Methods and issues in household surveys of vaccination coverage

Report to the Bill and Melinda Gates Foundation

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

November 2012

1. Independent consultant, La Londe les Maures, France
2. Consultant, North Bethesda, MD, USA*
3. Battelle Memorial Institute, Columbus, OH, USA

*Disclaimer: Dr. Izurieta currently works for the USA Food and Drug Administration, although he performed this work as an independent consultant. Therefore, the content of this paper reflects the views of the authors only and not the Food and Drug Administration.

Table of Contents

Summary and Recommendations

The prevalence of vaccination, or vaccination coverage, is an important indicator of public health and of the effectiveness of a country's public health infrastructure. Conceptually, vaccination coverage is a straightforward proportion: the number of persons in the target population who have been vaccinated in a given time period divided by the total size of that population, usually expressed as a percentage.

In low- and middle-income countries two methods are commonly employed to estimate coverage: indirect inference from administrative reports and direct inference using survey sampling. The quality of reporting and of estimation of the target population varies widely, within and between countries, making those estimates difficult to compare or sometimes even believe. Well-designed and executed surveys offer the potential for more accurate and comparable results.

There are two large well-designed survey programs that report on vaccination coverage in low and middle-income countries: USAID's Measure Demographic and Health Survey (DHS) and UNICEF's Multiple Indicator Cluster Survey (MICS). Both are multi-purpose country-wide household surveys of several thousand households conducted every few years with vaccination coverage forming only a small part of the survey agenda. They use probability sampling, i.e., the units are selected with known and nonzero probabilities. Two other methods are often used in smaller surveys that may be focused solely on vaccination coverage: the Expanded Program on Immunization's (EPI) cluster survey and Lot Quality Assurance Sampling (LQAS). These may yield results that are more geographically specific, sometimes down to the level of the health district or catchment area. They are often focused on near-term decision-making and less concerned with being comparable over time and location.

Some surveys collect blood specimens and assess immunity directly using serological assays. Exactly what can be measured and what can be inferred varies from antigen to antigen. Sero-surveys have an important role to play in assessing public health risk, but they do not address the full range of questions about immunization program effectiveness that can be addressed with good survey data on who received which vaccines, and when. Sero-surveys and vaccination-coverage surveys are likely to complement each other for the foreseeable future.

Surveys measuring vaccination coverage and/or serological evidence of immunity are susceptible to a set of challenges that include selection bias and information bias, which cannot be solved by increasing the sample size, as well as the precision of the coverage estimate (i.e. the width of the confidence intervals), which is determined by the survey sample size and sampling method. Common sources of errors are summarized in the table.

Selection bias can occur when the list of eligible respondents (the survey frame) excludes subpopulations and/or when field procedures use non-probability sampling methods or substitute a sampled household with one which is easier to reach. The populations likely to be missed in a vaccination coverage survey are also likely to be missed by vaccination teams, so selection bias most often inflates coverage estimates.

Information bias occurs when a child's vaccination status is misclassified due to mistakes on the vaccination record, transcription of data from the record, the way the questions were presented or the guardian's recall for children without a written record. There has been substantial reliance on the guardian's recall in recent surveys. Information bias may become more likely in the future as immunization schedules become more complex and variable.

Selection bias and information bias can affect any type of survey irrespective of sampling method, and potentially could be greater for larger multi-purpose surveys with long questionnaires, where field workers may be more tempted to cut corners, unless extremely good supervision and quality control are in place. Survey methods review should include questions such as: what proportion of the results was obtained from vaccination records as opposed to caregiver recall? How were missing data handled in the analysis? Were homes of absent responders revisited or replaced? How much opportunity did data collection teams have to select which homes and individuals entered the sample? The answers to questions like these help classify likely biases associated with a survey and help the person interpreting the results gauge whether any strong biases are likely to lead to the coverage being over- or under-estimated.

It is common to focus on the point estimate from the survey – the proportion of children who are estimated to be covered, but it is also important a) to consider uncertainty due to sampling design, which is expressed using a confidence interval, and b) to consider potential biases, which are not quantified, but must be inferred from what one reads about the survey protocol and its implementation.

Table: **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 minimise error
a. 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
b. Systematic error		
Selection bias - sampling frame	Depends on size of excluded population and difference in uptake of vaccination 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	Pre-select 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 under-estimate or over-estimate 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 Publicise 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 TT in adult women, verbal history usually under-estimates % 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, close supervision For TT, ask careful questions about <u>all</u> TT doses received in previous and current pregnancies and in campaigns (but this still does not account for DTP received in infancy). Sero-surveys play a useful role to measure 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 while enumerators can revisit household if necessary to correct data
Missing data	If non-random, biases results, often upwards	Conduct high-quality planning, training and supervision to reduce missing data Include appropriate statistical adjustment for missing data

Notes: HBR = Home-based record; TT = tetanus toxoid

When deciding which survey methodology is the most appropriate, decision makers should consider not only the specific information needed but also the speed with which the information is required. For instance, if the priority is to evaluate the success or failure by local area of a Measles/Rubella vaccination campaign while field operations are ongoing in that area, the survey design needs to take into account the potential tradeoffs between timeliness, precision, and ability to reduce bias.

In conclusion, while it is critical to recognize the limitations of coverage as the sole indicator of program success, there is great potential to use even imperfect coverage data together with other program indicators to improve program management at local level. For global policy-making, it is essential to assess the potential for bias in coverage estimates from whatever source and to pay attention to the confidence intervals, not only the point estimates, when using results. As a measure of population protection, coverage is limited by assumptions about vaccine effectiveness and thus is helpful but not sufficient, and additional information e.g., from effective vaccine management assessments, surveillance, outbreak investigations and special studies is needed to obtain a fuller picture of program impact. For some vaccines, biomarkers are already available which can be used to measure the prevalence of immunity in the target population while for others, better biomarkers are needed. There is currently tension between performance-based financing (PBF) systems which reward high coverage and efforts to improve the quality of coverage measurement. It is time to reward actions which improve the quality of data and not only a country's apparent coverage achievements.

Recommendations

1. Vaccination coverage should not be used alone as an indicator of program success as it does not necessarily equate to population protection. Action is needed by partners and country immunization programs to improve the interpretation and use of coverage data together with data on other indicators to improve program performance, e.g.:
    - Coverage of each vaccine and dropout rates – overall and among key subgroups
    - Timeliness of vaccination (receipt of vaccines within e.g. one month of recommended age)
    - Quality of vaccination (controlled temperature chain, safe injections, waste disposal)
    - Missed opportunities and strategies to reduce them
    - Disease surveillance & outbreak investigation (cases by age, vaccination status, time and place, surveillance quality indicators)

2. To reduce bias in coverage measurement by any method, primary recording of vaccination data must be improved. In the long term this will involve digital recording and data transmission. In the short-medium term in low income countries, paper-based recording must be improved. Record design will need to evolve rapidly to accommodate new and underutilised vaccine introduction and to allow recording of doses administered through Periodic Intensifications of Routine Immunization.

