality: more of the same is not enough. Lancet 362: 233–241.
11. Gwatkin DR (2000) Health inequalities and the health of the poor: what do we know? What can we do? Bull World Health Organ 78: 3–18.
12. Kawachi I, Adler NE, Dow WH (2010) Money, schooling, and health: mechanisms and causal evidence. Ann N Y Acad Sci 1186: 56–68.
13. Marmot Review (2010) Fair society, healthy lives: strategic review of health inequalities in England post 2010. London: Marmot Review.
14. Lee JH, Sadana R, editors(2011) Improving equity in health by addressing social determinants. Geneva: World Health Organization.
15. Victora CG, Barros AJ, Axelson H, Bhutta ZA, Chopra M, et al. (2012) How changes in coverage affect equity in maternal and child health interventions in 35 Countdown to 2015 countries: an analysis of national surveys. Lancet 380: 1149–1156.
16. World Health Organization (2012) Global health observatory. Available: http://www.who.int/gho/. Accessed 12 November 2012.
17. Countdown to 2015: Maternal, Newborn and Child Survival (2012) Countdown to 2015: Maternal, Newborn and Child Survival. Available: www.countdown2015mnch.org. Accessed 12 November 2012.
18. MEASURE DHS (2012) Demographic and Health Surveys. Available: http://www.measuredhs.com/ Accessed 12 November 2012.
19. UNICEF (2012) Statistics and Monitoring: Multiple Indicator Cluster Survey. Available: http://www.unicef.org/statistics/index_24302.html. Accessed 12 November 2012.
20. Gwatkin DR, Rutstein S, Johnson K, Suliman E, Wagstaff A, et al. (2007) Socio economic differences in health, nutrition, and population within developing countries. Washington: The World Bank.
21. Barros F, Victora C, Scherpbier R, Gwatkin D (2009) Inequities in the health and nutrition of children. In: Blas E, Sivasankara Kurup A, editors. Priority public health conditions: from learning to action on social determinants of health. Geneva: World Health Organization. pp. 50–75.
22. Barros AJ, Ronsmans C, Axelson H, Loaiza E, Bertoldi AD, et al. (2012) Equity in maternal, newborn, and child health interventions in Countdown to 2015: a retrospective review of survey data from 54 countries. Lancet 379: 1225–1233.
23. Krieger N, Williams DR, Moss NE (1997) Measuring social class in US public health research: concepts, methodologies, and guidelines. Annu Rev Public Health 18: 341–378.
24. Howe LD, Galobardes B, Matijasevich A, Gordon D, Johnston D, et al. (2012) Measuring socio-economic position for epidemiological studies in low- and middle-income countries: a methods of measurement in epidemiology paper. Int J Epidemiol 41: 871–886.
25. Filmer D, Pritchett L (1998) Estimating wealth effects without expenditure data—or tears: with an application to educational enrollments in states of India. Policy Research Working Paper WPS1994. Washington (District of Columbia): The World Bank. Available: http://documents.worldbank.org/curated/en/1998/10/442667/estimating-wealth-effects-without-expenditure-data-or-tears-application-educational-enrollments-states-india. Accessed 12 December 2012.
26. Jolliffe IT (2002) Principal component analysis. New York: Springer.
27. Boerma JT, Bryce J, Kinfu Y, Axelson H, Victora CG (2008) Mind the gap: equity and trends in coverage of maternal, newborn, and child health services in 54 Countdown countries. Lancet 371: 1259–1267.
28. Morris SS (2000) Validity of rapid estimates of household wealth and income for health surveys in rural Africa. J Epidemiol Community Health 54: 381–387.
29. Houweling TA, Kunst AE, Mackenbach JP (2003) Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter? Int J Equity Health 2: 8.
30. Wirth ME, Balk D, Delamonica E, Storeygard A, Sacks E, et al. (2006) Setting the stage for equity-sensitive monitoring of the maternal and child health Millennium Development Goals. Bull World Health Organ 84: 519–527.
31. Filmer D, Kinnon S (2008) Assessing asset indices. Policy Research Working Paper WPS4605. Washington (District of Columbia): The World Bank.
32. Kawachi I, Subramanian SV, Almeida-Filho N (2002) A glossary for health inequalities. J of Epidemiol Community Health 56: 647–652.
33. Victora CG, Fenn B, Bryce J, Kirkwood BR (2005) Co-coverage of preventive interventions and implications for child-survival strategies: evidence from national surveys. Lancet 366: 1460–1466.
34. Countdown to 2015 (2012) Building a future for women and children. The 2012 report. Geneva: World Health Organization. Available: http://countdown2015mnch.org/reports-and-articles/2012-report. Accessed 12 November 2012.
35. Harper S, Lynch J (2005) Methods for measuring cancer disparities: using data relevant to healthy people 2010 cancer-related objectives. Bethesda (Maryland): National Cancer Institute.
36. Measure DHS (2009) Nigeria: standard DHS, 2008. Available: http://www.measuredhs.com/what-we-do/survey/survey-display-302.cfm. Accessed 12 November 2012.
37. Harper S, Lynch J (2007) Selected comparisons of measures of health disparities: a review using databases relevant to healthy people 2010 cancer-related objectives. Bethesda (Maryland): National Cancer Institute.
38. Moser K, Frost C, Leon DA (2007) Comparing health inequalities across time and place—rate ratios and rate differences lead to different conclusions: analysis of cross-sectional data from 22 countries 1991–2001. Int J Epidemiol 36: 1285–1291.
39. Harper S, King NB, Meersman SC, Reichman ME, Breen N, et al. (2010) Implicit value judgments in the measurement of health inequalities. Milbank Q 88: 4–29.
40. Minujin A, Delamonica E (2003) Mind the gap! Widening child mortality disparities. J Hum Dev 4: 397–418.
41. Scanlon JP (2011). Measuring health disparities—summary. Available: http://www.jpscanlan.com/measuringhealthdisp.html. Accessed 15 November 2012.
42. O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M (2008) Analyzing health equity using household survey data. A guide to techniques and their implementation. Washington (District of Columbia): The World Bank. 220 p.
43. World Bank (2010) Health Equity and Financial Protection—tools for equity analyses [under section "Stata programs and Excel files to accompany Analyzing Health Equity Using Household Survey Data"]. Available: http://go.worldbank.org/YHORYCGNK0. Accessed 5 December 2012.
44. Deaton A (2001) Counting the world's poor: problems and possible solutions. World Bank Res Obs 16: 125–147.
45. Deaton A, Zaidi S (2002) Guidelines for constructing consumption aggregates for welfare analysis. Washington (District of Columbia): The World Bank. 108 p.
46. Macro International (1996) Sampling manual. DHS-III basic documentation, No. 6. Calverton (Maryland): Macro International.
47. World Health Organization (2005) The world health report 2005. Make every mother and child count. Geneva: World Health Organization.

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: Total Survey Error and the Interpretation of Intervention Coverage Estimates from Household Surveys

**Thomas P. Eisele[1]*, Dale A. Rhoda[2], Felicity T. Cutts[3], Joseph Keating[1], Ruilin Ren[4], Aluisio J. D. Barros[5], Fred Arnold[4]**

**1** Department of Global Health Systems and Development, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana, United States of America, **2** Center for Analytics and Public Health, Battelle Memorial Institute, Columbus, Ohio, United States of America, **3** Independent consultant, La Londe les Maures, France, **4** Demographic and Health Surveys, ICF International, Calverton, Maryland, United States of America, **5** Federal University of Pelotas, Pelotas, Brazil

**Abstract:** Nationally representative household surveys are increasingly relied upon to measure maternal, newborn, and child health (MNCH) intervention coverage at the population level in low- and middle-income countries. Surveys are the best tool we have for this purpose and are central to national and global decision making. However, all survey point estimates have a certain level of error (total survey error) comprising sampling and non-sampling error, both of which must be considered when interpreting survey results for decision making. In this review, we discuss the importance of considering these errors when interpreting MNCH intervention coverage estimates derived from household surveys, using relevant examples from national surveys to provide context. Sampling error is usually thought of as the precision of a point estimate and is represented by 95% confidence intervals, which are measurable. Confidence intervals can inform judgments about whether estimated parameters are likely to be different from the real value of a parameter. We recommend, therefore, that confidence intervals for key coverage indicators should always be provided in survey reports. By contrast, the direction and magnitude of non-sampling error is almost always unmeasurable, and therefore unknown. Information error and bias are the most common sources of non-sampling error in household survey estimates and we recommend that they should always be carefully considered when interpreting MNCH intervention coverage based on survey data. Overall, we recommend that future research on measuring MNCH intervention coverage should focus on refining and improving survey-based coverage estimates to develop a better understanding of how results should be interpreted and used.

*This paper is part of the* PLOS Medicine "*Measuring Coverage in MNCH" Collection*

## Introduction

Nationally representative household surveys are increasingly relied upon to measure maternal, newborn, and child health (MNCH) intervention coverage at the population level in low- and middle-income countries. These surveys include the Demographic and Health Survey/s (DHS), the Multiple Indicator Cluster Survey/s (MICS), the AIDS Indicator Survey/s (AIS), and the Malaria Indicator Survey/s (MIS). These surveys rely on scientific sampling methods, which require each element of the target population to have a known and non-zero probability of selection, to obtain point estimates of MNCH intervention coverage at the national and sub-national levels every 3–5 years [1]. Because accurate up-to-date sampling frames of individuals and households are often unavailable in many low- and middle-income countries, these surveys typically use a multi-stage cluster sample design. Clusters (primary sampling units) are selected at the first stage with a probability proportional to size strategy, with size being the estimated population size of the cluster. A constant number of households is then randomly selected at the second stage of selection from a sampling frame of households created from a complete enumeration of households in the selected clusters. Household surveys have become increasingly standardized in their sampling approaches and questionnaire designs to produce comparable results across countries and over time.

In using the coverage figures for MNCH interventions from household surveys for programmatic and policy decisions, it is important to remember that surveys only provide estimates of the

**Citation:** Eisele TP, Rhoda DA, Cutts FT, Keating J, Ren R, et al. (2013) Measuring Coverage in MNCH: Total Survey Error and the Interpretation of Intervention Coverage Estimates from Household Surveys. PLoS Med 10(5): e1001386. doi:10.1371/journal.pmed.1001386

**Academic Editor:** Nyovani Madise, Professor of Demography and Social Statistics, University of Southampton, United Kingdom

**Published** May 7, 2013



**Funding:** No specific funding was received for the writing of this paper. TPE and JK were funded by the Malaria Control and Evaluation Partnership in Africa (MACEPA), a PATH project, from funding from the Bill & Melinda Gates Foundation. This work was conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF. CHERG receives financial support from the Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. No funders were involved in in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** AIS, AIDS Indicator Survey/s; DHS, Demographic and Health Survey/s; EPI, Expanded Programme on Immunization; ITN, insecticide-treated mosquito net; LQAS, Lot Quality Assurance Sampling; MICS, Multiple Indicator Cluster Survey/s; MIS, Malaria Indicator Survey/s; MNCH, maternal, newborn, and child health.

* E-mail: teisele@tulane.edu

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

true population characteristics of interest. No matter how large a sample is drawn, and no matter how well a survey is designed or implemented, all survey point estimates have a certain level of error (total survey error) comprising sampling and non-sampling error, both of which need to be considered when interpreting survey results. Sampling error—the difference between the sample estimate of a population characteristic and the real value of the population characteristic—is the result of sampling from the population rather than taking measurements from the entire population [2]. Non-sampling error includes all other sources of error. For multi-stage cluster sampling of households in low- and middle-income countries, the most common types of non-sampling error are information bias and selection bias but coverage bias and non-response bias are also important issues.

In this review, which is part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, we discuss the important sampling error and non-sampling error issues involved in interpreting MNCH intervention coverage estimates derived from household surveys for decision making using relevant examples from national surveys to provide context.

## Sampling Error and Its Implications

### Confidence Intervals

We most often think of sampling error as the precision of a point estimate, represented by 95% confidence intervals. Confidence intervals are useful for two purposes. First, they characterize the precision of the estimate. Second, they provide context about whether estimated parameters are likely to be equal between two populations or time-points. In the absence of non-sampling error, if a large number of repeated samples is taken from the target population using the same sampling design, 95% of the resultant 95% confidence intervals about each sample point estimate will contain the true, albeit unknown, intervention coverage in the population [3]. In practice, the interpretation of confidence intervals is slightly different. Suppose, for example, we have a point estimate of intervention coverage of 70%, with a 95% confidence interval of 65%–75% in a program for improving MNCH. This confidence interval is typically taken to mean that we are 95% certain that the true intervention coverage in the population lies somewhere between 65% and 75%. Consider if the program had set a target of achieving 70% population coverage. In this case, one could not say with a high degree of confidence that the program had achieved its target coverage, even with a point estimate for coverage at 70% since fully half of the confidence limit lies below 70%.

### Sample Size and Sampling Design

The level of sampling error around a point estimate depends on the number of observations selected from the target population (sample size), the underlying prevalence and variance of the characteristic of interest in the target population, and the sampling design used in selecting the sample. In general, for a given sampling design, the precision of a point estimate improves with increasing sample size. Thus, a larger sample size yields a tighter confidence interval. However, given a specified sample size, the sampling design also directly affects the precision of an intervention coverage point estimate. In practice, data from nationally representative surveys nearly always come from a cluster survey rather than a simple random sample of households. In cluster surveys, the sampling error is affected by the number of respondents per group as well as the intracluster (or intraclass) correlation coefficient, which provides a measure of how similar the elements being analyzed in each cluster are with respect to a particular characteristic of interest [2]. In general, for a given sample size, the precision of a point estimate ascertained from a cluster sample design decreases as the sample size per cluster increases, and as the intracluster correlation coefficient increases.

### Reporting Sampling Errors and Confidence Intervals

Although the sample size and sampling design are frequently reported in survey methods, sampling errors represented by confidence intervals are rarely reported in the main body of reports of national surveys. DHS survey reports and reports for AIS and MIS surveys conducted through the DHS program include standard errors and accompanying confidence intervals of $\pm 2$ standard errors (slightly wider than the 95% confidence interval) for most key indicators in Appendix B only. In MICS survey reports, standard errors and 95% confidence intervals appear in Appendix S. Almost no MIS surveys conducted outside of the DHS program include sampling errors in the survey reports or appendices.

### Comparing Survey Results

One oft-cited advantage of large well-conducted surveys is that their results can be compared meaningfully both over time and between locations. In the absence of a formal hypothesis test, informal conclusions about the statistical significance of a difference are sometimes drawn by examining whether two confidence intervals overlap. Indeed, if two 95% confidence intervals do not overlap, then a formal hypothesis test would reject the null hypothesis of equality with a $p$-value below 5%. Unfortunately the converse is not true. If 95% confidence intervals overlap, it is still possible for the $p$-value from the hypothesis test to fall below 5% [4,5]. An example from the 2011 Ethiopia DHS report (presented in Box 1 and Figure 1) illustrates how difficult it can be to correctly interpret MNCH intervention coverage estimates without considering confidence intervals [6].