3. When surveys are conducted, efforts to ensure high-quality data must be made, including:
    - Using a probability sampling design (this requires a step of household listing prior to taking the sample)
    - Assessing the potential importance of populations who may have been missed in the sampling frame, e.g., those living in insecure areas; those living in temporary shelters; unofficial migrants. Surveys which aim to assess equity in coverage will generally need to make special efforts to reach such sub-populations.
    - Surveying children less than one year of age as well as those aged 12-23 months. Records are more likely to be available for younger children allowing assessment of indicators for which complete information on dates of birth and of vaccination are needed, e.g. median age at starting the vaccine series, proportion receiving a dose within one month of the recommended age. Furthermore, action can be taken (e.g., referral to the nearest health facility or outreach site) to increase coverage among those who are behind on vaccinations.

- Conducting revisits to households where a responsible adult is not available or the home-based record is not available but known to exist (e.g. is locked away temporarily)
- Seeking health-facility based records for children whose home-based record is unavailable
- Allowing adequate time for planning the survey, developing standard operating procedures and training interviewers
- Conducting close supervision of field activities
- Using high-quality data management and verification procedures

4. Assessment of the potential magnitude of bias in surveys is needed before results are interpreted. Quality control criteria leading to a decision to use or not results from surveys at national or global levels should be developed, endorsed by partners and ideally incorporated into the WUENIC process.

5. Partners should be encouraged to consider uncertainty in the coverage estimates when basing decisions such as PBF on coverage.

6. At sub-national level, ad-hoc sample surveys can be conducted but a more interesting alternative could be to conduct a rolling census of districts or sub-district, beginning with those considered to have high risk of low coverage. The census would obtain data on both the denominator and the vaccination status of children aged less than 2 years, and children behind on vaccination would be immediately referred for vaccination. Cohort analyses of vaccination coverage can then be conducted. This would be a prelude to development of digital registries, where such a rolling census would be one means of ensuring that registries capture all children in the area. We recommend evaluation of this approach in urban and rural areas of at least 2 countries.

7. Actions which will transform coverage measurement in the longer term include:
   - Roll-out digital recording of vaccination on clinic-based and national registries (i.e. building on work by Optimize and others; enrolment can be via birth registration and at first vaccination).
   - Work with vaccine manufacturers to incorporate bar-codes as well as heat- and freeze- indicators on vaccine vials to facilitate digital recording.
   - Evaluate the potential use of improved biomarkers to assess program effectiveness.

1. Introduction

The percentage of a population which has been vaccinated, called *vaccination coverage,* is an imperfect but helpful measure of public health and program effectiveness. Most importantly, it is a proxy for the prevalence of immunity and protection from risk from vaccine-preventable diseases (when combined with assumptions on vaccine effectiveness). Furthermore, it is a helpful measure to mark progress toward goals and targets, identify populations that are un- or under-vaccinated, infer that vaccines were used responsibly, and to measure the effectiveness of vaccination programs or even individual campaigns or vaccination teams. (Hadler, Dietz et al. 2008) Vaccination coverage is a tracer condition for results-based financing (RBF) (Brenzel et al 2009), an indicator of eligibility for Millennium Challenge Account Assistance (Millenium Challenge Corporation 2011), and a criterion for support for introduction of new vaccines – from 2012 onwards, countries who wish to receive GAVI support with the introduction of new vaccines need to have DTP3 coverage above 70%. (GAVI 2012a) Making funding decisions contingent on coverage potentially incentivises inflation of coverage figures, and there is wide recognition of the need to improve the reliability of coverage data. (Murray, Shengelia et al. 2003; Lim, Stein et al. 2008)

In this report we review coverage estimation methods, giving most attention to survey-based methods used in low and middle-income countries, which are often held as the "gold standard" source of coverage data , but which are subject to several sources of error that are important to be aware of and minimise. (Murray, Shengelia et al. 2003) We focus on evaluation of coverage of the routine immunization program but discuss some trade-offs which may be needed for evaluation of campaigns. We highlight potential pitfalls in surveys and propose strategies to improve coverage measurement, to inform public health practitioners and researchers who design and implement surveys as well as Ministry of Health officials and donors who interpret and use data from surveys.

1.1. What is vaccination coverage?

Conceptually, vaccination coverage is a straightforward proportion: the number of persons in the target population who have been vaccinated in a given time period divided by the total size of that population, usually expressed as a percentage. In low and middle-income countries, it is most often measured among children aged 12-23 months, the youngest annual cohort old enough to have completed the primary immunization series. Coverage is measured for each dose of each vaccine in the national immunization schedule, as well as the proportion of the population that has received all the indicated vaccines ("fully vaccinated child").

Coverage (%) = # doses administered*100
individuals

- denominator: target population during the period of interest
- numerator: number of doses of each antigen-dose combination administered to the target population during the period of interest

1.2. Desirable properties of a coverage estimate

A helpful estimate should be well-defined, accurate, precise, geographically specific, current, and comparable to other estimates over time or location. Survey designers need to know the end-users' perspectives, for instance, about *how precise* is precise enough, or *how geographically-specific* is specific enough. For example, program managers at local level usually require data quickly to help guide their actions, whereas at global level, more precise estimates are desired but longer delays are accepted before results are available to monitor overall progress towards global goals.

1.3. How is coverage measured?

Vaccination coverage is estimated either using available administrative data (routine reports and registries) or by conducting a survey. In some cases, administrative data are considered alongside survey data when these are available, to try to gain a comprehensive perspective, e.g., as done in the WHO-UNICEF coverage estimates (WUENIC) (Burton, Monasch et al. 2009).

1.3.1. Registries and routine reports

The reference standard for continuous administrative monitoring of vaccination coverage is based on a birth- and/or vaccination register aiming to document vaccination of each individual in each entire birth cohort. Denominators may derive from the same registry or from a separate vital statistics system. When well implemented they can provide data for coverage measurement and program management activities such as vaccine supply monitoring and requisitions, listing children due vaccines and sending reminders. Challenges include accounting for migration within and between countries, ensuring complete birth registration and reporting from public and private vaccination providers, avoiding record duplication and ensuring continuity after organizational changes e.g. changes in district boundaries. Registries have been used mostly in industrialized countries including some middle-income countries. (Ronveaux, Arrieta et al. 2009; Hull, Dey et al. 2011) Pilot studies of the use of registries are ongoing in

low and middle-income countries including Albania, Guatemala, India and Vietnam (PATH 2012) and in the long-term they are likely to be the best way to obtain continuous data for coverage measurement and program management.

More commonly, a rough estimate of vaccination 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 in the numerator and target population estimates in the denominator. (Hadler, Dietz et al. 2008) In many countries, however, the quality of reporting is low, numerators may be either inflated e.g., due to including doses outside the recommended age range, or too low e.g., if private practitioners do not report, and denominators are grossly inaccurate (Lim, Stein et al. 2008; Bosch-Capblanch, Ronveaux et al. 2009). Hence, wherever possible other data sources such as surveys are considered in the WHO-UNICEF coverage estimates (WUENIC) process . (Burton, Monasch et al. 2009)

1.3.2. Surveys

In this section we briefly describe four types of surveys that are commonly employed to estimate vaccination coverage. Two (DHS & MICS) are very large, multi-purpose surveys which are admirably comparable across time and location, and the other two (EPI cluster, and Lot Quality Assurance Sampling (LQAS)) are employed usually for more focused, timely, tactical purposes at national and sub-national levels. Features of these surveys are compared in Table 1.