### Sampling Error and Data Disaggregation

When analyzing survey data for MNCH coverage indicators, sampling errors and the accompanying confidence intervals are influenced, at times substantially, by changes in sample size owing to disaggregating the data by socio-behavioural or demographic characteristics of interest, such as where the respondents live or their household socioeconomic status. While DHS and MICS surveys typically report the confidence intervals around many key indicators in their appendices, they do not report confidence intervals of disaggregated point estimates. To illustrate why it is important to consider the sampling error of such estimates, we present summary sampling characteristics of established malaria control coverage indicators from the 2007 Zambia DHS survey (Table 1) [7]. Point estimates were weighted to account for unequal probability sampling and any differential non-response. Standard errors were estimated to account for correlated data at the primary sampling unit level using the Huber White Sandwich estimator, which accounts for the loss of precision due to increased intracluster correlation coefficient.

For assessing progress of malaria control efforts in Zambia, the standard indicator for household insecticide-treated mosquito net (ITN) coverage shows that 52% of households possess at least one ITN, with a 95% confidence interval of $\pm 2.5$ percentage points (Table 1). The interpretation for programmatic purposes, based on information from the sample of 7,164 households, would be that one has 95% confidence that the true (albeit unknown) proportion of households in Zambia with at least one ITN lies somewhere between 50% and 55%, in the absence of non-sampling error. By contrast, the indicator used for assessing progress in the coverage of access to antimalarial treatment of fevers in children (38%) has a

### Box 1. An Illustration of the Difficulty in Correctly Interpreting MNCH Intervention Coverage Estimates without Considering Confidence Intervals

The 2011 Ethiopia DHS survey report states "The percentage of children age 12–23 months who were fully vaccinated at the time of the survey increased from 14% in 2000 to 20% in 2005 and 24% in 2011—a 70% increase over 10 years and a 19% increase in the 5 years preceding the 2011 survey" [6]. The main text of the Ethiopia report does not present confidence intervals, but Appendix B shows that for the 2011 survey the approximate 95% confidence interval for the immunization coverage of 24% was 21%–28%. In the 2005 survey, the confidence interval around the point estimate of 20% was 17%–23%, and in 2000 the confidence interval around the point estimate of 14% was 12%–16% [33]. Although a statistical test for significance would be needed to draw a strong conclusion, the difference between 2005 and 2011 (20% up to 24%) is not likely to be statistically significant at the 5% level because the confidence intervals overlap (see Figure 1). It would have been helpful for the body of the report to note the *p*-value associated with the 19% increase, or to say that the 2005 to 2011 increase is not statistically significant at the 5% level. This example demonstrates the importance of considering confidence intervals when survey-based estimates are used to draw strong conclusions or to make important programmatic decisions.

95% confidence interval of ±3.8 percentage points, because this sample is limited to only those children with fever in the past 2 weeks (5,844 children). When disaggregated by household socioeconomic status, the 95% confidence intervals increase considerably. For example, based on a subsample of 106 children, the proportion with a fever in the past 2 weeks in the wealthiest households who received an antimalarial was estimated to be 40%, with a 95% confidence interval of ±10 percentage points, which would be interpreted to mean that one can only be 95% confident that the true antimalarial coverage lies somewhere between 30% and 50% in this population in Zambia. This level of uncertainty clearly poses challenges in the interpretation of the point estimate for decision making.

## Non-sampling Error and Its Implications

Unlike sampling error, the direction and magnitude of non-sampling error is almost always unmeasurable, and therefore unknown. Non-sampling error cannot be controlled directly by sample size or by type of probability sampling design used.

Non-sampling error is more insidious than sampling error and can develop at many stages of the survey. Study planners need to anticipate as many threats to the validity of the survey results as possible and put careful controls in place to limit the magnitude of non-sampling error, such as the strict use of probability sampling, sensitization to increase response rates, high quality training of data collectors, and implementation of quality control measures for field work. Another paper in this collection on vaccination

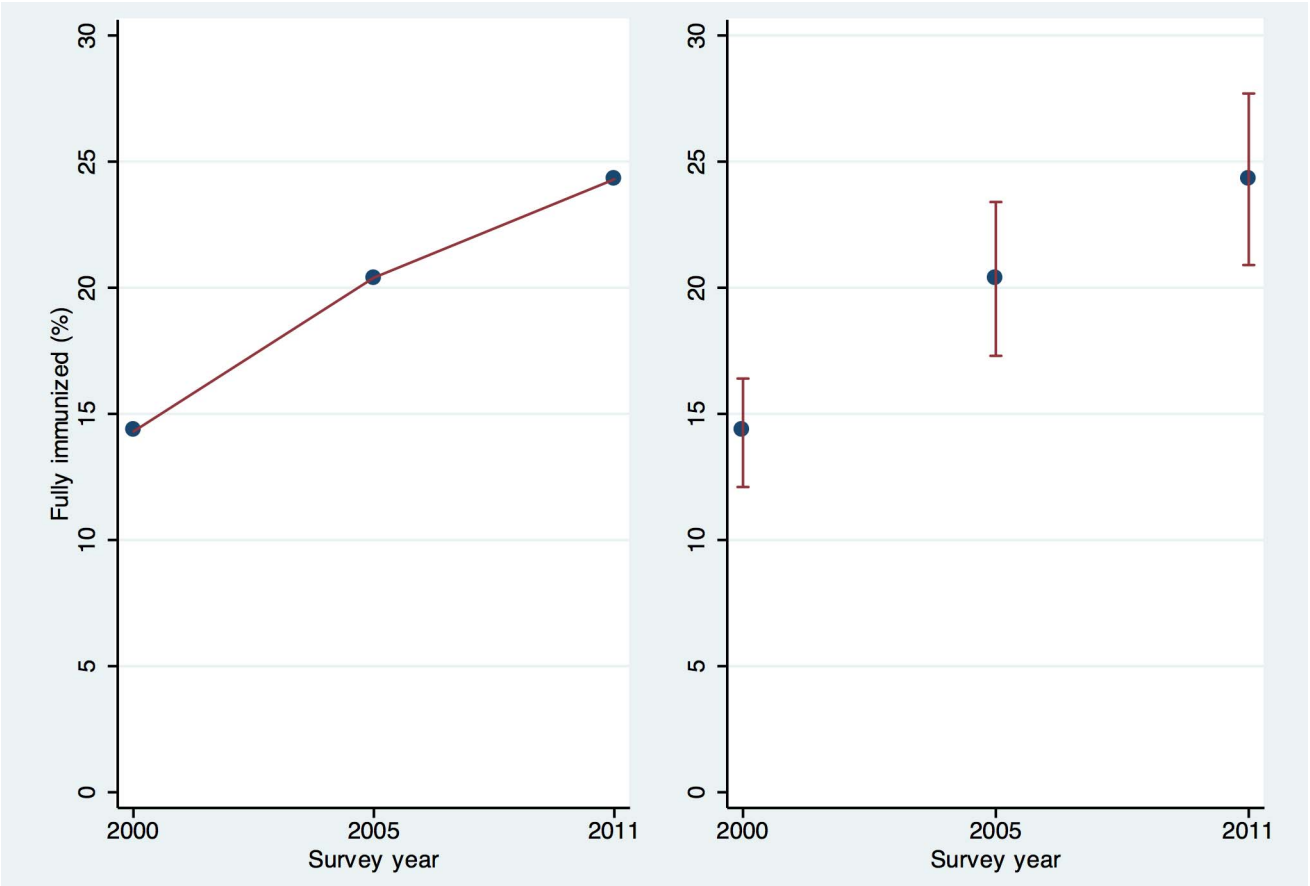


**Figure 1. Two ways of looking at coverage of full immunization in Ethiopia—with and without confidence intervals.** Ethiopia DHS surveys 2000, 2005, 2011 [33].
doi:10.1371/journal.pmed.1001386.g001

**Table 1.** Sampling characteristics of selected point estimates from the 2007 Zambia DHS Survey [7].

| Indicator | Percent Point Estimate | Percent Standard Error | 95% CI | Sample Size |
|---|---|---|---|---|
| **Percent of households with ≥1 ITN** | 52.2 | 1.24 | 49.8–54.7 | 7,164 |
| **Percent of children <5 years old with fever in past 2 weeks** | 17.7 | 0.71 | 16.4–19.2 | 5,844 |
| **Percent of children <5 years old with fever who received any antimalarial** | 38.2 | 1.85 | 34.6–41.9 | 1,034 |
| Wealth quintile[a] | | | | |
| Lowest | 35.3 | 3.85 | 27.7–43.0 | 219 |
| Second | 42.6 | 3.60 | 35.5–49.7 | 228 |
| Middle | 36.6 | 3.44 | 29.8–43.4 | 253 |
| Fourth | 37.5 | 3.15 | 31.3–43.8 | 228 |
| Highest | 39.6 | 5.02 | 29.5–49.7 | 106 |

Point estimates in this table may vary slightly from the point estimates reported in the 2007 Zambia DHS survey because of slight differences in inclusion criteria during analysis, although all are within 1%.

[a]There are different numbers of children in each wealth quintile because wealth quintiles are calculated at the household level for all persons in the household and not for subgroups.

doi:10.1371/journal.pmed.1001386.t001

coverage provides more details of sources of non-sampling error and measures to mitigate against them [8].

Importantly, although sampling error is generally reduced by increasing sample size, non-sampling error can have an inverse relationship with sample size. Together, these factors make interpreting MNCH intervention coverage estimates obtained from household surveys very challenging in the presence of substantial non-sampling error. Non-sampling errors are also a major concern when comparing survey results from different surveys using different survey instruments or survey methodology, or inadequate levels of quality control. Below we present examples of the more common types of non-sampling errors and biases that threaten MNCH intervention coverage estimated from household surveys.

## Information Error and Information Bias

Information error and information bias (also referred to as measurement error and bias) are common in household survey estimates of MNCH intervention [8–13]. Information error arises from errors in measuring MNCH intervention coverage, which often occur when the respondents do not know the exact answer to the survey questions, yet provide answers anyway. Information bias arises from systematic (i.e., non-random) errors in measuring MNCH intervention coverage and includes recall bias and social-desirability bias. Although information error typically results in higher variance, and as a result decreased precision of the point estimate, information bias results in an overestimate or underestimate of the population point estimate. The real problem is that one does not know which way this type of error is biasing the results.

The measurement of vaccination coverage provides a good example of how information error and bias can affect the ability of decision makers to use information from surveys to evaluate the trends in coverage over time. As described by Cutts and colleagues elsewhere in this collection [8], information on vaccination coverage is usually kept in home-based vaccination records but once a child is past infancy, parents may not retain these records. If a vaccination card is not available, DHS and MICS surveys request information from each mother on the vaccinations received by children born in the past 5 years. Because home-based vaccination records are less likely to be available for children born 4 years ago than for those born 1 year ago, recall error or bias may be more likely for older than for younger children, which reduces the ability of decision makers to draw strong conclusions about trends.

Even with the best designed surveys and questionnaires, recall error and bias can affect MNCH intervention coverage estimates [14–16]. For example, for estimating ITN coverage, a key indicator for malaria control programs [17], surveys must ascertain the type of net used and, for nets that are not long-lasting insecticidal nets, when the net was procured and treated with insecticide to distinguish nets that are ITNs from those that are not. Data from Eritrea show substantial date heaping (rounding off of time since an event by respondents) at 12 months for both net procurement and insecticide re-treatment. Date heaping affects whether nets meet the definition of an ITN and therefore has the potential to bias coverage estimates [18].

Measurement of ITN use by children <5 years old also provides a good example of how social desirability bias can impact MNCH coverage estimates. Suppose a national survey obtained an estimate of ITN use among children of 60%, with a 95% confidence interval of 55%–65% but that respondents in selected households tended to tell household interviewers that their ITNs were being used more often than they really were because the respondents feel socially obligated to respond positively to questions about recent net use, a situation that is thought to occur [19]. In this scenario, social desirability bias results in systematic over-reporting of the use of ITNs. Even in a well-designed and implemented survey with a sampling error ±10%, we cannot be 95% certain that the true proportion of child ITN use falls within the calculated 95% confidence interval of 55%–65%. In fact, we would need to use expert but subjective judgment as to the direction and the magnitude of any bias when interpreting this coverage estimate and the degree to which bias may affect results.

## Selection Bias

Another common form of non-sampling error is selection bias, which occurs when potential respondents do not all have a known and non-zero probability of selection. In such a case, no matter how large a sample is taken, survey point estimates will always differ from the true, yet unknown, population characteristic under study [3]. Common sources of selection bias in surveys that measure MNCH intervention coverage include: (1) failure to

### Box 2. Selection Bias Resulting from a Two-Stage Cluster Survey Protocol Where Complete Enumeration of Households within Primary Sampling Units Is Not Undertaken (Original EPI Method)

Rapid surveys to assess sub-national intervention coverage in low- and middle-income countries often follow a variation of the "30×7 EPI" protocol where 30 villages are selected at random from a list of villages within a given area, typically with the probability of selection based on the estimated relative number of households in each villages provided by local officials or census data (probability proportional to size sampling). Within each village (cluster), the survey team selects seven households for interview to ascertain information on intervention coverage, most typically immunization coverage [8]. The procedure used for the selection of households in each cluster often involves beginning at a central area of the selected village, counting all households in a randomly selected direction from the center of the village to the village edge (a.k.a., "spin the bottle technique"), selecting a starting household randomly and then proceeding always to the next closest house until seven children aged 12–23 months have been found [20,22]. This protocol can be more prone to selection bias than the household enumeration protocol used by DHS and MICS surveys for several reasons.

- If all households within selected villages are not enumerated to create an accurate sampling frame from which the seven households are selected at the second stage, one cannot document that each household in the village has a known and non-zero probability of selection, which is a requirement for probability sampling.
- It is rare that the true number of households in a selected village is equal to the estimated numbers used when selecting the 30 villages. This is not an issue if a true count of households in each of the 30 villages is obtained after selection, which then allows the survey data to be weighted during analysis to adjust for discrepancies in estimated and actual village sizes. However, if an accurate count of households is not obtained, the survey data cannot be weighted, and selection bias due to unequal probability of household selection cannot be ruled out.
- There is no requirement to document potentially eligible individuals in each household and conduct revisits. Any household where respondents are absent at the time of the visit is simply replaced by the nearest household having an eligible individual and respondent.

include some eligible respondents in the sampling frame, (2) failure to account properly for the sampling design in analyzing survey data because the sampling design parameters (e.g., cluster size) have not been reported, and (3) the use of non-probability sampling designs.