The MEASURE **Demographic and Health Survey (DHS)** is a household survey conducted in developing countries with funding from the United States Agency for International Development (USAID). It uses a standardized multi-stage cluster sampling design and standardized questionnaires to obtain data on a wide range of topics related to population health and nutrition. (Rutstein and Rojas 2006) Surveys are repeated in DHS countries about every five years with interim measures of important indicators. Data and reports from DHS surveys are freely available, with preliminary reports available about 3 months after survey completion and final reports after a longer time delay. (Measure DHS 2012)

The United Nations Children's Fund (UNICEF) **Multiple Indicator Cluster Survey (MICS)** is another household survey. MICS have been conducted multiple times in over 60 countries since 1995. Its design is very similar to DHS although in the past it allowed more flexibility in sampling method. MICS surveys, datasets, and reports are freely available. (Ahmed, Ali et al. 2009; UNICEF 2012) Currently, the fourth round of MICS is in progress, with results from all 65 surveys expected before the end of 2012.

Both DHS and MICS are probability sample surveys, that is one in which households have known and nonzero probabilities of being selected. They aim to be representative of the entire population of a country. They use a "standard segment" sampling design, as follows:

- Three-stage sampling; can have implicit stratification (which is achieved by ordering the list of census enumeration areas (EAs) according to desired strata, e.g.,first list urban then rural areas)
- Selection of enumeration areas by probability proportional to estimated size (PPES) sampling
- Mapping and segmentation in larger EAs with more than one standard segment (a "standard segment" is a population of about 500 persons)
- Selection of one segment at random (in small EAs, the segment is the EA)
- Listing of households in sample segments
- Systematic random selection of sample households in segments, usually with a country-wide total of at least 4,000 households.

The World Health Organization's **Expanded Programme on Immunization (EPI) Cluster Survey** was developed in the late 1970's as a practical tool to provide a quick estimate of coverage to within +/- 10 percentage points of the point estimate, in 12-23 month old children and their mothers with a modified cluster design of 30 clusters and a non-probability sample of seven children in each cluster using the "random walk" method. (Henderson and Sundaresan 1982) This survey is 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. Several adaptations have been evaluated to incorporate probability sampling at the final stage (Turner, Magnani et al. 1996; Milligan, Alpha et al. 2004; Myatt, Feleke et al. 2005; Grais, Rose et al. 2007; Luman, Worku et al. 2007). The recent guidelines on the conduct of hepatitis B cluster surveys, a companion manual to the original WHO EPI survey guidelines, emphasize that probability sampling is needed for scientifically robust estimates of coverage. (World Health Organization 2011a) The main design difference between the EPI survey (if probability sampling is used) and the DHS or MICS is that EPI surveys focus specifically on vaccination data while DHS and MICS have long questionnaires covering a wide range of population and health topics and include a much larger total sample size. Field implementation of EPI surveys is variable and often done without external technical assistance (TA), while the DHS and MICS are highly standardized and have substantial TA and quality control.

Lot Quality Assurance Sampling (LQAS) is a small sample technique that was developed for the manufacturing industry in the first half of the twentieth century and applied to public health in the late

1980s and is used increasingly in immunization. (Dodge and Romig 1959; Lemeshow and Taber 1991; Robertson, Anker et al. 1997; Robertson and Valadez 2006) It uses a stratified sampling approach, considering coverage in each so-called *lot,* which might be a district, health unit or catchment area. It does not estimate percent coverage for each lot, but rather classifies each lot as having either *adequate* or *inadequate* coverage. LQAS surveys employ a wide range of survey sizes. Some plans require only a day, or less, to survey a catchment area whereas others sample many clusters per lot, and require several days. In many cases data can be combined across multiple lots (strata) to estimate coverage for a region quite precisely. A clustered version, where LQAS sampling is "nested" within a cluster survey, is used to evaluate whether neonatal tetanus is likely to have been eliminated (defined as incidence of <1/1000 live births) in a country (Cotter, Bremer et al. 2003), to evaluate coverage of yellow fever vaccine (Pezzoli, Pineda et al. 2009), to evaluate coverage of a meningococcal vaccine campaign (Kim, Pezzoli et al. 2012) and to monitor polio vaccination coverage after supplementary immunization activities. (World Health Organization 2012a)

Table 1. Characteristics of Common Surveys used to measure vaccination

Survey characteristic	DHS	MICS	EPI	LQAS
Primary objectives	Collect information on a wide range of population, health & nutrition topics, plus additional optional modules	Collect information on population health,child protection and child development.	Estimate vaccination coverage	Classify lots (catchment areas) into two groups: those with so-called *adequate* and 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 usually villages or urban neighborhoods, selected using PPES*	Classic method uses simple random sampling (SRS) within a lot; when lots are large, cluster sampling is sometimes employed
Household selection	HH selected randomly based on a complete household listing and mapping in the sample clusters.	Current practice is random selection of HH based on a complete listing and mapping of enumeration areas	Varies; usually non-probability; select first HH randomly; select neighboring HH until data are obtained from enough HH to enroll 7 children	When cluster sampling is used, select first HH randomly and then move in a consistent direction sampling every *k*th HH
Total sample size	Based on desired precision for key indicators at the regional level. Number of children age 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 >2000 women and several hundred children aged 12-23 months.	Usually 30 clusters of 7 children aged 12-23 months each = 210 children; sized to yield estimate of ± 10% assuming design effect of 2	Varies greatly; 19 respondents per lot is a common size with SRS; 50 or 60 is common when using cluster sampling

Respondents	All men and women 15-49; vaccination data on children < 5 years if biological mother is interviewed and WCBA	All women 15-49; vaccination data on children < 5 years if primary caretaker is interviewed and WCBA	Mothers or primary caretaker of children 12-23 months	Varies: 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 Under-5: 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 & 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 field and analyze

Notes: MCH = Maternal & Child Health; PPES = Probability Proportional to Estimated Population Size; HH = household; WCBA = women of childbearing age

2. Survey design and implementation

When surveys use standardized techniques and best practices they can yield estimates that are appropriate for comparison over time or between locations. Vaccination surveys can, however, be subject to the threats to validity that are common to survey designs, especially selection bias and information bias. Sample size is another key component of survey design as it affects the precision of the estimates, as discussed in section 3.

2.1. Selection bias

Selection bias may occur due to use of an outdated or non-representative sampling frame (e.g., failure to capture unofficial residences, homeless families and temporary visitors; inaccurate census possibly influenced by political issues (Borgdorff and Walker 1988); changes in boundaries or fertility patterns and permanent or temporary migration that make extrapolations from the latest census unreliable), use of non-probability sampling, or poor field worker practices such as substituting a selected household with one of easier access. In general, the more flexibility the field worker has regarding household selection, the greater the chance of selection bias The "random walk" method, where a starting household is chosen by a simple method and then adjacent households are visited until 7 children aged 12-23 months have been enrolled, has intrinsic geographic bias, allows field workers to select households rather than this being part of the initial sampling process, does not document reasons for non-participation, and cannot adjust for biases due to use of out of date estimates of size for selection of clusters using probability proportional to estimated size (PPES). 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 not found by interviewers because they work in the fields all day also lack time to attend vaccination clinics), replacement is likely to result in bias, probably upwards.

Surveys of the vaccination status of living individuals are inherently biased since death is more likely in unvaccinated than in vaccinated children. In high infant mortality rate settings, this bias may be substantial.

2.2. Information bias

Information bias may derive from the respondent or the field worker. In community-based surveys in low- and middle-income countries, vaccination status is usually determined by transcription of

vaccination dates from the home-based vaccination record (HBVR) to a paper or digital questionnaire. When the HBVR is not available (it may be lost, locked away, or difficult to find and interviewers may not allow time for its retrieval), questions are asked to the parent or guardian of the child to obtain a verbal history of vaccination. Occasionally, health-facility-based records may be sought.

2.2.1. Accuracy and availability of data on the home-based vaccination record

Although HBVR-based information is usually considered to be reliable, mistakes can be made during primary recording of data on vaccinations received or date of birth when the child attended for vaccination, during later transcription of data when old, lost or damaged vaccination records are replaced, or during transcription of data onto the questionnaire. Since primary recording is often done in batches at a site in the health centre which is separate from the vaccination station, vaccinations may be recorded on a card without the vaccine being given (e.g., because the vaccine supply at the clinic ran out but the clerk had already recorded the vaccination). (Cutts, Smith et al. 1990) In some countries, health workers write in pencil the date when the next vaccine is due and this can be confused with actual vaccinations received if interviewers are not alert. When new vaccines are introduced in the national schedule, old HBVRs may continue in circulation requiring health workers to improvise in their recording. (See Figure 1.) There is further confusion as to whether or not to record vaccines administered during campaigns such as “vaccination weeks” on the HBVR. (World Health Organization 2012d) Mistakes in recording or transcribing dates of birth and of each vaccine and dose on the home-based record and/or on the survey questionnaire can affect later calculations of valid doses received and hence coverage.

For vaccines administered to adults, such as tetanus toxoid (TT) given to adult women to prevent neonatal tetanus, home-based records are rarely available. Women who are currently pregnant may have a record of TT received during the current pregnancy but rarely have records of doses received previously. Since a lifetime total of 5 or 6 (depending on schedule) doses of TT-containing vaccines is considered to protect women through their childbearing years, (World Health Organization 2006)only recording TT received in the most recent pregnancy generally underestimates the proportion of women protected.

Fig.1. Several instances of improvisation on a vaccination card. (Courtesy of Carolina Danovaro, PAHO.)

2.2.2. Accuracy and availability of data on health facility (HF)-based records and their consultation in surveys

Health facility-based records include registers, tally sheets, a clinic-based copy of the vaccination record and others. In principle, HF-based records could be sought for children whose HBVR is not available. While common in developed countries, this is rarely done in surveys in low income countries. Even if HF records were consulted, errors are common for multiple reasons (Ronveaux, Rickert et al. 2005; Bosch-Capblanch, Ronveaux et al. 2009)), e.g.,

- Poor form design (e.g., lack of space; outdated records meaning improvisation is needed for new vaccines);
- mistakes in recording data (e.g., wrong vaccine dose, wrong date, inclusion of children outside the target age group etc)
- transcription errors when copying data from temporary records used at outreach sites into the permanent register
- deliberate falsification of records, (e.g., due to real or perceived pressure to meet targets)
- Poor storage leading to lost or damaged records and registers.

2.2.3. Reliability of the verbal history of vaccination

The reliability of a history of vaccination elicited from the mother may vary with the mother's education, health worker's diligence in explaining what vaccinations have been given at each visit, the interviewer's skills, carefulness, neutral demeanor and use of appropriate language, the recall period, the complexity of the vaccination schedule, and the length of the questionnaire and resulting interview fatigue. (Gareaballah and Loevinsohn 1989; Valadez and Weld 1992; Langsten and Hill 1998) The respondent may know that a certain vaccine (e.g.,DTP) has been received but be unable to recall how many doses the child received. The time available to take a careful history may be much shorter in a long survey such as the DHS which may have questionnaires of over 100 pages than an EPI survey with a one or two page questionnaire.

Problems both for recording vaccinations and for eliciting a verbal history of vaccination from caregivers are likely to become greater as the number of vaccines in the National Immunization Program (NIP) increases and schedules become more complex. When the EPI coverage survey was introduced in the 1980s, the infant EPI schedule comprised 5 visits which lent themselves to straightforward questions to the mother (Table 2).

Table 2. Illustrative questions used in the past to elicit a verbal history of vaccination for vaccines in 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 Diphtheria-tetanus-pertussis (DTP) (subcutaneous or intramuscular injection, usually in the thigh) and oral polio vaccine (OPV) (oral)	Did the child receive an injection in the leg (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	
14 weeks	DTP, OPV 3	
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?

Current schedules are much more complex and vary over time and between countries, making standardization of questions much more difficult. For example, DTP-containing vaccines (DTPCV) are generally administered following a 6, 10, 14 week schedule in low income countries of Africa and Asia (though some begin the schedule at age 8 weeks), and single-antigen measles vaccine at age 9 months. In most middle and high-income countries, corresponding ages are 2, 4, 6 months for the DTP-containing vaccines and 12 months for 1st dose measles-mumps-rubella vaccine. (World Health Organization 2012c) The scheduled number of doses for rotavirus vaccine varies according to the manufacturer, and pneumococcal conjugate vaccine schedules vary from administration simultaneously with DTPCV to having longer intervals between doses, with the 3rd dose administered simultaneously with measles vaccine. It is unlikely that mothers will know which of the DTP-containing vaccines (DTP, pentavalent, hexavalent, etc.) the child received, or be able to distinguish between injections of these, pneumococcal or meningococcal vaccines. In countries where pneumococcal and yellow fever vaccines are scheduled simultaneously with measles vaccine, confusion between them is also likely. TT-containing vaccines are increasingly delivered through a mixture of routine and campaign strategies throughout life, and not only during pregnancy, making it extremely difficult to obtain information on the lifetime total of doses received. (World Health Organization 2006)

Problems will be compounded for estimating coverage of vaccination at older ages (e.g., measles second dose, booster doses of DTP-containing vaccines, and HPV to girls aged between 10-14 years). For surveys this can mean sampling additional cohorts. For HPV especially it entails decisions regarding who the respondent will be (the girl or the adult caretaker, both), access during the day (usually when interviews are conducted) to girls likely to be in school, etc.

3. Data analysis and reporting issues

3.1. Outcomes

Traditionally, vaccination coverage surveys have evaluated the proportion of persons who have been vaccinated, by card only or 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, including only those DTP doses with a minimum interval of 28 days and measles at a minimum age of 270 days. When dates of vaccination are obtained from the HBVR or HF-record, dosage timing can also be reviewed to provide a more detailed or nuanced perspective on the effectiveness of vaccination programs. Doses administered before the minimum age, or before the appropriate intra-dose interval has elapsed are considered to be *invalid* and should be repeated. Doses administered later than recommended leave children susceptible to infection for longer than is necessary and indirectly put others at risk as well. Timeliness can be illustrated through simple graphs of the distribution of age at receipt of each dose compared to the national schedule (Jahn et al 2008, Clark et al 2009) or by time-to-event curves of the cumulative coverage by age (Jahn et al 2008, Dayan, Shaw et al. 2006). The mean number of extra days or weeks that children remained under-vaccinated and at risk of disease [37,39,40] and risk factors for delay in vaccination can be assessed (Dayan, Shaw et al. 2006, Luman et al 2005).