Coverage bias results when the sampling frame excludes some members of the target population or includes ineligible persons who are not part of the target population. It can occur when certain districts of a country are excluded from a survey because of security considerations. It can also occur when members of the target population are homeless, nomadic, or institutionalized and therefore not identified with one of the physical homes used in the multistage household sampling designs common in low- and middle-income countries.

The household sampling methods designed by the World Health Organization for measuring child immunization coverage under the Expanded Programme on Immunization (EPI) provide a good example of the importance of considering sampling design when interpreting survey data. In this program, 30 clusters are selected and seven children within each cluster are selected for assessment of their immunization status [20,21]. Variations of this protocol are widely used to measure sub-national intervention coverage in low- and middle-income countries as they require considerably fewer resources to implement than the well-established protocols employed in DHS and MICS surveys [8,22], which enumerate all eligible households in the selected cluster and then make random selections from that updated list. While improvements have been made over the years, the EPI survey method does not typically include the development of an accurate sampling frame within each of the selected 30 clusters for selecting the seven children, and thus the survey design does not ensure that each child sampled has a known non-zero probability of selection [23]. Moreover, sampling weights cannot be used when analyzing these data to account for unequal probabilities of child selection, which requires accurate counts of the households in each cluster. As a result, selection bias cannot be ruled out, and may actually be quite likely when using traditional EPI 30×7 cluster survey methods [24]. Box 2 presents a hypothetical example of these issues.

MNCH parameters are also sometimes assessed using "rapid" survey protocols that are designed to use minimal samples sizes to obtain local estimates of intervention coverage in comparison with a desired value for programmatic purposes. The small sample size survey technique called Lot Quality Assurance Sampling (LQAS) is a good example [25]. There are several varieties of this survey technique with varying levels of rigor [26–31], all of which seek to

### Box 3. Recommendations for Interpreting MNCH Coverage Indicators Measured from Household Surveys

- Program managers and policy-makers must consider both sampling and non-sampling error when using MNCH coverage estimates for programmatic decision making and assessing progress against coverage targets.
- Study planners need to anticipate as many threats to the validity of the survey estimates as possible and put careful controls in place to limit the magnitude of non-sampling error.
- To help with interpreting MNCH intervention coverage estimates, additional validation research that identifies the sources, direction, and magnitude of non-sampling error of these estimates is needed.
- All survey reports that present MNCH intervention coverage estimates should provide detailed descriptions of the sampling design, sample size, survey instruments, quality controls, data analysis, and data collection protocols to improve the transparency, consistency, and interpretability of estimates.
- To allow for the consideration of important sources of non-sampling error in interpreting MNCH coverage estimates, survey reports should include a detailed limitations section that explicitly lists possible sources of non-sampling error, and should speculate about their direction and magnitude where possible.

## Key Points

- Household surveys are the best tool we currently have for measuring MNCH intervention coverage at the population level in low- and middle-income countries and are central to national and global decision making.
- All estimates of MNCH intervention coverage obtained from household surveys have a certain level of error (total survey error) comprising sampling and non-sampling error, both of which must be considered when interpreting survey results.
- Sampling error (the precision of a point estimate) is measurable and is represented by the 95% confidence intervals, which characterize the precision of the estimate and provide context about whether estimated parameters are likely to be equal between two populations or time-points.
- Information error and bias are common sources of non-sampling error in household survey estimates; survey methods should be reported in enough detail to allow assessment of the potential for non-sampling error, the direction and magnitude of which is almost always unmeasurable.
- The focus of future research for measuring MNCH intervention coverage should be on refining and improving survey-based coverage estimates and developing a better understanding of how results should be interpreted and used.

classify districts as having either adequate or inadequate intervention coverage. LQAS surveys are usually short, consisting of a few questions focused on a small number of binary outcomes. They are quick to administer and analyze, sometimes taking only 1–2 days per district. When implemented using best practices, study teams select households randomly from an up-to-date sampling frame so the results are representative of the population. At other times, however, respondents are identified using a convenience sample approach in which only one person is selected in a truly random fashion and subsequent respondents are selected from nearby households. Additionally, in some LQAS protocols field staff stop collecting data when it becomes clear that the final few respondents will not affect the district's adequate/inadequate classification. In all cases where an LQAS design does not allow each respondent to have a known and non-zero probability of selection, the sampling is not truly random and the validity of results is threatened by selection bias.

Non-response bias is also a common source of selection bias. Non-response bias results when the answers to survey questions differ between selected respondents who participate in the survey and selected respondents who choose not to participate, or are unable to participate, in the survey. The response rate is measured as the proportion of individuals selected to take part in a survey who actually complete the questionnaire. The potential for non-response bias goes up as the response rate goes down. This type of bias is common in phone and internet-based surveys—interview methods more commonly used in high-income countries where populations have access to such technology. In such surveys, it is not uncommon to have response rates as low as 50%, which must be corrected for by using survey weighting and other techniques. In low- and middle-income countries where survey interviews are conducted in person with individuals in selected households, non-response bias is usually less of an issue. The protocol used by DHS, MICS, AIS, and MIS surveys for limiting non-response bias requires interviewers to return to selected households at least three times if selected individuals in the household are unavailable for the interview [1]. This practice greatly limits household and individual non-response, with response rates among eligible women of reproductive age in these surveys typically exceeding 95%. However, non-response bias can be a serious issue for estimates of MNCH intervention coverage obtained from surveys that do not follow this protocol. For example, EPI surveys do not require multiple visits to selected households and do not document how many potentially eligible households were skipped because the respondent was absent. [23]. Both DHS and MICS surveys report the response rates of households and of eligible women and men who were selected to participate in the survey. This practice should be followed in all survey reports, and where response rates are below 85%–90%, surveys results should be interpreted with caution.

## Recommendations

Our major recommendations for interpreting MNCH coverage estimates are summarized in Box 3, but here we will briefly draw together the recommendations we make in this review.

MNCH intervention coverage estimates should be interpreted taking into account the confidence interval around the estimates, especially when assessing trends over time in intervention coverage, or comparing coverage across population subgroups or between geographic regions. To allow this to occur, we recommend that confidence intervals around key health indicators are included in the tables in the main body of all survey reports (not just in the appendices of DHS and MICS reports as at present). MNCH intervention coverage estimates with a high degree of uncertainty (e.g., because they are based on small subpopulations) should be interpreted with caution. As surveys are designed to provide valid intervention coverage point estimates only for survey domains or at higher levels, coverage estimates below the survey domain level should also be interpreted with extreme caution (e.g., coverage estimates at the district level when the survey was designed to yield valid estimates only at higher administrative levels).

To limit information error and bias, we recommend that extensive pretesting is conducted prior to survey implementation. Where possible, the results of pretesting should be used to adapt survey questionnaires to local norms and cultures, while safeguarding the key structure of the questions needed to ascertain standardized MNCH indicators in a consistent manner across countries and over time.

To limit selection bias, we recommend that MNCH estimates are obtained using well-established multi-stage sampling protocols, such as those used by DHS and MICS surveys, which create accurate sampling frames of households for second stage selection, weight data for unequal probability of selection, and return to households at least three times to limit non-response bias.

Importantly, we commend recent calls for establishing reporting guidelines for survey research [32]. We recommend that all survey reports and papers presenting MNCH intervention coverage estimates follow similar recommendations to improve transparency, consistency, and interpretability of estimates, even if detailed methods have to be included as a web appendix (see Box 3).

Finally, to allow for the consideration of important sources of non-sampling error in interpreting MNCH coverage estimates, survey reports should include a limitations section that explicitly lists possible sources of non-sampling error (see Box 3). Descriptive statistics on data quality that impact non-sampling error should be considered when interpreting MNCH coverage estimates, including household, individual and question non-response, socio-

demographic composition of respondents, completeness of reporting, and date and age heaping. The DHS provides such information in an appendix to their country reports; other survey reports should do so as well.

## Conclusions

Estimates of MNCH intervention coverage obtained from household surveys are increasingly relied upon to assess progress in program effectiveness, and are a key metric in assessing Millennium Development Goals 4 and 5 for reducing neonatal, child, and maternal deaths worldwide. Surveys are the best tool we have for measuring MNCH intervention coverage at the population level and are central to national and global decision making but are subject to both sampling and non-sampling error. Non-sampling error is more insidious than sampling error and is difficult to quantify. It can, therefore, render the interpretation of MNCH intervention coverage estimates challenging. To this end, validation research that identifies the sources, direction, and magnitude of non-sampling error of MNCH intervention coverage estimates is urgently needed. Research presented elsewhere in this collection provides an excellent start [8–13], but future research must focus on refining and improving survey-based coverage estimates, and on developing a better understanding of how results on MNCH intervention coverage should be reported and interpreted.


## Acknowledgments

We wish to thank members of the working group on Improving Coverage Measurement through Large-scale Surveys for Maternal, Newborn and Child Health Interventions and the Child Health Epidemiology Reference Group (CHERG) for their comments, collaboration, and encouragement throughout this process. Jennifer Bryce is thanked for her comments on the draft manuscript.


## Author Contributions

Analyzed the data: TPE. Wrote the first draft of the manuscript: TPE. Contributed to the writing of the manuscript: DAR FTC JK RR AJDB FA. ICMJE criteria for authorship read and met: TPE DAR FTC JK RR AJDB FA. Agree with manuscript results and conclusions: TPE DAR FTC JK RR AJDB FA.

## References


1. Hancioglu A, Arnold F (2013) Measuring coverage in MNCH: Tracking progress in health for women and children using DHS and MICS household surveys. PLoS Med 10: e1001391. doi:10.1371/journal.pmed.1001391
2. Groves RM, Fowler FJ, Couper MP, Lepkowski JM, Singer E, et al. (2009) Survey methodology. Hoboken (New Jersey): John Wiley and Sons.
3. Levy PS, Lemenshow S (1999) Sampling of populations: methods and applications. New York: John Wiley and Sons Inc.
4. Payton ME, Greenstone MH, Schenker N (2003) Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? J Insect Sci 3: 34.
5. Schenker N, Gentleman JF (2001) On judging the significance of differences by examioning the overlap between confidence intervals. Am Stat 55: 182–186.
6. International CSAEaI (2012) Ethiopia Demographic and Health Survey 2011. Addis Ababa, Ethiopia and Calverton (Maryland): Central Statistical Agency and ICF International.
7. Central Statistics Authority (CSO), Ministry of Health (MOH), Tropical Disease Research Centre (TDRC), University of Zambia Macro International Inc. (2009) Zambia Demographic and Health Survey 2007. Calverton (Maryland): CSO and Macro International Inc.
8. Cutts FT, Izurieta H, Rhoda D (2013) Measuring coverage in MNCH: Design, implementation, and interpretation challenges associated with tracking vaccination coverage using household surveys. PLoS Med 10: e1001404. doi:10.1371/journal.pmed.1001404
9. Fischer Walker CL, Fontaine O, Black RE (2013) Measuring coverage in MNCH: Current indicators for measuring coverage of diarrhea treatment interventions and opportunities for improvement. PLoS Med 10: e1001385. doi:0.1371/journal.pmed.1001385
10. Campbell H, el Arifeen S, Hazir T, O'Kelly J, Bryce J, et al. (2013) Measuring coverage in MNCH: Challenges in monitoring the proportion of young children with pneumonia who receive antibiotic treatment. PLOS Med 10: e1001421. doi:10.1371/journal.pmed.1001421
11. Eisele TP, Silumbe K, Yukich J, Hamainza B, Keating J, et al. (2013) Measuring coverage in maternal and child health: Accuracy of measuring diagnosis and treatment of childhood malaria from household surveys in Zambia. PLoS Med 10: e1001417. doi:10.1371/journal.pmed.1001417
12. Hazir T, Bequem K, el Arifeen S, Khan AM, Huque MH, et al. (2013) Measuring coverage in MNCH: A prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia. PLoS Med 10: e1001422. doi:10.1371/journal.pmed.1001422
13. Stanton CK, Rawlins B, Drake M, dos Anjos M, Cantor D, et al. (2013) Measuring coverage in MNCH: Testing the validity of women's self-report of key maternal and newborn health interventions during the peripartum period in Mozambique. PLoS ONE 8: e60694. doi:10.1371/journal.pone.0060694
14. Suarez L, Simpson DM, Smith DR (1997) Errors and correlates in parental recall of child immunizations: effects on vaccination coverage estimates. Pediatrics 99: E3.
15. Valadez JJ, Weld LH (1992) Maternal recall error of child vaccination status in a developing nation. Am J Public Health 82: 120–122.
16. Rhodes AE, Lin E, Mustard CA (2002) Self-reported use of mental health services versus administrative records: should we care? Int J Methods Psychiatr Res 11: 125–133.
17. RBM (2008) Guidelines for Core Population-based Indicators Calverton, Maryland: Roll Back Malaria, MEASURE *Evaluation*, World Health Organization, UNICEF. New York: UNICEF.
18. Eisele TP, Macintyre K, Yukich J, Ghebremeskel T (2006) Interpreting household survey data intended to measure insecticide-treated bednet coverage: results from two surveys in Eritrea. Malar J 5: 36.
19. Skarbinski J, Winston CA, Massaga JJ, Kachur SP, Rowe AK (2008) Assessing the validity of health facility-based data on insecticide-treated bednet possession and use: comparison of data collected via health facility and household surveys–Lindi region and Rufiji district, Tanzania, 2005. Trop Med Int Health 13: 396–405.
20. Lemeshow S, Robinson D (1985) Surveys to measure programme coverage and impact: a review of the methodology used by the expanded programme on immunization. World Health Statistics Quarterly 38: 65–75.
21. Henderson RH, Sundaresan T (1982) Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. Bull World Health Organ 60: 253–260.
22. Cromwell EA, Ngondi J, McFarland D, King JD, Emerson PM (2012) Methods for estimating population coverage of mass distribution programmes: a review of practices in relation to trachoma control. Trans R Soc Trop Med Hyg 106: 588–595.
23. Turner AG, Magnani RJ, Shuaib M (1996) A not quite as quick but much cleaner alternative to the Expanded Programme on Immunization (EPI) Cluster Survey design. Int J Epidemiol 25: 198–203.
24. Luman ET, Worku A, Berhane Y, Martin R, Cairns L (2007) Comparison of two survey methodologies to assess vaccination coverage. Int J Epidemiol 36: 633–641.
25. Robertson SE, Valadez JJ (2006) Global review of health care surveys using lot quality assurance sampling (LQAS), 1984–2004. Soc Sci Med 63: 1648–1660.
26. Rhoda DA, Fernandez SA, Fitch DJ, Lemeshow S (2010) LQAS: user beware. Int J Epidemiol 39: 60–68.
27. Pezzoli L, Tchio R, Dzossa AD, Ndjomo S, Takeu A, et al. (2012) Clustered lot quality assurance sampling: a tool to monitor immunization coverage rapidly during a national yellow fever and polio vaccination campaign in Cameroon, May 2009. Epidemiol Infect 140: 14–26.
28. Pezzoli L, Andrews N, Ronveaux O (2010) Clustered lot quality assurance sampling to assess immunisation coverage: increasing rapidity and maintaining precision. Trop Med Int Health 15: 540–546.
29. Olives C, Pagano M (2010) Bayes-LQAS: classifying the prevalence of global acute malnutrition. Emerg Themes Epidemiol 7: 3.
30. Hedt BL, Casey O, Pagano M, Valadez JL (2008) Large country-lot quality assurance sampling: a new method for rapid monitoring and evaluation of health, nutrition and population programs at sub-national levels. Washington (D.C.): The World Bank. p 1–61.
31. Olives C, Valadez JJ, Brooker SJ, Pagano M (2012) Multiple category-lot quality assurance sampling: a new classification system with application to schistosomiasis control. PLoS Negl Trop Dis 6: e1806. doi:10.1371/journal.pntd.0001806
32. Bennett C, Khangura S, Brehaut JC, Graham ID, Moher D, et al. (2011) Reporting guidelines for survey research: an analysis of published guidance and reporting practices. PLoS Med 8: e1001069. doi:10.1371/journal.pmed.1001069
33. Central Statistical Agency [Ethiopia] and ICF International (2012) Ethiopia Demographic and Health Survey 2011. Addis Ababa, Ethiopia and Calverton (Maryland): Central Statistical Agency and ICF International. Available: http://www.measuredhs.com/pubs/pdf/FR255/FR255.pdf. Accessed 11 December 2012.