3.2. Bias

One important factor associated with survey-related inference is the sampling weight assigned to each response. Those weights are related to the probability of being selected into the sample and are based on population estimates. In many cases cluster surveys are designed to be *self-weighting,* which is to have equivalent weights. If the population estimates are wrong or outdated, or if they are based on an age group other than children 12-23 months, then the weights may be incorrect and may result in a biased estimate of coverage. If a household listing step is included in the survey preparation and sampling stages, the number of eligible households and number of interviewed households is known in each cluster and appropriate weights can be calculated and used to derive national estimates and confidence intervals. DHS and MICS nowadays recommend calculating and using sampling weights for each interviewed household and individual.

If respondents are not selected randomly and if the same forces that influence participation in the survey also influence participation in vaccination (i.e., safety of workers, or pockets of persons who are hard to reach), then the coverage estimates can be biased upwards.

3.3. Precision of survey estimates

The standard error of the coverage estimate is traditionally used to report a 95% confidence interval around the point estimate. The confidence interval for a proportion is affected by the sample size, the sampling design, and the underlying proportion itself. In the absence of meaningful strata, a simple random sample (SRS) drawn from the population of interest would result in the narrowest confidence interval for a set sample size but the SRS is impractical, so cluster sampling is usually employed to keep survey costs and timeframe manageable. When other factors are held constant, narrower CIs result from larger samples, samples that are spread across more clusters and proportions that fall closer to 0 or 100% than to 50%.

Application of standard statistical techniques to estimate CIs has been challenged for surveys such as the EPI which use non-probability sampling of households within each cluster (Bennett, Radalowicz et al. 1994), although simulations of results from EPI surveys showed that in general they gave confidence intervals within the desired precision of $\pm$ 10 percentage points. (Lemeshow, Tserkovnyi et al. 1985) Some of the variations on the EPI cluster survey take a probability sample (e.g., SRS or systematic random sample) within each cluster. (Turner, Magnani et al. 1996; Milligan, Alpha et al. 2004; Myatt, Feleke et al. 2005; Grais, Rose et al. 2007; Luman, Worku et al. 2007) A study in two districts of Ethiopia showed that systematic random sampling could be substituted for the EPI method of household selection with no increased cost and little increased time or complexity (Luman, Worku et al. 2007), and this makes it possible to calculate sampling weights and construct meaningful CIs.

LQAS is not primarily designed to define coverage with a given precision, but rather to classify areas according to adequate or inadequate coverage. In order to make classifications with a small sample size, LQAS has a central range of prevalence that is not excluded by either classification. That is to say that *adequate* means either high or medium coverage, and *inadequate* means either low or medium. Neither classification excludes the medium category, which is sometimes called the *gray area*. For fixed values of alpha and beta (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 precise conclusion about what the coverage is likely to be. (See Figure 2.) When data are combined across numerous lots, it is possible to estimate a region-wide proportion and confidence interval using formulae from stratified sampling. LQAS sample sizes are usually selected based on alpha and beta in individual lots rather than the width of the region-wide CI.

It may be worthwhile to point out one popular LQAS design that is usually misinterpreted by its proponents. The decision rule of 13 or more vaccinated children among a sample size of 19, used for example in (Valadez and Devkota 2002; Valadez, Weiss et al. 2002; Valadez, Weiss et al. 2003; Weiss, Burnham et al. 2009), allows the following conclusions:

- If fewer than 13 children are vaccinated, the coverage is classified as *inadequate* and one can conclude that coverage is *below 80%* with at least 90% certainty (Figure 2b).
- If 13 or more vaccinated children are identified among the 19, the lot is classified as *adequate* and one can conclude that coverage is *above 50%* with at least 90% certainty.

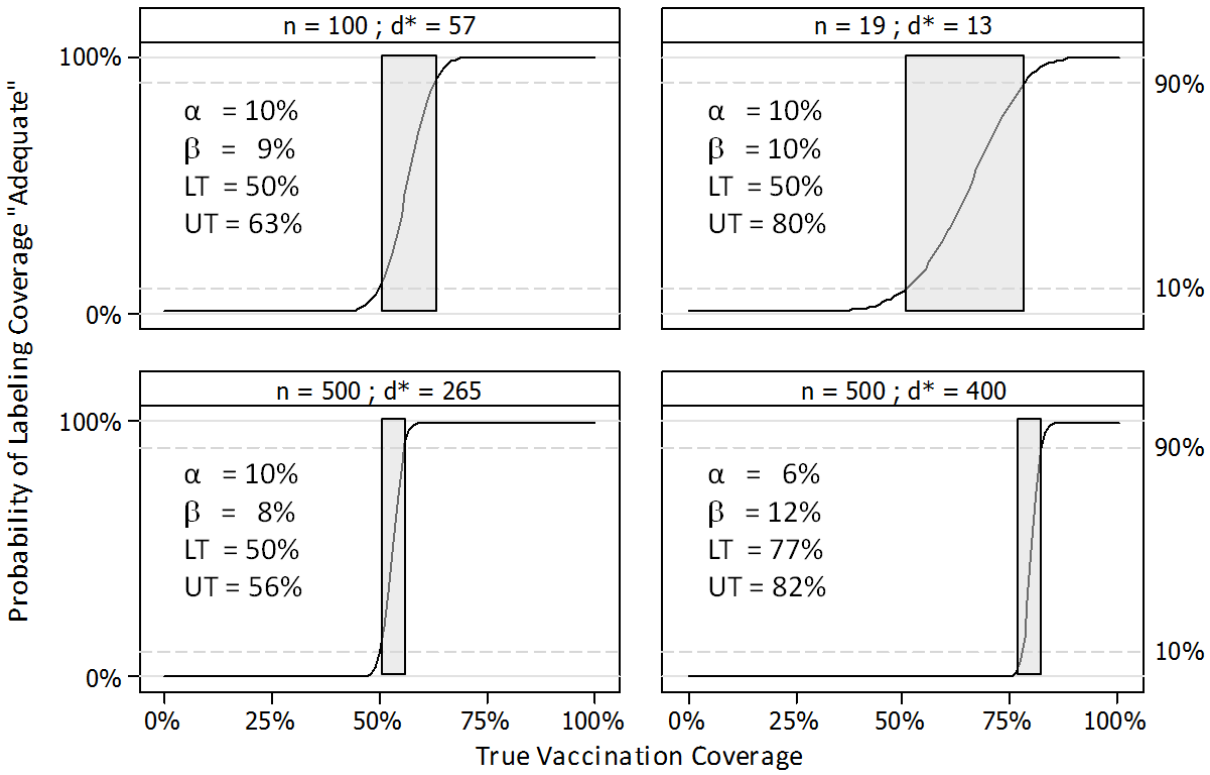
Authors such as Valadez and Weiss have encouraged the interpretation that 13 or more vaccinated children means > 80% coverage with the 13/19 design, but this incorrectly excludes the gray area. It is absurd to conclude with 90% confidence from a sample proportion of 13/19=68.4% that the population proportion is > 80%, but that is indeed a pervasive practice. LQAS designs can be powerful and cost effective, but they must be interpreted correctly. See (Rhoda, Fernandez et al. 2010) for additional cautions on LQAS design and interpretation in public health.