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: Evaluation of Community-Based Treatment of Childhood Illnesses through Household Surveys

**Elizabeth Hazel*, Jennifer Requejo, Julia David, Jennifer Bryce**

Institute for International Programs, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America

**Abstract:** Community case management (CCM) is a strategy for training and supporting workers at the community level to provide treatment for the three major childhood diseases—diarrhea, fever (indicative of malaria), and pneumonia—as a complement to facility-based care. Many low- and middle-income countries are now implementing CCM and need to evaluate whether adoption of the strategy is associated with increases in treatment coverage. In this review, we assess the extent to which large-scale, national household surveys can serve as sources of baseline data for evaluating trends in community-based treatment coverage for childhood illnesses. Our examination of the questionnaires used in Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) conducted between 2005 and 2010 in five sub-Saharan African countries shows that questions on care seeking that included a locally adapted option for a community-based provider were present in all the DHS surveys and in some MICS surveys. Most of the surveys also assessed whether appropriate treatments were available, but only one survey collected information on the place of treatment for all three illnesses. This absence of baseline data on treatment source in household surveys will limit efforts to evaluate the effects of the introduction of CCM strategies in the study countries. We recommend alternative analysis plans for assessing CCM programs using household survey data that depend on baseline data availability and on the timing of CCM policy implementation.

*This paper is part of the* PLOS Medicine "*Measuring Coverage in MNCH" Collection*

## Introduction

Most low- and middle-income countries are making slow progress in addressing child mortality—too slow to achieve Millennium Development Goal 4 by 2015 [1]. Diarrhea, pneumonia, and malaria account for 37% of under-five deaths worldwide [2], with only about one-third of children with these illnesses receiving appropriate treatment [3]. To address this disease burden and treatment gap, governments and donors in 52 of the Countdown to 2015 priority countries had adopted community case management (CCM) of childhood illness strategies by 2011 [1,4]. CCM aims to extend the treatment of childhood illnesses from health facilities into communities [5–7] by training and supporting existing or newly recruited community health workers to provide treatment for neonatal conditions and simple cases of childhood pneumonia, diarrhea, and malaria at the community level and to refer cases of more severe illness. The underlying assumption of CCM is that the expansion of treatment capabilities to community health workers will result in increases in access to and coverage of treatment, especially for children living in households far removed from existing health facilities [6].

Clearly, it is essential that countries introducing CCM carefully assess its contribution to increased treatment coverage for childhood illnesses. To do this, population-level data on the place of treatment and the type of health provider are needed. Where these data exist, they can serve as a baseline for evaluations of the contribution of CCM to treatment coverage going forward. Where they do not exist, evaluators will need alternative analytical designs. In practice, routine health information systems in low- and middle-income countries are often weak and cannot consistently provide valid coverage data for these treatment indicators. Often the best available source of coverage data in these countries is nationally representative household surveys [8]. Two major programs generate the household-level survey data needed to measure coverage for maternal, newborn, and child health interventions in low- and middle-income countries—the Demographic and Health Surveys (DHS), supported by USAID [9], and the Multiple Indicator Cluster Surveys (MICS), supported by UNICEF [10]. However, although questions on the coverage of care seeking and appropriate treatment of childhood illnesses have been included in DHS and MICS protocols in the past, it may not be possible to use these data to determine whether a treatment was delivered by a health facility or by a community-based health worker, information that is needed to assess the success of CCM.

In this review, which is part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, we assess the extent to which

**Citation:** Hazel E, Requejo J, David J, Bryce J (2013) Measuring Coverage in MNCH: Evaluation of Community-Based Treatment of Childhood Illnesses through Household Surveys. PLoS Med 10(5): e1001384. doi:10.1371/journal.pmed.1001384

**Academic Editor:** Nyovani Madise, Professor of Demography and Social Statistics, University of Southampton, United Kingdom

**Published** May 7, 2013



**Funding:** This work was conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF, with financial support from The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** CCM, community case management; DHS, Demographic and Health Survey(s); MICS, Multiple Indicator Cluster Survey(s)

* E-mail: ehazel@jhsph.edu

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

existing household surveys provide the data needed to measure trends in coverage for the correct management of childhood illnesses by place of treatment (health facility or community) and type of provider. In addition, we recommend alternative analysis plans that might be used in settings where baseline data are insufficient to measure trends in treatment coverage by the place and type of provider.

## Assessing the Surveys

For our assessment, we focused on Ethiopia, Ghana, Malawi, Mali, and Niger, five countries where the Catalytic Initiative to Save a Million Lives, a partnership of donors and United Nations agencies, is supporting CCM as a strategy to accelerate coverage for the treatment of childhood illnesses [11]. In total, we reviewed the locally adapted questionnaires from nine DHS and MICS surveys conducted in these countries since 2005 along with the most recent DHS and MICS core questionnaires.

We assessed each survey to determine whether it could provide information on the place of treatment—in a health facility or in the community—for a child reported to have symptoms of pneumonia, fever, and/or diarrhea, and information about the provider of the treatment. We defined symptoms of pneumonia as cough and rapid or difficult breathing, although DHS refers to these symptoms as "symptoms of acute respiratory infection" and MICS refers to them as indicating "suspected pneumonia." Fever is used as a symptom of malaria in CMM strategies. We reviewed question wording, question placement, skip patterns, and the sample surveyed for each survey questionnaire. Table S1 shows all the information on place of treatment and care seeking for these three major childhood illnesses collected by the surveys in each country.

We assumed that CCM provided no treatments in the study countries as a part of the formal health system prior to the adoption of illness-specific CCM policies. Table 1 shows a summary of national CCM policies, the cadre of the CCM worker, and the date of reported CCM policy implementation for the five countries. We show the date of national policy implementation rather than policy adoption, because procurement issues often prevented full CCM implementation at the time of adoption of the policy (for example, treatment of diarrhea with zinc often lagged behind treatment with oral replacement salts). We examined the locally adapted questionnaire for each survey to assess whether the data set could be used to report on three constructs (care seeking, appropriate treatment, and place of treatment) for each of three childhood illnesses (pneumonia, diarrhea, and fever). We assigned each survey a score of "yes" or "no" for each of the three constructs and illnesses. We assigned a "partial" score if some information was available but was not sufficient to determine the exact source of care or treatment.

Care seeking refers to reports by caregivers about whether, and if so where, they took the sick child for care. We assigned a "yes" score to a survey for care seeking if there were specific questions about where or from whom the mother or caregiver sought advice or treatment that allowed us to determine whether care was sought from a health facility or at the community level.

We defined "appropriate" treatment as the first-line treatment recommended by the CCM policy in each country (Table 1). We assigned a "yes" score to a survey for appropriate treatment only if

**Table 1.** Summary of national CCM policies, cadre of worker, and date of policy implementation for the five countries.

| Country | Cadre of CCM Worker | Child Illness | 1st-Line Treatment | Date of CCM Policy Implementation |
|---|---|---|---|---|
| Ethiopia | Health extension worker | Diarrhea | ORS | 2004 [18] |
| | | | Zinc | 2012[a] |
| | | Pneumonia | Cotrimoxazole | 2011 [19] |
| | | Fever | Artemether/lumefantrine/Chloroquine | 2004 [18] |
| Ghana | Community-based agent | Diarrhea | ORS | 2004 [20] |
| | | | Zinc | 2010 [19] |
| | | Pneumonia | Amoxicillin | 2010 [19] |
| | | Fever | Chloroquine | 2003 [20] |
| | | | ACT | 2007 [20] |
| Malawi | Health surveillance assistants | Diarrhea | ORS | 2008 [19] |
| | | | Zinc | 2010[b] |
| | | Pneumonia | Cotrimoxazole | 2008 [19] |
| | | Fever | Coartem | 2008 [19] |
| Mali | *Agents de santé communautaires* | Diarrhea | ORS | 2010 [19] |
| | | | Zinc | 2010 [19] |
| | | Pneumonia | Amoxicillin[c] | 2010 [19] |
| | | Fever | ACT | 2010 [19] |
| Niger | *Agents de santé communautaires* | Diarrhea | ORS | 2006 [21] |
| | | | Zinc | 2008 [21] |
| | | Pneumonia | Cotrimoxazole | 2008 [21] |
| | | Fever | ACT | 2008 [21] |

[a]Personal communication, Tedbabe Degefie.
[b]Personal communication, Humphreys Nsona.
[c]Personal communication, Hamadoun Sangho.
ACT, artemisinin combination therapy; ORS, oral replacement salts.
doi:10.1371/journal.pmed.1001384.t001

the survey included specific questions on treatment options. Appropriate treatments included oral replacement salts and/or zinc for diarrhea (depending upon when countries implemented policies for zinc), antibiotics for pneumonia, and specific antimalarials (sulfadoxine and pyrimethamine, artemisinin-combination therapies, etc.) for fever.

Place of treatment refers to the location (health facility or community) where the treatment was "received," as reported by the child's mother or caregiver. In this context, "received" may mean that the mother or caregiver received either the actual medicine or a prescription for the medicine to be filled elsewhere. Response options to these questions varied across the surveys, but generally included both public and private health facilities, pharmacies and drug venders/shops, and, sometimes, specific community-based providers such as community health workers and/or traditional healers. We assigned a "yes" score for place of treatment if we were able to determine unambiguously whether the child received the treatment at a health facility or at the community level. We noted whether information was available to determine the specific community site of treatment (e.g., mobile clinic) and the specific community treatment provider (e.g., community health workers).

In addition, for surveys that did not include information on the place of treatment, we were sometimes able to draw plausible inferences about this construct based on information about care seeking or intention to treat and appropriate treatment. Thus, if the country policy at the time of the survey allowed treatment by community health workers (as opposed to referral to a health facility), it may be possible to assume that children reported as having been taken to a community health worker received treatment for their illness from that individual. Moreover, for surveys without direct questions on place of treatment, inferences based on care seeking can be strengthened if data are available on appropriate treatment. In other words, in a context where government policies support CCM, a mother who reports seeking care for her child from a community health worker *and* who reports receiving appropriate treatment can be assumed to have benefitted from CCM with more certainty than a mother who reports only that she sought care from a community health worker.

## Do Household Surveys Contain Baseline Data on CCM?

Table 2 summarizes the availability of information on care seeking, treatment, and point-of-treatment for each of the three childhood illnesses in the large-scale household surveys conducted in the five countries that we studied. We grouped the results by type of survey (DHS versus MICS) to reduce redundancy.

### DHS Surveys

The Mali 2006 and Niger 2006 surveys occurred before the implementation of the national CCM policy (Table 1). Ghana (2008) and Ethiopia (2005) had implemented CCM policies at the time of the surveys for diarrhea (excluding zinc treatment) and fever, but not for pneumonia. Ethiopia had implemented the CCM policy for pneumonia by the time of the 2011 DHS. Malawi was implementing CCM policies for all three illnesses at the time of the 2010 DHS.

All the DHS surveys included specific questions on care seeking. Response options for the site of care included a mix of both community and health facility sites and providers in the public and private sectors. The surveys for Ethiopia, Ghana, Mali, and Malawi included a community health worker as a possible response option for the care-seeking questions for the three childhood illnesses reviewed. The Niger 2006 DHS listed a community health worker for diarrhea illness care seeking only.

Treatment questions were available in all the DHS surveys with the exception of pneumonia treatment in the 2006 surveys conducted in Mali and Niger. In addition, there were no questions on zinc treatment for diarrhea in these two surveys; however, these countries had not incorporated zinc into their national policies by the time of the surveys.

Four of the DHS surveys (Ghana 2008, Malawi 2010, and Ethiopia 2005 and 2011) and the phase 6 core questionnaire included a follow-up question for mothers reporting that care had been sought at more than one site that was designed to determine the site where care was sought first for each illness.

Explicit place of treatment information was not available from DHS surveys conducted since 2005 in any of five countries we considered and was also not included in the DHS phase 6 core questionnaire. However, as explained earlier, we found that information on place of treatment could be gleaned from the analysis of care-seeking data collected in DHS surveys, although the results were limited in important ways. For example, independent question sequences in the DHS surveys asked a mother about whether (and where) she sought advice or treatment for her child, and about the treatment received. Both questions offered locally adapted response options that included both facility and community sites, so one could assume that if the child received appropriate treatment, it was through the reported site or provider. Because the care-seeking question allowed for multiple responses, the sequence through which care was sought from specific sites was not captured (except in the questionnaires that include a follow-up question to identify first site of care seeking). Thus, there was no way to determine through the existing questions what treatment, if any, was given at the site(s) where care was sought, unless only one source of care seeking was mentioned, in which case one could presume (possibly incorrectly) the treatment was received there.

### MICS Surveys

Ghana had implemented CCM policies for diarrhea and fever treatment at the time of its MICS survey, but neither Mali not Malawi had implemented a national CCM policy at the time of their surveys (Table 1).

Questions about care seeking for pneumonia were included in all four MICS surveys examined here (Table 2). The response options included community health workers in the Ghana and Malawi surveys and in the core MICS4 questionnaires. Information on care seeking for diarrhea was not captured by these surveys. The Ghana and Malawi surveys as well as the MICS4 core questionnaire partially captured information on care seeking for fever through questions about whether a child with fever was taken to a health facility.