LQAS

- Uses small sample sizes
- Classifies lots as *adequate* or *inadequate*
  - *Adequate* means not low; (**may be medium**)
  - *Inadequate* means not high; (**may be medium**)
  - **Neither classification excludes medium (*gray area)***

Size and position of gray area can be tuned by selecting different designs

Figure 2. Operating characteristic curves for four LQAS sampling plans. In the nomenclature of this paper, values of coverage to the left of the gray area are *low;* those to the right of the gray area are *high;* and values of coverage in the gray area are *medium.* In each panel, the curve indicates the probability of finding *d** or more vaccinated children in a random sample of size *n*. LT = lower threshold; lots with coverage ≤ LT will be classified as having *inadequate* coverage with probability ≥ (1-α). UT = upper threshold; lots with coverage ≥ UT will be classified as having *adequate* coverage with probability ≥ (1-β). The *gray area* is the region where LT < coverage < UT; in that region the probability of concluding that coverage is adequate lies between α and (1-β); 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**.



3.4. Cluster surveys: Design Effect (DEFF) & intracluster correlation coefficient (ICC)

Because in general, individuals living in one cluster of a population are more likely to be homogeneous than persons from different clusters, a cluster sample will have a wider confidence interval than a SRS, by a factor known as the square root of the *design effect (DEFF)* which relates to the average correlation between responses from two respondents from the same cluster (ICC):

Cluster sample CI width with N respondents = SRS CI width with N respondents * sqrt(DEFF),

where DEFF = (1 + (m-1) x ICC) and m= the average number of respondents per cluster. (Bennett, Woods et al. 1991)

Note that if ICC=0 then DEFF=0, and the cluster survey yields CIs as narrow as those from a SRS. ICC varies by survey response variable, by population, and even by time period within a population. DEFF varies by ICC and by the size of the sample from each cluster; as m increases, so does DEFF. The EPI cluster survey assumes that the DEFF will be no worse (no larger) than 2 when m = 7, which implies an ICC ≤ 1/6. It is very helpful for coverage surveys to report the values of m that they employ and of DEFF or ICC that they observe, so subsequent survey designers can calculate what sample sizes will meet their goals for precision. The relationship between m and DEFF should also be considered. Although such guidance is missing from the 2005 WHO manual on cluster surveys (World Health Organization 2005), it is included in the more recent companion manual on sample design and procedures for Hepatitis B immunization surveys. (World Health Organization 2012b)

3.5. Missing data

Missing data arise for at least three reasons in coverage surveys. First, some respondents fail to give a valid answer to some questionnaire items. Second, interviewers sometimes fail to follow the questionnaire's skip patterns correctly, inadvertently skipping questions they should have asked. Third, some respondents may refuse to participate or may not be reachable after having been selected into the sample. Missing data on individual survey questions is known as *item nonresponse* and missing data on an entire respondent is known as *unit nonresponse.*

The consequences of missing data vary depending on the whether the cause of missingness is related to the topic being studied by the survey. Under many circumstances data may be said to be *missing at random* because the reason for missingness is unrelated to the survey topic. Data that are

missing at random are a disappointing inconvenience that makes survey estimation and inference results less precise, but they do not bias the results. Under other circumstances, the reasons the data are missing are related to the topic under study and are not explained by the observed data; in those cases the results of the survey are mathematically biased by virtue of not incorporating those missing outcomes. Consider, for example, a caregiver who refuses to participate in a vaccination survey because she routinely refuses to have her children vaccinated and fears some sort of punishment. Or consider a family where everyone in the household, including the children, works in the fields all day. That family is less likely to have time to attend a clinic for vaccination or to be at home when an immunization team comes around and less likely to be at home when a coverage survey team comes around.

In the analysis phase, missing data are handled in a variety of ways. Most commonly analyses are restricted to respondents with complete data; if data are not missing completely at random, then the results will be biased. Analyses of DHS and MICS surveys prior to 2002 showed that maternal recall 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 (Brown et al 2002). Inclusion of children without HBRs is only possible for calculation of percentage coverage and not for assessing timeliness of vaccination. Alternatively, analysts may impute (fill in) hypothetical, plausible values for missing responses, e.g., by taking the mean of the observed values or the same value as the record before the missing value. Imputing a single value fails to account correctly for the uncertainty associated with selecting an arbitrary (though perhaps plausible) value to impute; these methods should be avoided.

More sophisticated methods combine available information from survey responses with assumptions to generate information about the missing value or about the reason the data are missing, and can account properly for the uncertainty due to missingness. (Schafer 1997; Little and Rubin 2002) These methods include integrating over a likelihood function, or imputing numerous values for each missing datapoint, which is known as *multiple imputation*. Some high profile public-use datasets, like the U.S. National Health and Nutritional Examination Survey (NHANES), generate multiply imputed datasets for public release and provide documentation of the steps used for imputation and guidelines for downstream analysis. (Schafer 2001)

DHS and MICS do adjust sampling weights for non-response, but do not employ multiple imputation in vaccination coverage estimates or in their publicly released datasets. Inference about child health,

including vaccination coverage, in those surveys uses the survey sampling weights associated with the caregiver who answered the questions. Different coverage rates can be estimated for children whose vaccination cards were available versus those who relied upon caregiver recall.

3.6. How do consumers use the estimates of uncertainty?

Although survey results should always be reported with their respective confidence intervals, in the field of vaccination coverage, the point estimates garner most of the attention. Summaries that describe achievements and trends often fail to point out uncertainty (Duclos, Okwo-Bele et al. 2009; Brown, Burton et al. 2011) as WUENIC estimates are generally used which lack a way of expressing uncertainty even when they are based on survey results. Performance-based financing compensates countries for the estimated number of additional children immunized, ignoring any uncertainty whatsoever.

The EPI cluster survey was designed to yield prevalence precision $\pm$ 10% back in an era when coverage estimates were typically < 70%. Now that estimates are higher, (farther from 50%) the CIs will be narrower, and thus should be calculated and reported so that readers do not assume they are $\pm$ 10%. The reader should also take care to understand whether probability sampling was used in an EPI coverage survey.

When LQAS is used to classify catchment areas, a result of *adequate* is sometimes cause to divert special resources elsewhere. (Valadez 1991; Valadez 2004) Doing so can have important consequences on the district that loses those resources, so it is very important for LQAS consumers to understand that classification errors are likely with small sample sizes, and to require strong evidence in order to declare intervention success and withdraw special attention. Unfortunately, some investigators and some freely available training materials have made misleading claims about what it is prudent to conclude with small sample LQAS designs, suggesting that a classification of *adequate* means that the prevalence is high rather than high or medium. (Rhoda, Fernandez et al. 2010) Such a claim may be very attractive to program managers, but it is not in the interest of public health to declare programmatic success and possibly divert resources elsewhere prematurely. LQAS designs should protect the people being studied and should draw conclusions that are strongly supported by the data at hand; the gray area should not be excluded from the list of likely prevalence values.

4. Discussion

Vaccination coverage is an important indicator used not only to monitor immunization programs but also health systems performance at national and global levels. Its accurate measurement is critical to ensure that performance-based financing rewards real performance, and to guide program implementation. At national and sub-national levels, identifying low coverage areas is important to trigger action to reach underserved children, who are often those at highest risk of mortality should they acquire a vaccine-preventable infection. (Rheingans, Atherly et al. 2012) Vaccination coverage is a key component of cohort analyses which estimate the build-up of susceptible children after a measles campaign and identify when follow-up campaigns are needed. (de Quadros, Olive et al. 1996) Inflated coverage data contribute to delays in implementing follow-up campaigns and allows measles outbreaks to occur with substantial avoidable mortality. (Simons, Ferrari et al. 2012) Accurate recording of an individual's vaccination status is also essential in order for health workers to determine eligibility for vaccination at each health center visit, and to evaluate vaccine effectiveness using case-control or cohort approaches. (Orenstein, Bernier et al. 1985)

Vaccination coverage is clearly defined and understood by health workers and policy makers, making it perhaps the simplest indicator of child health. Nonetheless, its measurement is beset with challenges. (Murray, Shengelia et al. 2003; Lim, Stein et al. 2008) Estimates reported by countries to international agencies using administrative data are frequently inflated to unfeasible levels, and PBF may encourage over-reporting. Surveys are traditionally considered a more accurate method of measurement, but they have many sources of error and it is essential to review the methods of any survey before assuming that its results are accurate. Particularly important questions include:

1. How likely is information bias to have biased results?

    - For what proportion of subjects were HBVRs available, and is there any external information on the quality of recording in the relevant country (e.g., data quality assessments)?