All the MICS surveys included questions about the treatments received for children reporting symptoms of pneumonia, diarrhea, or fever. The Malawi and Ghana 2006 surveys did not include questions on zinc; however, a policy recommending treatment of diarrhea with zinc had not been implemented at the time of the surveys in the two countries. Two country surveys (Malawi 2006 and Mali 2009) and the MICS4 core questionnaire asked a follow-up question for children reported to have received treatment for fever to identify which drugs were given or prescribed through a health facility, which could be used to assess the validity of responses.

Finally, Table 2 shows that information on the proportion of children treated by a community health worker was captured in the 2006 Ghana MICS survey for the three major illnesses. Partial

**Table 2.** Information on care seeking, treatment, and point-of-treatment available from the large-scale household surveys for the five countries.

| Country and Survey | Care Seeking: | | | Appropriate Treatment: | | Place of Appropriate Treatment: | |
|---|---|---|---|---|---|---|---|
| | Available | CCM Health Provider Listed as Possible Response | Additional Information | Available | Additional Information | Available | Treatment Provider (CCM) |
| **Ethiopia 2005 DHS** | Yes for all illnesses | Community health agent | Indicates first care-seeking site | Yes for all illnesses | For **pneumonia** and **fever** illnesses: asks if drug was obtained outside the home. | No | N/A |
| **Niger 2006 DHS** | Yes for all illnesses | For **diarrhea** illness: *Agent santé communautaire*<br>For **pneumonia**/**fever**: no community health worker listed | No | Yes, for **diarrhea** and **fever** only | For **fever** illness: asks if drug was obtained outside the home. | No | N/A |
| **Mali 2006 DHS** | Yes for all illnesses | *Agent santé communautaire* | No | Yes, for **diarrhea** and **fever** only | For **fever** illness: asks if drug was obtained outside the home. | No | N/A |
| **Ghana 2008 DHS** | Yes for all illnesses | Government and nongovernment fieldworker | Indicates first care-seeking site | Yes for all illnesses | For **pneumonia** and **fever** illnesses: Asks if drug was obtained outside the home. | No | N/A |
| **Malawi 2010 DHS** | Yes for all illnesses | Government and nongovernment health surveillance assistants | Indicates first care-seeking site | Yes for all illnesses | No | No | N/A |
| **Ethiopia 2011 DHS** | Yes for all illnesses | Government health post/ health extension workers | Indicates first care-seeking site | Yes for all illnesses | No | No | N/A |
| **DHS Phase 6 Core 2008–2013** | Yes for all illnesses | Government and nongovernment fieldworker | Indicates first care-seeking site | Yes for all illnesses | No | No | N/A |
| **Ghana 2006 MICS** | Yes, for **pneumonia** only; Partial[a] for **fever** only | For **pneumonia** illness: village health worker<br>For **fever** illness: no community health workers listed | For **fever** illness: asks if child was taken to a health facility; no specific sites are listed | Yes for all illnesses | No | Yes for all illnesses | Village health worker |
| **Malawi 2006 MICS** | Yes, for **pneumonia** only; Partial[a] for **fever** only | For **pneumonia** illness: village health worker<br>For **fever** illness: no community health workers listed | For **fever** illness: asks if child was taken to a health facility; no specific sites are listed | Yes for all illnesses | No | Partial[a], for **fever** only | For **fever** illness: asks about specific medicines that were given at the health facility |
| **Mali 2009 MICS** | Yes, for **pneumonia** and **fever** only | For **pneumonia** and **fever** illnesses: no community health workers listed | No | Yes for all illnesses | No | Partial[a], for **fever** only | For **fever** illness: asks about specific medicines that were given at the health facility |
| **MICS4 Core (2009–2011)** | Yes, for **pneumonia** only; Partial[a] for **fever** only | For **pneumonia** illness: village health worker | For **fever** illness: asks if child was taken to a health facility; no specific sites are listed | Yes for all illnesses | No | Partial[a], for **fever** only | For **fever** illness: asks about specific medicines that were given at the health facility |

[a]Partial indicates that some information is available but not enough to determine exact source of care or treatment.

doi:10.1371/journal.pmed.1001384.t002

questions on place of treatment for fever were included in the Mali (2009) and Malawi (2006) MICS surveys and in the MICS4 core questionnaire.

## An Alternative Analysis Plan for Program Implementers and Evaluators

This analysis shows that, with the exception of the 2006 Ghana MICS survey, comprehensive baseline data on the place and provider of appropriate treatment of childhood pneumonia, diarrhea, and fever are not available from the major household surveys conducted in the study countries before 2010. However, evaluators can still answer some important questions about the effectiveness of CCM in the study countries, especially if future surveys are designed to capture these data. For instance, one of the major global questions is how the introduction of CCM could affect care seeking at health facilities. Evaluators can compare survey reports of care seeking from all sources at baseline (prior to CCM program implementation) with the sum of care seeking rates from health facilities and from community-based health workers at midline (during CCM program implementation) and endline (the time of the CCM program evaluation). They can then use routine data on utilization to assess whether CCM has contributed to overall increases in treatment coverage. This approach is limited in most low- and middle-income countries due to the poor quality of routine data on service utilization [8]. Another possible approach would be to compare illness-specific care-seeking rates from health facilities at baseline with the sum of care-seeking rates from health facilities and from community-based health workers at midline and endline.

Figure 1 provides a flow chart of alternative analysis plans designed to determine the effects of CCM on treatment coverage, treatment source, and care-seeking source, depending on the types of baseline data that are available in a given context. Consider, for example, a hypothetical setting where a DHS or MICS survey collects information on both care seeking and treatment but not on place of treatment prior to implementation of a CCM program (baseline). After this survey is conducted, a CCM policy is implemented, and the government requests a time-trend analysis on the impact of the CCM program on childhood illnesses. If an endline survey collects comparable information on care seeking and treatment but also collects information on place of treatment, then a time-trend analysis is possible for the change in overall treatment coverage. Changes in care seeking outside health facilities can also be assessed to determine the impact of the CCM program on care-seeking patterns. Moreover, in this hypothetical setting, because it can be assumed that treatment through CCM-trained community health worker at baseline is

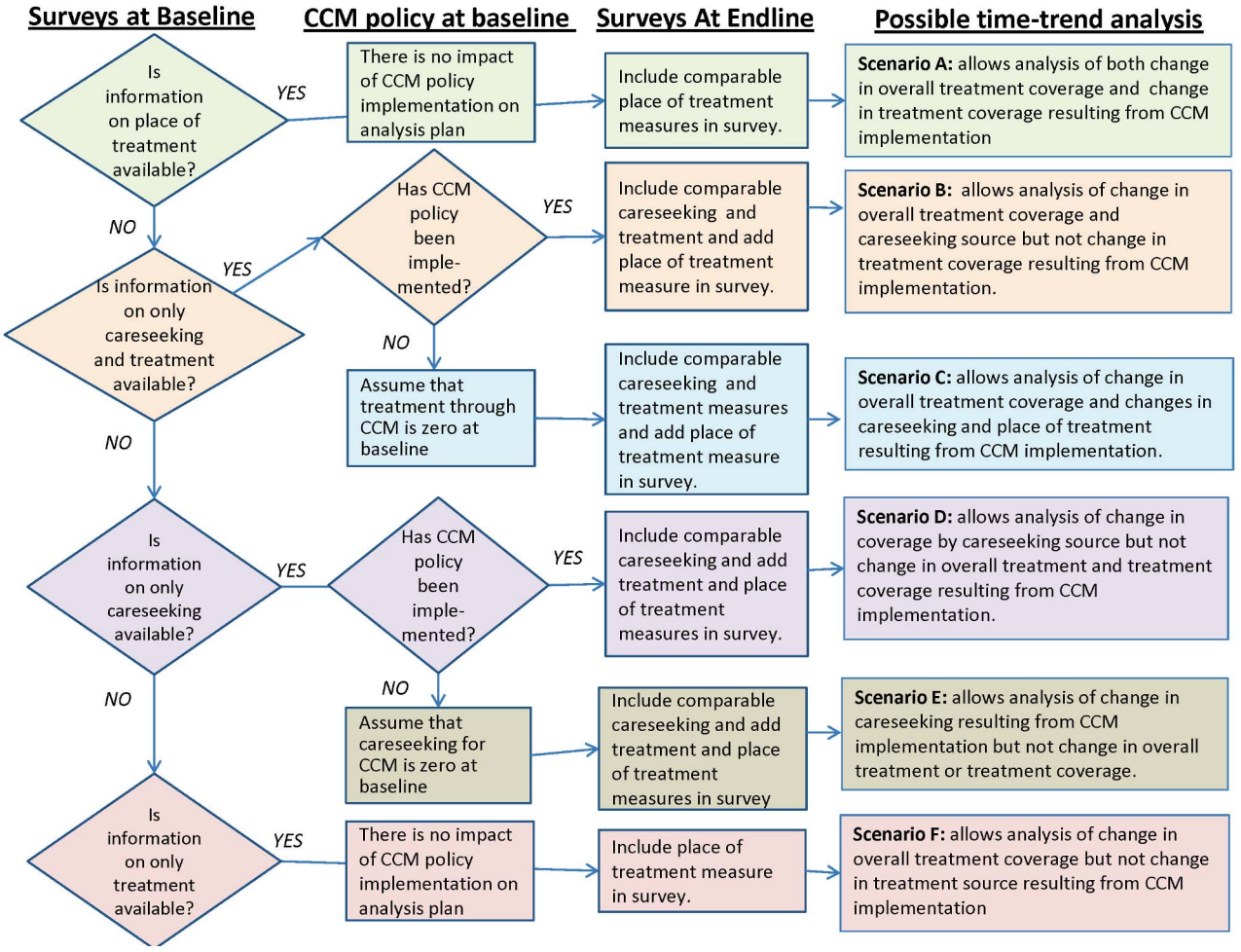


**Figure 1. Analysis plan flow chart.** Decision flow chart for six scenarios of time-trend analysis options depending on baseline data availability and timing of CCM policy implementation. CCM policies have been implemented at endline in all cases.
doi:10.1371/journal.pmed.1001384.g001

zero, increases in community health workers as the point of treatment can be assessed (Figure 1, Scenario C).

By contrast, consider a setting where there is information only on the treatment available at baseline and not on either the place of treatment or care seeking. In this scenario, the timing of the CCM policy implementation is irrelevant since no measure of treatment or care-seeking source is collected through the baseline survey. If information on place of treatment, as well as on the actual treatment, is collected in an endline survey, then time-trend analysis is possible for the changes in overall treatment coverage that is related to the implementation of CCM but not for the change in overall treatment source (Figure 1, Scenario F).

Importantly, given that the CCM policy for the three major childhood illnesses may be implemented at different times and the variation in the availability of care seeking, treatment and place of treatment data by illness, evaluators and program implementers may need to investigate alternative illness-specific analyses.

## Limitations in Measuring the Source of Treatment through Household Surveys

The sample sizes needed to determine whether changes in treatment rates for specific childhood illnesses are statistically significant will vary depending on the prevalence of each disease (or its presumed prevalence based on respondents' reports of signs and symptoms) and estimated levels of appropriate treatment at baseline. Increasing the sample sizes will increase the costs and logistical challenges of capturing the information through a DHS/MICS survey, as discussed in another paper in the Collection [12], as will looking at place of treatment by wealth status to determine whether CCM and other service delivery strategies designed to reach the poor are effective in reducing inequities. Survey planners will need to consider whether the results obtained from such analyses will justify the resources required, and evaluators of CCM will need to interpret inadequately powered analyses with care.

There are other important limitations in using household surveys to measure appropriate treatment of childhood illnesses [13–15]. For example, the analyses rely on respondents' ability to recognize, recall, and report signs and symptoms correctly and to be able to accurately recall and report care-seeking patterns, where treatments were obtained, and when/how often they were given to the child. Work is underway to improve these measurements, as reflected by the other papers in this Collection. Currently, DHS and MICS surveys can provide only limited information on whether the child was appropriately assessed and on whether adequate treatment was given by the health worker. Household surveys may be analyzed in conjunction with community health worker quality-of-care surveys, however, to provide information on correct assessment, classification, and treatment [16].

## Recommendations for Future Survey Protocols

The introduction of CCM provides a good example of the need for flexibility and continuous evolution in the major household surveys used to assess intervention coverage in low- and middle-income countries. Before 2005, few governments had authorized community health workers to provide treatment for childhood illnesses, and exceptions were limited to the provision of oral replacement salts for diarrhea. By 2011, 52 of the Countdown to 2015 priority countries had adopted CCM policies and had moved forward with implementation [4]. The widespread adoption of this and any new strategy, combined with a growing recognition of the need for evidence-based evaluations of program effectiveness, underline the continuing need for modifications in household survey protocols so that the population impact of specific strategies and interventions can be evaluated. We recommend strongly that all future coverage surveys include standard questions on the place and provider of treatment. Notably, when we shared our findings with technical staff at both DHS and MICS, they received our results positively and agreed to consider including questions capturing information on both place and provider of treatment in future surveys. The responsiveness of these survey programs to our suggestions should help to ensure the adequate capture of important changes in service delivery at a population level.

Importantly, additional questions incorporated into DHS and MICS surveys will need to be adapted to each country context. For example, it will be important to conduct a pretest to determine whether respondents are able to identify the community health worker trained to provide CCM in their area. Program implementers can contribute to the validity of such measurements by introducing strategies that will help child caregivers remember and report accurately about care received from a CCM-trained community health worker. In an assessment of village health workers delivering CCM in Bangladesh, for example, each worker was given a bright pink bag that was shown in pretests to be easy for mothers to remember, and to help them distinguish the CCM worker from other community-level workers [17]. Similar context-specific strategies to increase the salience and recall of an encounter with a community health worker trained in CCM are likely to increase the validity of place of treatment reports.

Those interested in evaluating CCM must also consider the potential role of informal providers in providing treatments for childhood illnesses, and must ensure that context-specific response options are included in the survey protocol to separate community health workers and informal providers at the community level. Finally, globally, more studies are needed to explore care seeking. In particular, care seeking that involves the informal sector needs to be better studied, and the impact of such care-seeking behavior on CCM program needs to be investigated. Information from such studies, and consideration of the other recommendations we make

### Key Points

- Low-income countries are increasingly adopting community case management (CCM) as a strategy for increasing the coverage of appropriate treatments for childhood illnesses.
- CCM program managers need to evaluate the effectiveness of their programs through time trend analyses that investigate whether treatment is being received at the community level.
- Population-based household surveys such as the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) are currently the only available means of obtaining data on treatment coverage by source of treatment in low-income countries.
- In an assessment of nine DHS/MICS surveys in five countries that are currently implementing CCM, we found that, although data on care seeking and treatment coverage are available, there is limited information on treatment source.
- We recommend that treatment source questions be included in future survey protocols, a recommendation that technical staff at DHS and MICS are now considering; we also recommend alternative analysis plans that implementers and evaluators may use to assess CCM programs.

above, will ensure that the questions included in future surveys are correctly designed to provide the information that evaluators of CCM and other strategies need.

## Supporting Information

**Table S1 Summary of illness care seeking, treatment, and place of treatment available by survey.** Details of questions used to measure care seeking, treatment, and place of treatment by country-specific survey and the core questionnaires. (DOCX)

## Acknowledgments

Many individuals contributed to this work, and we mention some of them here. Mark Young of UNICEF-New York raised the initial question about whether it would be possible to assess trends in place of treatment in the Catalytic Initiative to Save a Million Lives countries, which stimulated this work. Technical staff from DHS and MICS reviewed early drafts of this report and provided valuable feedback: Fred Arnold, Tessa Wardlaw, and Holly Newby. Members of the Catalytic Initiative to Save a Million Lives evaluation team based at the Institute for International Programs at the Johns Hopkins Bloomberg School of Public Health reviewed drafts and provided helpful comments, including Agbessi Amouzou, Robert Black, Kate Gilroy, Ken Hill, Marjorie Opuni, Cesar Victora, and others, as did Shams el Arifeen of ICDDR,B. We would also like to acknowledge the Core Group on Coverage Measurement, Child Health Epidemiology Reference Group (CHERG) for continuously striving to improve child health coverage measures. Members include Jennifer Bryce, Fred Arnold, Ann Blanc, Attila Hancioglu, Holly Newby, Jennifer Harris Requejo and Tessa Wardlaw many of whom, as mentioned above, provided valuable feedback on the paper.

## Author Contributions

Analyzed the data: EH JR JD JB. Wrote the first draft of the manuscript: EH. Contributed to the writing of the manuscript: EH JR JD JB. ICMJE criteria for authorship read and met: EH JR JD JB. Agree with manuscript results and conclusions: EH JR JD JB.

## References

1. Requejo J, Bryce J, Victora C, Countdown to 2015 group. Countdown to 2015 for Maternal and Child Survival: the 2012 Report. World Health Organization and UNICEF 2012. Available: http://www.countdown2015mnch.org/documents/2012Report/2012-Complete.pdf. Accessed 1 November 2012.
2. Liu L, Johnson HL, Cousens S, Perin J, Scott S, et al. (2012) Global, regional and national causes of child mortality 2000–2010: an updated systematic analysis. Lancet 379 (9832): 2151–2161.
3. Bhutta ZA, Chopra M, Axleson H, Berman P, Boerma T, et al. (2010). Countdown to 2015 decade report (2000–10): taking stock of maternal, newborn, and child survival. Lancet 375: 2032–2044.
4. de Sousa A, Tiedje KE, Recht J, Bjelicd I, Hamere DH. (2012) Community case management of childhood illnesses: policy and implementation in Countdown to 2015 countries. Bull WHO 90: 183–190.
5. Sixty-Third World Health Assembly (WHA63.24). Accelerated progress towards achievement of Millennium Development Goal 4 to reduce child mortality: prevention and treatment of pneumonia, 2010. Available: http://apps.who.int/gb/ebwha/pdf_files/WHA63-REC1/WHA63_REC1-P2-en.pdf. Accessed 14 June 2012.
6. Marsh D, Gilroy K, Van de Weerdt, R Wansi E, Qazi S (2008) Community case management of pneumonia: at a tipping point? Bull WHO 86: 381–389.
7. UNICEF (2010) Narrowing the Gaps to Meet the Goals. Available: http://www.unicef.org/publications/files/Narrowing_the_Gaps_to_Meet_the_Goals_090310_2a.pdf. Accessed 14 June 2012.
8. Countdown to 2015, Health Metrics Network, World Health Organization and UNICEF (2011) Monitoring maternal, newborn and child health: Understanding key progress indicators. Available: http://www.who.int/healthmetrics/news/monitoring_maternal_newborn_child_health.pdf.
9. MEASURE DHS Demographic and Health Surveys. Available: http://www.measuredhs.com/. Accessed 14 June 2012.
10. UNICEF. Statistics and Monitoring: Multiple Indicator Cluster Survey. Available: http://www.unicef.org/statistics/index_24302.html. Accessed 14 June 2012.
11. Canadian International Development Agency: The Catalytic Initiative to Save a Million Lives. Available: http://www.acdi-cida.gc.ca/acdi-cida/acdi-cida.nsf/eng/NAD-1249841-JLG#cidaclf2fix4. Accessed 14 June 2012.
12. Hancioglu A, Arnold F (2013) Measuring coverage in MNCH: Tracking progress in health for women and children using DHS and MICS household surveys. PLoS Med 10: e1001391. doi:10.1371/journal.pmed.1001391.
13. Campbell H, el Arifeen S, Hazir H, O'Kelly J, Bryce J, et al. (2013) Measuring coverage in MNCH: Challenges in monitoring the proportion of young children with pneumonia who receive antibiotic treatment. PLoS Med 10: e1001421. doi:10.1371/journal.pmed.1001421.
14. Eisele T, Rhoda DA, Cutts FT, Keating J, Ren R, et al. (2013) Measuring coverage in MNCH: Total survey error and the interpretation of intervention coverage estimates from household surveys. PLoS Med 10: e1001386. doi:10.1371/journal.pmed.1001386.
15. Fischer Walker CL, Fontaine O, Black RE (2013) Measuring coverage in MNCH: Current indicators for measuring coverage of diarrhea treatment interventions and opportunities for improvement. PLoS Med 10: e1001385. doi:10.1371/journal.pmed.1001385.
16. Gilroy KE, Callaghan-Koru J, Cardemil CV, Nsona H, Amouzou A, et al. (2012) Quality of sick child care delivered by Health Surveillance Assistants in Malawi. Health Policy Plan 2012: 1–13. Available: http://heapol.oxfordjournals.org/content/early/2012/10/13/heapol.czs095.full.pdf#page = 1&view = FitH. Accessed 5 November 2012.
17. Arifeen SE, Hoque DME, Akter T, Rahman M, Hoque ME, et al. (2009). Effect of the Integrated Management of Childhood Illness strategy on childhood mortality and nutrition in a rural area in Bangladesh: a cluster randomised trial. Lancet 374: 393–400.
18. Federal Ministry of Health and Regional Health Bureaus. (2008). Ethiopia Health Sector Development Programme HSDP III Mid-term Review. Available: http://www.moh.gov.et/mobile/Resources/Main%20Report%20Final%2012[1].07.pdf. Accessed 14 June 2012.
19. Oliver K, Young M, Oliphant N, Diaz T, Kim J (2012) Review of Systematic Challenges to the Scale-up of Integrated Community Case Management. Maternal, Newborn and Child Health Working Paper. Available: http://www.unicef.org/infobycountry/files/Analysis_of_Systematic_Barriers_cover_1163.pdf. Accessed 14 June 2012.
20. Institute for International Programs - Johns Hopkins University (2008) Final Report for the Retrospective Evaluation of the Accelerated Child Survival and Development Program: Ghana. Available: http://www.jhsph.edu/dept/ih/IIP/projects/acsd/Ghana.pdf. Accessed 14 June 2012.
21. Amouzou A, Oumarou H, Bensaid K, the Niger Countdown Case Study Working Group (2012) Reduction in child mortality in Niger: a Countdown to 2015 country case study. Lancet 380: 1169–1178.

OPEN ACCESS Freely available online

PLOS | MEDICINE

Review

# Measuring Coverage in MNCH: Challenges and Opportunities in the Selection of Coverage Indicators for Global Monitoring

**Jennifer Harris Requejo[1]*, Holly Newby[2], Jennifer Bryce[1]**

**1** Institute for International Programs, Department of International Health, The Johns Hopkins University, Baltimore, Maryland, United States of America, **2** Statistics and Monitoring Section, Division of Policy and Strategy, UNICEF, New York, New York, United States of America

**Abstract:** Global monitoring of intervention coverage is a cornerstone of international efforts to improve reproductive, maternal, newborn, and child health. In this review, we examine the process and implications of selecting a core set of coverage indicators for global monitoring, using as examples the processes used by the Countdown to 2015 for Maternal, Newborn and Child Survival and the Commission on Accountability for Women's and Children's Health. We describe how the generation of data for global monitoring involves five iterative steps: development of standard indicator definitions and measurement approaches to ensure comparability across countries; collection of high-quality data at the country level; compilation of country data at the global level; organization of global databases; and rounds of data quality checking. Regular and rigorous technical review processes that involve high-level decision makers and experts familiar with indicator measurement are needed to maximize uptake and to ensure that indicators used for global monitoring are selected on the basis of available evidence of intervention effectiveness, feasibility of measurement, and data availability as well as programmatic relevance. Experience from recent initiatives illustrates the challenges of striking this balance as well as strategies for reducing the tensions inherent in the indicator selection process. We conclude that more attention and continued investment need to be directed to global monitoring, to support both the process of global database development and the selection of sets of coverage indicators to promote accountability. The stakes are high, because these indicators can drive policy and program development at the country and global level, and ultimately impact the health of women and children and the communities where they live.

*This paper is part of the* PLOS Medicine "*Measuring Coverage in MNCH" Collection.*

## Introduction

Global monitoring of coverage for maternal and child health interventions involves the collection and analysis of a limited set of quantitative indicators to assess progress, and is central to international efforts to improve reproductive, maternal, newborn, and child health (RMNCH). Decision makers use results from global monitoring to set priorities and to determine where to allocate resources [1]. Coverage measures are a major focus of global monitoring because they can change much more rapidly in response to policy and program interventions than measures of impact (e.g., mortality, morbidity, fertility, nutritional status). Coverage refers to the proportion of a population in need of a public health intervention that actually receives it [2].

In this article, which is part of the *PLOS Medicine* "Measuring Coverage in MNCH" Collection, we review the steps involved in producing data of adequate quality for use in global monitoring. We also provide a critical analysis of the sets of coverage indicators included in the Countdown to 2015 for Maternal, Newborn and Child Survival and the Commission on Information and Accountability for Women's and Children's Health initiatives (Table 1), which are referred to as "Countdown" and "the Commission," respectively, throughout the rest of this review. Because global monitoring results affect the lives of women and children, it is critical that the "right" indicators are assessed and correctly interpreted. Thus, our aim in this article is to recommend improvements in the process used to select sets of coverage indicators in global monitoring efforts moving forward to and beyond 2015.

## The Process of Global Monitoring

Five iterative steps are required to generate and use coverage data for global monitoring (Figure 1).

First, global consensus indicators must be defined. The characteristics of a "good" indicator for global monitoring include: high validity (the extent to which the indicator is a true and accurate measure of the phenomenon under study); reliability (the

**Citation:** Requejo JH, Newby H, Bryce J (2013) Measuring Coverage in MNCH: Challenges and Opportunities in the Selection of Coverage Indicators for Global Monitoring. PLoS Med 10(5): e1001416. doi:10.1371/journal.pmed.1001416

**Academic Editor:** Nyovani Madise, Professor of Demography and Social Statistics, University of Southampton, United Kingdom

**Published** May 7, 2013



**Funding:** This work was conducted under the auspices of the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF, with financial support from The Bill & Melinda Gates Foundation through their grant to the US Fund for UNICEF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations: iERG**, independent Expert Review Group; RMNCH, reproductive, maternal, newborn, and child health; MDG, Millennium Development Goals.

* E-mail: jrequejo@jhsph.edu

**Provenance:** Submitted as part of a sponsored Collection; externally reviewed.

extent to which indicator measurements are consistent and dependable across countries and over time); the ability to detect change within a reasonable period and as a result of program implementation; and the ability to produce data that are easily interpreted and therefore useful in guiding program change [3,4].

Second, each indicator must be measured at the country level using standard methods that produce complete, comparable, high-quality data that are nationally representative. Producing such data requires specialized technical inputs in sampling and survey design, thorough training and supervision of those who collect the data, and close attention to quality in data entry, cleaning, weighting, and tabulation. Another paper in this Collection reviews the Demographic and Health Surveys and the Multiple Indicator Cluster Surveys, the two international household survey programs that produce the majority of coverage data used in global monitoring [5].

Third, country-level data must be compiled at the global level. This step involves checking data quality and the consistency of indicator measurement and often requires the recalculation of indicators from raw datasets. This process is labor intensive and contributes to the time lag between completion of surveys and the public availability of compiled databases. Various agencies (including UNICEF, WHO and UNFPA, and Save the Children) lead this step for different indicators.

The fourth step entails organizing the global databases and conducting another round of data quality checking. For some indicators (e.g., immunization coverage), this step includes a consultative process that involves the United Nations and independent technical groups, who develop estimates based on the combination of survey and program data. Country consultations are held on these adjusted indicators prior to their inclusion in publicly accessible databases.

Alteration of global databases to incorporate new interventions and improved measurement approaches requires adjustments in steps one through four. Decisions to introduce changes to global databases need to be made through a consultative process so that reporting requirements remain feasible and relevant for countries, trend analyses remain possible, and consensus is reached on indicator importance for informing programs and policies.

The fifth and final step is the completion of specific accountability analyses. This step requires agreement on which countries will be included, which indicators will be highlighted, what additional information is needed to interpret the results, and the identification of key messages for target audiences. These analyses must take into account any limitations of the data that are identified in earlier steps to ensure that correct interpretations are made for public health programming.

## Defining Sets of Indicators for Global Monitoring: Countdown and the Commission

Over the years, several initiatives, including Health for All by the Year 2000 [6], the World Summit for Children [7], and the Millennium Development Goals (MDGs) [8,9], have monitored sets of indicators to assess global progress in health. Critiques of these initiatives (e.g., [10,11]) stress that sets of global indicators represent more than the sum of the individual measurements, and often reflect broader aspirational concepts such as human development and human rights. The eight MDGs, for example, have been considered as a stimulus to poverty reduction strategies [12], official development assistance and political consensus [13], and increased monitoring of development projects [10]. Others argue that the definition of the MDGs is too narrow and leaves some important areas associated with development unrepresented, that the synergies across the MDGs are not sufficiently clear, and that they are particularly silent on equity [11].