    - For those with no HBVR, were other written documents (e.g., health facility registers) consulted, and is there any external information on the quality of recording in the relevant country (e.g., data quality assessments)?

    - How likely is it that the verbal history was reliable? Factors which may make it less reliable include a complex and/or recently changed immunization schedule, a long recall period, a long multi-purpose questionnaire, persons other than the mother being interviewed, and respondents having low formal education.

2. How representative was the survey sample of the target population?

    - Were large parts of the population missed from the sampling frame thought to have a different probability of vaccination (e.g., conflict-affected areas; areas with large populations of unofficial residents who may not be included in census data)?

3. How likely was selection bias to have occurred?

    - Were households selected using probability sampling methods?

    - Were revisits conducted and if so, were these done at times most likely to locate a knowledgeable respondent?

    - How closely were field workers supervised and what quality control procedures were done?

4. What quality control procedures were done during data management and survey analysis?

5. How precise are the estimates?

    - Is the confidence interval presented and is this used appropriately in the survey report?

    - In LQAS surveys, is the gray area (neither adequate nor inadequate) clearly described and correctly interpreted?

Two of the most important potential sources of error are information bias leading to misclassification of vaccine status and selection bias. Information bias can be reduced by seeking more than one source of information on vaccination and triangulating information. (Luman, Ryman et al. 2009) The magnitude of potential misclassification is illustrated by the low HBVR availability in recent surveys in countries which contribute the most to global immunization coverage estimates (Brown 2012). In India, WHO-UNICEF ("WUENIC) coverage estimates since 2009 have been based on a 2008 survey in which cards were seen for only 52% of children. (World Health Organization 2011b) In Nigeria, in the 2009 national immunization coverage survey 40% of 19551 children had cards; DTP1 coverage was only 29% if only card-documented vaccination was accepted but 73% including a verbal history of vaccination. In the 2007 DHS, cards were seen for only 26% of 4945 children. (World Health Organization 2011c) Similar differences have been noted in many other countries. Records of TT may be available for the most recent pregnancy but not for doses in previous pregnancies, in mass campaigns, or doses of DTP in early childhood. Serological surveys show that vaccination coverage surveys tend to underestimate the prevalence of protection against neonatal tetanus.

To reduce information bias, improved recording of vaccination and access to records is vital; in the long term this will likely involve computerized registries which can also improve program management such as tracking activities, stock keeping and vaccine supply. In the shorter term, much more attention is needed to designing appropriate vaccination records (Usman, Rahbar et al. 2011), encouraging families to keep them, ensuring that survey enumerators conduct revisits if a card is potentially available, adding questions about the condition of the vaccination card (e.g., is it the original record or a replacement; does it have adequate space for recording all vaccines; is it easily legible), and referring to health facility-based records at least on a subsample of children. Discrepancies between information from different sources should be reported in survey results, and discussion of discrepancies at local and national levels should help to improve recording practices. In multi-indicator surveys, vaccination coverage could be cross-tabulated against coverage of other interventions for persons with and without a HBVR; if the verbal history is reliable the same associations should be found in both groups.

Selection bias may arise from the use of suboptimal sampling frames, non-probability sampling methods, and/or enumerator procedural failures. Groups which may be omitted from sampling frames include those living in conflict-affected areas , geographically remote or inaccessible areas, rapidly growing urban areas (Bharti, Tatem et al. 2011), street children, and recent migrants (temporary or permanent), all of whom may be at greater risk of vaccine-preventable disease morbidity and mortality. Depending on the context and purposes of the survey, special effort may be needed to update the population data used for the sampling frame or to conduct special surveys among these groups; the need for this could be assessed initially by modeling the potential effect of plausible variations in each of these factors under different assumptions about coverage in those included or excluded from the frame.

Biomarkers are a potential tool to monitor program performance but have several limitations for validating coverage. Currently, serological assays capable of distinguishing immunity generated following vaccination from that following “natural” infection are only available for tetanus toxoid (since infection does not generate lasting immunity) and subunit vaccines , where vaccines induce only antibody to antigens included in the vaccine and not to other antigens found in the whole organism. For hepatitis B, for example, vaccine induces antibody only to the surface antigen whereas infection induces long-lasting antibody to both surface and core antigens. (Cutts and Hall 2004) Both tetanus toxoid and hepatitis B vaccines are given in multi-dose schedules and a large proportion of individuals have antibodies after two doses, however (subsequent doses being given to ensure long-lasting immunity), thus detection of antibodies to the relevant antigen does not indicate reliably how many doses have

been received. (Tapia, Pasetti et al. 2006) Furthermore, absence of antibody does not mean that the child had never been vaccinated; seronegativity could instead be observed after only one dose or after multiple doses of vaccine whose potency had been reduced by freezing. These biomarkers are therefore potentially useful to estimate population-level protection (Fortuin, Maine et al. 1995) but not necessarily to validate coverage measurements. For vaccines such as measles and rubella in settings such as the Americas where infection has been reduced to very low levels, serological surveys could theoretically be used to validate coverage, (though assumptions about vaccine effectiveness would still be needed), but disease surveillance is considered a more cost-effective way to monitor sustained success. The development of antibody assays on oral fluid samples for measles (Nigatu, Nokes et al. 1999) (Nigatu et al 1999) and tetanus (Tapia et al 2006) may make surveys with repeated sample collection more acceptable and allow evaluation of measles vaccine campaigns. (Nigatu, Samuel et al. 2008)

Large multi-purpose surveys like DHS have the goal of providing a coverage estimate from a probability sample that can be used for meaningful comparisons over time or between locations. Those survey projects usually take a year or more to complete and make results available. In other cases the estimates are needed quickly for a particular purpose. Smaller surveys may be less precise but allow greater attention to reducing non-sampling errors especially if they are single-purpose, and if probability sampling is used and survey procedures implemented rigorously, high-quality coverage estimates can be obtained. The EPI cluster sample is practicable to conduct at sub-national level and can yield useful and valid results especially if done according to the updated WHO guidance including appropriate calculation of sample size and use of probabilistic sampling (WHO 2012b). Whatever type of survey is done, all efforts must be made to reduce bias and this may mean that additional skilled resources are provided to DHS or MICS survey teams to enable a focus on obtaining documentation of vaccination for the maximum number of children.