The Countdown and Commission initiatives include sets of indicators that monitor MDG 4 (reduce child mortality) and MDG 5 (improve maternal health). Both initiatives address some of the limitations of the MDG framework, at least for the health of women and children, by embracing the holistic concept of the continuum of care and stressing the interrelatedness of the two MDGs. Both also complement the MDG indicators with recommendations that a more expansive set of RMNCH coverage indicators are analyzed nationally and by key equity considerations to promote accountability. Table 2 shows the coverage indicators in the two initiatives and compares them to the

**Table 1.** Countdown to 2015 for Maternal, Newborn and Child Survival and the Commission on Information and Accountability for Women's and Children's Health.

| *Countdown to 2015 for Maternal, Newborn and Child Survival* | *Commission on Information and Accountability for Women's and Children's Health* |
|---|---|
| **Aim** | **Aim** |
| Focuses on coverage and uses country-specific data to stimulate and support country progress towards the health related MDGs, particularly MDG 4 and MDG 5. | To develop a framework for global reporting, oversight and accountability on women's and children's health. |
| **Organizational structure** | **Organizational structure** |
| A global movement of academics, governments, representatives of multilateral and bilateral agencies, professional associations, non-governmental organizations and other members of civil society. It has a governance structure that manages the work and inputs from over 70 members. | Time-limited group developed following the launch of the Global Strategy for Women's and Children's Health in 2010. Progress in implementing its recommendations is overseen by an independent Expert Review Group (iERG). |
| **Countries** | **Countries** |
| The 75 countries where more than 95% of all maternal and child deaths occur. | The 75 countries where more than 95% of all maternal and child deaths occur. |
| **Products, reporting, and dissemination** | **Reporting and dissemination** |
| Periodic reports (in 2005, 2008, 2010 and 2012) and country profiles on key aspects of reproductive, maternal, newborn and child health. Advocacy materials and peer-reviewed articles. | In May 2011, the Commission launched its report, *Keeping Promises, Measuring Results*. The iERG will report annually until 2015 on progress in implementing its 10 recommendations. The first report was published in September, 2012. |

doi:10.1371/journal.pmed.1001416.t001

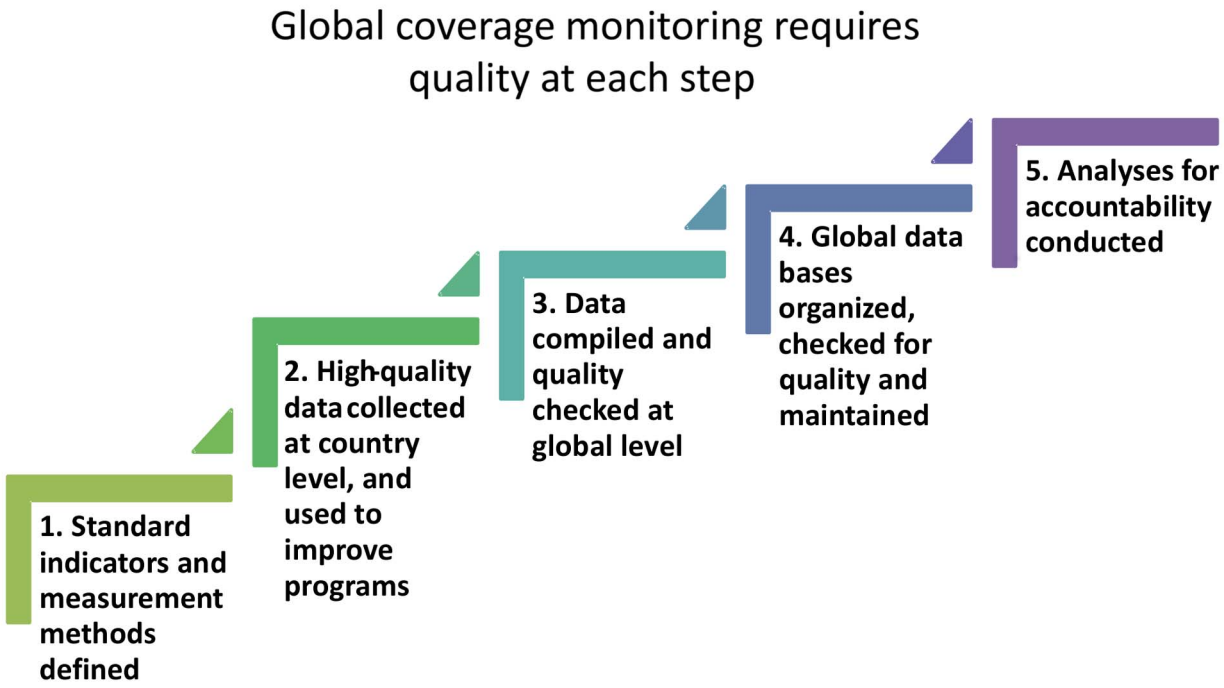


**Figure 1.** The five-step process for global monitoring of intervention coverage.
doi:10.1371/journal.pmed.1001416.g001

RMNCH coverage indicators included in MDG monitoring. In the next two subsections, we describe the technical and political considerations that drove the selection of the indicator sets tracked by Countdown and the Commission.

## Selection of Indicator Sets by Countdown

Since its inception in 2003 by the Bellagio Study Group on Child Survival [14], Countdown (Table 1) has produced periodic reports and country profiles on key aspects of RMNCH, and has been widely recognized for its role in promoting the use of coverage data to hold stakeholders to account for global and national action.

Countdown tracks progress in the 75 countries where more than 95% of all maternal and child deaths occur [15]. It synthesizes data on coverage of lifesaving interventions across the continuum of care, highlighting successes and missed opportunities. Countdown also tracks key determinants of coverage—equity patterns across population groups, health system factors, supportive policies, and available financial resources—and takes into consideration the role of broader contextual factors in driving coverage change [16].

Countdown is not involved in steps one through four in the global monitoring process (Figure 1). These steps are implemented by UNICEF, in consultation with various technical groups, resulting in a public access database (childinfo.org) that is updated annually and that contains the most recent estimates for coverage indicators for all countries. Countdown abstracts coverage data from this database and conducts a further check on internal and external validity by looking for out-of-range values or inconsistencies with other national survey data. Countdown then presents these data on country profiles and carries out secondary analyses to produce reports. This last stage includes a series of consultations with the Countdown membership to agree on what the coverage results mean for global public health, what key messages can be drawn to spur action, and how (and to whom) those messages can best be communicated.

Countdown has defined criteria to guide the selection of interventions for which it will track coverage. The most important criterion is the availability of internationally accepted evidence demonstrating intervention effectiveness in reducing maternal, newborn or child mortality and feasibility for delivery at scale in low- and middle-income countries. In addition, each intervention tracked by Countdown must be associated with a "good" coverage indicator as defined earlier. Countdown has reviewed its coverage indicators three times. The review process includes an assessment of experience in the last reporting cycle, consideration of new interventions and their associated coverage indicators against the selection criteria, and an open solicitation of proposals for changes.

Countdown reports on the number of countries with recent data for each indicator that it tracks. In its 2012 report, this ranged from only four countries with data for postnatal care for the baby to 73 countries with comparable data on measles [17]. The report also showed that only 29 Countdown countries had conducted a household survey during 2009–2011. These findings are presented to emphasize the need for more data collection efforts in the 75 countries as a prerequisite to improved accountability.

## Selection of Indicator Sets by the Commission

In 2010, a global strategy led by the United Nations called "Every Woman, Every Child" gave rise to a time-limited Commission on Information and Accountability for Women's and Children's Health. The Commission's mandate was to develop a framework for global reporting, oversight and accountability on women's and children's health [18].

The Commission used a two-step process to select a set of 11 core indicators, including three impact and eight coverage

**Table 2.** Coverage Indicators for Global Monitoring of RMNCH: Millennium Development Goal Framework, Countdown to 2015 for Maternal, Newborn and Child Survival, and the Commission for Information and Accountability for Women's and Children's Health, 2012.

| Coverage Indicator | | Millennium Development Goal[a] | Countdown to 2015 | Commission | Issues in Indicator Comparability across Initiatives |
|---|---|---|---|---|---|
| *Pre-pregnancy* | | | | | |
| 1) | Demand for family planning satisfied | | X | X | |
| 2) | Contraceptive prevalence rate | X | | | Countdown includes this indicator in a supplemental webannex to its report |
| 3) | Unmet need for family planning | X | | | Countdown includes this indicator in a supplemental webannex to its report |
| *Pregnancy* | | | | | |
| 4) | Antenatal care (at least one visit) with a skilled provider | X | X | | |
| 5) | Antenatal care (four or more visits) by any provider, skilled or unskilled | X | X | X | Commission indicator is defined as skilled provider only. Data are not currently available through international household survey programs for skilled provider. |
| 6) | Intermittent preventive treatment of malaria for pregnant women | | X | | |
| 7) | Neonatal tetanus protection | | X | | |
| 8) | Prevention of mother-to-child transmission of HIV | | X | X | The Commission combines the two HIV indicators; MDG 6B called for the achievement, by 2010, of universal access to treatment for HIV/AIDS for all those who need it. Target indicators for prevention of mother-to-child-transmission of HIV or antiretrovirals for pregnant women are not listed in the MDG framework. |
| 9) | Eligible HIV+ pregnant women receiving anti-retroviral therapy for their own health | | X | X | |
| *Birth* | | | | | |
| 10) | Skilled attendant at birth | X | X | X | |
| 11) | Cesarean section rate | | X | | |
| *Postnatal* | | | | | |
| 12) | Early initiation of breastfeeding | | X | | |
| 13) | Postnatal visit for mother | | X | X | The Commission reports on the two postnatal care indicators as a composite measure. Data may be available through international household survey programs on the composite measure in the current and future survey rounds. |
| 14) | Postnatal visit for baby | | X | X | |
| *Infancy* | | | | | |
| 15) | Exclusive breastfeeding | | X | X | |
| 16) | Introduction of solid, semi-solid, or soft foods | | X | | |
| 17) | Diphtheria-tetanus-pertussis (three doses) | | X | X | |
| 18) | Measles immunization | X | X | | |
| 19) | *Haemophilus influenzae* type b immunization (three doses) | | X | | |
| 20) | Vitamin A supplementation (two doses) | | X | | |

**Table 2.** Cont.

| Coverage Indicator | | Millennium Development Goal[a] | Countdown to 2015 | Commission | Issues in Indicator Comparability across Initiatives |
|---|---|---|---|---|---|
| *Childhood (under the age of 5 years)* | | | | | |
| 21) | Children sleeping under insecticide-treated nets | X | X | | |
| 22) | Children with fever receiving first line antimalarial treatment | X | X | | |
| 23) | Careseeking for pneumonia | | X | | |
| 24) | Children with suspected pneumonia receiving antibiotic treatment | | X | X | |
| 25) | Oral rehydration therapy with continued feeding | | X | | |
| 26) | Oral rehydration salts | | X | | |
| 27) | Improved drinking water sources | X | X | | |
| 28) | Improved sanitation facilities | X | X | | |

[a]includes only MDG target indicators related to RMNCH.

doi:10.1371/journal.pmed.1001416.t002

measures. Seven of the core coverage indicators are measured primarily through household surveys, and are described in Table S1 according to the criteria for what makes a ''good'' coverage indicator. National program records aggregated from facility records and modeling techniques are used to generate estimates for the remaining core coverage indicator—prevention of mother-to-child transmission of HIV and antiretroviral therapy for pregnant women. To select its core indicators, the Commission first convened a Working Group on Accountability for Results to prepare a background paper on recommendations for a set of indicators and measurement needs for women's and children's health. The background paper [19] specified that the core set of indicators should be limited in number to reduce the reporting burden on countries, should be reflective of the continuum of care, and should have strong political and public health significance across countries. The Working Group reviewed the MDG and Countdown indicators and recommended a set of coverage indicators from these. The Commission's final report – *Keeping Promises, Measuring Results* – incorporates these recommendations [18]. Other papers in this Collection examine the performance of several of these indicators with the aim of improving their measurement and interpretation [20–24]. Countdown, in collaboration with the Health Metrics Network, has also produced a report that describes the program relevance and measurement limitations of the coverage indicators selected by the Commission [25].

The Commission report also called for the creation of an independent Expert Review Group (iERG) to report annually until 2015 on progress in implementation of the Commission's 10 recommendations, which include improved measurement of the core coverage indicators in the 75 countries. Their first report indicated that, in 2012, only 11 of the 75 countries had recent data for all eight core coverage indicators, that there were gaps in coverage along the continuum of care, and that the poorest groups disproportionately experienced the lowest levels of coverage. The iERG recommended expanded commitment and capacity to evaluate RMNCH initiatives in order to help countries set priorities and allocate resources accordingly [26].

## Tensions and Compromises in the Selection of Global Monitoring Indicators

The selection of core coverage indicators by Countdown and by the Commission illustrates four tensions inherent in the process of selecting indicator sets for global coverage monitoring.

A first tension is between the desire to have ***comprehensive information*** about the policies and programs relevant to the topic of interest (i.e., poverty reduction, RMNCH), and the simultaneous need to ***keep the number of indicators small*** to minimize the reporting burden on countries and to avoid information overload. Resolution of this tension entails making hard choices about which indicators are left out. Indicators by definition are signals of the need to investigate a phenomenon more thoroughly. That this concept is poorly understood is reflected by the importance often mistakenly accorded to individual indicators in decision-making processes. Those engaged in global monitoring must continue to educate their target audiences about the appropriate use of indicators as signals that can trigger the need for further investigation and as signals that should be interpreted in the context of more comprehensive information.

Both Countdown and the Commission faced the challenge of selecting a core set of coverage indicators that represent the continuum of care as well as a balance between preventive and curative interventions. One way in which both initiatives achieved this was by selecting indicators of service contacts that reflect major dimensions of the continuum (e.g., antenatal care for the pregnancy period, skilled attendant at birth for labor and delivery, and postnatal care visits for the postnatal period). Although measurable through household surveys and readily understood, these contact measures do not necessarily reflect receipt of recommended interventions, which limits their usefulness for programmatic purposes.

A second strategy used by Countdown, but not by the Commission, to address this first tension was to expand the number of core coverage indicators, and to track coverage determinants (e.g., health systems factors, policies, and financial

data). Starting in 2008, Countdown extended its country profiles to two pages and created web-based supplemental tables to provide additional information. Countdown aims to be responsive to new evidence and regularly reviews its indicator set to ensure that it captures the best information available on RMNCH coverage. This iterative approach can, however, result in the selection of too many indicators and a consequent loss of focus and an inability to generate well-targeted key messages. Within the Commission, by contrast, the power to make decisions about the indicator set rested firmly with the seven Commissioners, informed by technical experts through the Working Group. This structure allowed the Commission to keep the total number of indicators small and its effort focused, but may have resulted in critical gaps. The Commission set, for example, includes only one indicator for prevention of childhood illnesses (diphtheria-tetanus-pertussis vaccination) and one indicator for case management of childhood illnesses (antibiotic treatment of childhood pneumonia). The set does not address interventions related to malaria or diarrhea, two major causes of child deaths in most of the 75 countries. The Commission acknowledges, therefore, that its indicator set is not sufficient for country-level monitoring.

A second tension is between ***what we know is important*** for improving public health and ***what can be measured reasonably well*** given available data sources and methods. A good example of compromise is the inclusion by both Countdown and the Commission of an indicator on coverage for antibiotic treatment of childhood pneumonia. Pneumonia, the number one killer of children worldwide [27] can be treated effectively and at relatively low cost with a course of antibiotics [28] but, as reported elsewhere in this Collection [20,21] and summarized in Table S1, the current global coverage indicator for treatment does not produce accurate results. Its measurement limitations suggest that additional related indicators may need to be collected until data collection methods are improved (i.e., care seeking for pneumonia). Countdown takes this approach and tracks both careseeking and treatment indicators. Another example is the Commission's antenatal care indicator ("four or more antenatal care visits from a skilled provider"). Although this indicator could potentially provide more programmatically useful data than the Countdown and MDG target indicator (which specifies any provider), in practice, information on the type of provider for each of four or more visits cannot be measured through household survey interviews because of recall issues. The inclusion of this indicator in the Commission set may spur efforts to collect these data, but comparable data from the 75 countries are currently not available. These examples demonstrate that when there is not a single indicator that is technically sound and useful for guiding programs for a given topic area, global monitoring initiatives can opt to collect a set of indicators related to the topic or provide adequate documentation on how to interpret a less-than-ideal indicator.

A third tension is between focusing on coverage for ***interventions that address the highest disease burden*** and ensuring that the indicator is ***relevant to as many countries as possible***. Two good examples of this tension concern malaria and HIV/AIDS. In two-thirds of the 75 countries covered by Countdown and the Commission, at least 75% of the population is at risk of *Plasmodium falciparum* malaria transmission. Interventions that effectively prevent deaths from malaria, such as the provision and use of insecticide-treated nets, antimalarials, and intermittent preventive treatment for pregnant women, have been scaled up rapidly in many of these countries [17]. For HIV/AIDS, 21 of the Countdown/Commission countries are priority countries for the elimination of mother-to-child transmission of HIV because of high levels of HIV seroprevalence in women. Prevention of mother-to-child-transmission of HIV with antiretroviral therapy drugs has been shown to be effective and is being rapidly scaled up in these countries [29]. Countdown reports on coverage of all these interventions; the Commission only reports on coverage of HIV/AIDS interventions (Table 2). These different choices show that the selection processes for global monitoring are not strictly based on the technical merits of individual indicators.

A fourth tension is between the need for ***timely data to guide decision making*** and the ***cost and resources required to conduct frequent surveys***. Frequent and high-quality coverage data are essential for program monitoring and require regular implementation of household surveys that meet at least minimum quality standards. In their reports, both Countdown and the iERG highlight data gaps for the indicators they track across and within the 75 countries they cover, and make strong arguments for frequent—even annual—collection of coverage data. Although a criterion for indicator inclusion in Countdown is data availability in most of the 75 countries to enable results to inform programs and policies, exceptions have been made for a few indicators because of their public health importance. Coverage for postnatal care for newborns, for example, has been included in Countdown reporting in the last two cycles [17,30] based on its importance for neonatal survival, even though fewer than 10 countries had data to report. It is now also one part of a composite indicator recommended by the Commission. This move towards reporting postnatal care for newborns has raised the visibility of postnatal care and has flagged the need to increase data collection efforts. Importantly, investments in data collection efforts to address such gaps must always be based on careful consideration of the timeframes required for detecting changes in specific indicators. Table S1 details variations in the ability of the core Commission indicators to detect change over time. Some indicators, such as exclusive breastfeeding and proportion of demand for family planning satisfied, are highly responsive to changes as they reflect current coverage. Others, such as skilled attendant at birth, are based on recall periods of two and five years (for the Multiple Indicator Cluster Surveys and Demographic and Health Surveys, respectively). Building capacity at the country level on understanding variations in the responsiveness of coverage indicators is essential for planning data collection efforts in view of resource constraints and for interpreting coverage levels and trends.

## Conclusions and Recommendations for Global Monitoring

Global monitoring efforts should produce timely results that can be used to support sound policy and programmatic decisions. Efforts are needed to generate better coverage data so that the indicators selected for global monitoring meet all technical requirements and to improve country capacity to measure, report on, and use them. In this article, we have reviewed the steps involved in global monitoring, the processes used by Countdown and the Commission to select a subset of coverage indicators for tracking coverage, and key tensions associated with selecting coverage indicators for global monitoring.

Several lessons gleaned from the Countdown and Commission indicator selection processes can be applied to the work now underway to define goals for the post-2015 era. First, the indicator selection process should be guided by the technical merits of individual indicators as well as by the underlying aspirational goal or broader agenda at the heart of the monitoring effort. Both the Countdown and Commission selection processes were driven in large part by the need to ensure that the indicator set adequately

## Key Points

- Tracking coverage of interventions proven to reduce maternal, newborn, and child mortality is central to global monitoring efforts.
- Effective global monitoring depends on a five-step process that ensures the generation of high-quality data.
- Sets of coverage indicators selected for global monitoring play a key role in driving policy and programmatic decisions at the global and national levels, but it is essential that other related information is considered when making these decisions.
- Correct interpretation of levels and trends of coverage indicators depends upon awareness of their strengths and limitations, and on the commitment and ability of stakeholders to use them for decision making.
- Efforts are needed to improve country capacity to measure and report on core coverage indicators through household surveys and routine reports; moreover, a rigorous and inclusive technical review process is needed to ensure that indicators for post-2015 global monitoring are chosen on the basis of evidence of effectiveness, feasibility of regular measurement, and programmatic relevance, and to maximize uptake of the indicators.

captures information across the continuum of care. Results can be used to identify major gaps and successes along the continuum, and to hold all partners to account for progress in achieving MDG4 and MDG5.

Second, any core set of coverage indicators needs to be interpreted within the context of information on inputs, processes and outputs related to program and policy implementation as well as broader social, economic, political, and environmental information that might affect coverage levels and trends. Countdown routinely tracks determinants of coverage in its analyses, and the Commission is clear on the need for additional information to supplement its coverage results, particularly at the national level.

Third, coverage indicators need to be selected through a rigorous and transparent process that involves consultation with a wide range of stakeholders. The process should include assessment of intervention effectiveness, data availability, and quality. The measurement limitations of the indicators also need to be identified so that they can be taken into consideration when interpreting results.

Our recommendation for the post-2015 agenda-setting process is that the example of the Commission should be followed, but that an additional step that involves critical review of the proposed indicators should be added to ensure that the indicators reflect the best balance between feasibility of measurement, data availability, and the broader agenda-setting functions needed from the set as a whole [13]. Such a review process should also stimulate further investment in a program of measurement research to improve the "value" of coverage indicators. For example, Table S1 shows that the development of methods for capturing information on interventions received during service contacts (e.g., antenatal care, skilled attendant at birth, and postnatal care) that is representative at the population level requires further work. Support for ongoing efforts to ensure consistency of measurement between the major international survey programs and over time is also needed for the production of comparable data.

Initiatives like Countdown with a more expansive indicator set should continue to serve as a resource for higher-level global monitoring efforts and to guide program development and implementation at national and sub-national levels. The bottom line, however, is that effective global monitoring depends on all five steps in the global monitoring process, and it is imperative that the whole process receives sufficient resources to allow it to respond to the future health needs of mothers and children.

## Supporting Information

**Table S1** Assessment of the seven core coverage Commission indicators measured primarily through household surveys.
(DOC)

## Acknowledgments

We would like to thank Dr. Fred Arnold, Measure Demographic and Health Surveys, for his careful review and insightful comments on various drafts of this manuscript.

## Author Contributions

Analyzed the data: JHR HN JB. Wrote the first draft of the manuscript: JHR JB. Contributed to the writing of the manuscript: JHR HN JB. ICMJE criteria for authorship read and met: JHR HN JB. Agree with manuscript results and conclusions: JHR HN JB.

## References

1. Bryce J, Victora CG, Boerma JT, Peters DH, Black RE (2011) Evaluating the scale-up to MDGs 4 and 5: a common framework. Int Health 3: 139–146.
2. Amouzou A, Habi O, Bensaïd K, and the Niger Countdown Case Study Working Group (2012) Reduction in child mortality in Niger: A Countdown to 2015 country case study. Lancet 380: 1169–1178.
3. Gakido E, Cowling K, Lozano R, Murray CJL (2010) Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: A systematic analysis. Lancet 380: 959–974.
4. Bryce J, Nguyen-Dinh P, Roungou JB, Naimoli J, Breman J (1994) Evaluation of Malaria Control Programs in Africa. Bull World Health Organ 72: 371–381.
5. Hancioglu A, Arnold F (2013) Measuring Coverage in MNCH: Tracking Progress in Health for Women and Children Using DHS and MICS Household Surveys. PLoS Med 10: e1001391. doi:10.1371/journal.pmed.1001391
6. World Health Organization (1982) Global Strategy for Health for All by the Year 2000. Geneva: WHO. Available: http://whqlibdoc.who.int/publications/9241800038.pdf. Accessed 24 January 2013.
7. World Summit for Children, 1990. Available: http://www.unicef.org/wsc/. Accessed 12 November 2012.
8. Millennium Development Goals (2011) Accountability commission for health of women and children. Available: http://www.who.int/topics/millennium_development_goals/accountability_commission/en/. Accessed 23 January 2013.
9. Millennium Development Goals. Official list of MDG indicators. Available: http://mdgs.un.org/unsd/mdg/Host.aspx?Content = Indicators/OfficialList.htm. Accessed 23 January 2012.
10. Vandemoortele J (2009) The MDG conundrum: Meeting the targets without missing the point. Development Policy Review 27 (4): 355–371.
11. Waage J, Banerji R, Campbell O, Chirwa E, Collendez G, et al. (2010) The Millennium Development Goals: a cross-sectoral analysis and principles for goal setting after 2015. Lancet DOI:10.1016/S0140-6736(10)61196-8.
12. Fukuda-Parr S, Hulme D (2009) International Norm Dynamics and 'The End of Poverty: Understanding the Millennium Development Goals (MDGs)', WPI Working Paper 96, Manchester: Manchester University. Available: http://www.eadi.org/fileadmin/MDG_2015_Publications/fukuda-parr_and_hulme_2009_international_norm_dynamics.pdf. Accessed 24 January 2013.
13. Sumner A, Tiwari M (2010) Global Poverty Reduction to 2015 and Beyond: What has been the Impact of the MDGs and What are the Options for a Post-2015 Global Framework? IDD Working Paper 348. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.2040-0209.2010.00348_2.x/pdf. Accessed 9 November 2012.
14. Bellagio Study Group on Child Survival (2003) Knowledge into action for child survival. Lancet 362: 323–27.
15. Countdown to 2015 writing group (2012) Building a future for women and children. Lancet 379: 2121–2122.

16. Bhutta Z, Chopra M, Axelson H, Berman P, Boerma T, et al (2010) Countdown to 2015 Decade report (2000–2010): taking stock of maternal, newborn, and child survival. Lancet 375: 2032–2044.
17. Requejo J, Bryce J, Victora C and the Countdown writing team (2012) Building a future for women and children: Countdown 2012 report. Geneva: WHO.
18. Commission on Information and Accountability for Women's and Children's Health (2011) Keeping promises, measuring results. Available: http://www.everywomaneverychild.org/images/content/files/accountability_commission/final_report/Final_EN_Web.pdf. Accessed 12 November 2012.
19. Working Group on Accountability for Results. Available: http://www.who.int/topics/millennium_development_goals/accountability_commission/Working_Group_on_Results_Final_Paper.pdf. Accessed 23 January 2013.
20. Campbell H, el Arifeen S, Hazir H, O'Kelly J, Bryce J, et al. (2013) Measuring coverage in MNCH: Challenges in monitoring the proportion of young children with pneumonia who receive antibiotic treatment. PLoS Med 10: e1001421. doi:10.1371/journal.pmed.1001421
21. Hazir T, Bequem K, el Arifeen S, Khan AM, Huque MH, et al. (2013) Measuring coverage in MNCH: A prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia. PLoS Med 10: e1001422. doi:10.1371/journal.pmed.1001422
22. Cutts FT, Izurieta H, Rhoda DA (2013) Measuring coverage in MNCH: Design, implementation and interpretation challenges associated with tracking vaccination coverage using household surveys. PLoS Med 10: e1001404. doi:10.1371/journal.pmed.1001404
23. Moran A, Kerber K, Sitrin D, Guenther T, Morrissey CS, et al. (2013) Measuring coverage in MNCH: Indicators for global tracking of newborn care. PLoS Med 10: e1001415. doi:10.1371/journal.pmed.1001415
24. Stanton CK, Rawlins B, Drake M, Anjos M, Cantor D, et al. (2013) Measuring coverage in MNCH: Testing the validity of women's self-report of key maternal and newborn health Interventions during the peripartum period in Mozambique. PLoS ONE 8: e60694. doi:10.1371/journal.pone.0060694
25. Countdown to 2015 and Health Metrics Network (2011) Monitoring maternal, newborn and child health: Understanding key progress indicators. Geneva: WHO. Available: http://www.who.int/healthmetrics/news/monitoring_maternal_newborn_child_health.pdf. Accessed 23 January 2013.
26. Independent Expert Review Group (2012) Every woman, every child: from commitments to action. Geneva: WHO. Available: http://www.who.int/woman_child_accountability/ierg/reports/2012/IERG_report_low_resolution.pdf. Accessed 12 November 2012.
27. Liu L, Johnson H, Cousens S, Perin J Scott S, et al (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 379 (9832): 2151–2161.
28. United Nations Children's Fund (2012) Pneumonia and diarrhea: Tackling deadliest diseases for the world's poorest children. New York: UNICEF. Available: http://www.unicef.org/media/files/UNICEF_P_D_complete_0604.pdf. Accessed 23 January 2013.
29. Joint United Nations Programme on HIV/AIDS (2011) Countdown to Zero: Global plan towards the elimination of new HIV infections among children by 2015 and keeping their mothers alive 2011–2015. UNAIDS. Available: http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/20110609_JC2137_Global-Plan-Elimination-HIV-Children_en.pdf. Accessed 23 January 2013.
30. Requejo J, Bryce J, and the Countdown writing team (2010) Countdown to 2015 Decade Report (2000–2010): Taking stock of maternal, newborn, and child survival. Geneva: WHO. Available: http://www.countdown2015mnch.org/documents/2010Report/2010_ReportAndProfiles.pdf. Accessed 23 January 2013.


PLOS is a nonprofit organization that accelerates progress in science and medicine by leading a transformation in research communication

U.S. HEADQUARTERS
1160 Battery Street
Koshland Building East, Ste. 100
San Francisco, CA 94111 US
+ 1 415.624.1200 PHONE
+ 1 415.546.4090 FAX

EUROPEAN EDITORIAL OFFICE
Carlyle House
Carlyle Road
Cambridge CB4 3DN UK
+ 44 0.1223.442.810 PHONE
+ 44 0.1223.442.833 FAX

Published May 2013

info@plos.org
plos.org