Finally, some estimates are needed on a very rapid basis (e.g., during campaigns, while the vaccination team is still in the area), to decide whether to release a vaccination team to move to the next district or have them re-canvas an area that was inadequately covered the first time through. In those cases the balance between obtaining a statistically valid estimate or adopting a purposive sampling method focused on areas thought most likely to have low coverage needs to be considered. Some evaluators have found such a purposive sampling method to work well, giving same-day results at low-cost. (Izurieta, Venczel et al. 2003) Recent experience with rapid monitoring of polio campaigns

using purposive sampling elsewhere, however, has been less positive and current WHO plans are to use LQAS for polio campaign monitoring. It is essential that LQAS results be interpreted appropriately with due recognition of the gray area.

In conclusion, while it is critical to recognize the limitations of coverage as an indicator of program success, there is great potential to use even imperfect coverage data together with other program indicators to improve program management. For global policy-making, it is essential to assess the potential for bias in coverage estimates from whatever source. The inclusion beginning in 2011 of a "grade of confidence" (GOC) in the WUENIC in country reports produced by WHO-UNICEF is a first step[1]. The GOC reflects the degree of empirical support upon which the estimates are based and is not a judgment of the quality of data reported by national authorities. Further work to incorporate assessment of the quality of surveys when these are available will increase the helpfulness of the GOC. Policy makers also need to pay attention to the confidence intervals, not only the point estimates, when using survey results.

As a measure of population protection, coverage is limited by assumptions about vaccine effectiveness and thus is helpful but not sufficient, and additional information on effective vaccine management assessments, surveillance, outbreak investigations and special studies is needed to obtain a fuller picture of program impact. (Lowther, Curriero et al. 2009) For some vaccines, biomarkers are already available which can be used to measure the prevalence of immunity in the target population while for others, better biomarkers are needed. There is currently tension between PBF systems which reward high coverage and efforts to improve the quality of coverage measurement. It is time to reward actions which improve the quality of data and not only a country's apparent coverage achievements.

[1] A score of 3 indicates that the estimate is supported by reported data (R+), coverage recalculated with an independent denominator from the World Population Prospects from the UN Population Division (D+) and at least one supporting survey within 2 years (S+). While well supported, the estimate still carries a risk of being wrong. A score of 2 indicates that the estimate is supported by at least one data source; [R+], [S+], or [D+]; and none of the data sources challenges the estimate. A score of 1 indicates that there are no directly supporting sources of data.

5. References

Ahmed, S., D. Ali, et al. (2009). Evaluation of UNICEF Multiple Indicator Cluster Surveys Round 3 (MICS3) Final Report. Report to UNICEF under contract no. EO/ICC/2007/04, John Snow, Inc. (JSI).

Bennett, S., A. Radalowicz, et al. (1994). "A computer simulation of household sampling schemes for health surveys in developing countries." International Journal of Epidemiology **23**(6): 1282-1291.

Bennett, S., T. Woods, et al. (1991). "A simplified general method for cluster-sample surveys of health in developing countries." World Health Statistics Quarterly **44**(3): 98-106.

Bharti, N., A. J. Tatem, et al. (2011). "Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery." Science **334**(6061): 1424-1427.

Borgdorff, M. W. and G. J. Walker (1988). "Estimating vaccination coverage: routine information or sample survey?" Tropical Medicine & Hygiene **91**(1): 35-42.

Bosch-Capblanch, X., O. Ronveaux, et al. (2009). "Accuracy and quality of immunization information systems in forty-one low income countries." Tropical Medicine & International Health **14**(1): 2-10.

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 August 29, 2012.

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. 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 August 29, 2012.

Brown, D. W., A. Burton, et al. (2011). "A mid-term assessment of progress towards the immunization coverage goal of the Global Immunization Vision and Strategy (GIVS)." BMC public health **11**(1): 806.

Brown DW (2012) Child Immunization Cards: Essential Yet Underutilized in National Immunization Programmes. Open Vaccine Journal 5: 1-7.

Burton, A., R. Monasch, et al. (2009). "WHO and UNICEF estimates of national infant immunization coverage: methods and processes." Bulletin of the World Health Organization **87**(7): 535-541.

Clark A, Sanderson C (2009) Timing of children's vaccinations in 45 low-income and middle-income countries: an analysis of survey data. The Lancet 373: 1543-1549.

Cotter, B., V. Bremer, et al. (2003). "Assessment of neonatal tetanus elimination in an African setting by lot quality assurance cluster sampling (LQA-CS)." Epidemiology and infection **130**(2): 221-226.

Cutts, F. and A. Hall (2004). "Vaccines for neonatal viral infections: hepatitis B vaccine." Expert Review of Vaccines **3**(4): 349-352.

Cutts, F. T., P. G. Smith, et al. (1990). "Field evaluation of measles vaccine efficacy in Mozambique." American Journal of Epidemiology **131**(2): 349-355.

Dayan, G. H., K. M. Shaw, et al. (2006). "Assessment of Delay in Age-appropriate Vaccination Using Survival Analysis." American Journal of Epidemiology **163**(6): 561-570.

de Quadros, C. A., J. M. Olive, et al. (1996). "Measles Elimination in the Americas Evolving Strategies." The Journal of the American Medical Association **275**(3): 224-229.

Dodge, H. F. and H. G. Romig (1959). Sampling Inspection Tables - Single and Double Sampling. New York, John Wiley & Sons, Inc.

Duclos, P., J. M. Okwo-Bele, et al. (2009). "Global immunization: status, progress, challenges and future." BMC International Health and Human Rights **9**(Suppl 1): S2.

Fortuin, M., N. Maine, et al. (1995). "Measles, polio and tetanus toxoid antibody levels in Gambian children aged 3 to 4 years following routine vaccination." Transactions of the Royal Society of Tropical Medicine and Hygiene **89**(3): 326-329.

Gareaballah, E. T. and B. P. Loevinsohn (1989). "The accuracy of mother's reports about their children's vaccination status." Bulletin of the World Health Organization **67**(6): 669.

GAVI. (2012a). "Country Eligibility policy - Finance & Programmatic policies." Retrieved 6/29/2012, 2012, from http://www.gavialliance.org/about/governance/programme-policies/country-eligibility/.

Grais, R. F., A. M. C. Rose, et al. (2007). "Don't spin the pen: two alternative methods for second-stage sampling in urban cluster surveys." Emerging Themes in Epidemiology **4**(8).

Hadler, S. C., V. Dietz, et al. (2008). Vaccination programs in developing countries. Vaccines. S. A. Plotkin and W. A. Orenstein. Philadelphia, Saunders.

Henderson, R. H. and T. Sundaresan (1982). "Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method." Bulletin of the World Health Organization **60**(2): 253.

Hull, B., A. Dey, et al. (2011). "Immunisation coverage annual report, 2009." Communicable Diseases Intelligence **35**(2): 132-148.

Izurieta, H., L. Venczel, et al. (2003). "Monitoring Measles Eradication in the Region of the Americas: Critical Activities and Tools." Journal of Infectious Diseases **187**(Supplement 1): S133-S139.

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

Kim, S. H., L. Pezzoli, 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**(1): e29116.

Langsten, R. and K. Hill (1998). "The accuracy of mothers' reports of child vaccination: evidence from rural Egypt." Social Science & Medicine **46**(9): 1205-1212.

Lemeshow, S. and S. Taber (1991). "Lot quality assurance sampling: single-and double-sampling plans." World Health Stat Q **44**(3): 115-132.

Lemeshow, S., A. G. Tserkovnyi, et al. (1985). "A computer simulation of the EPI survey strategy." International Journal of Epidemiology **14**(3): 473-481.

Lim, S. S., D. B. Stein, et al. (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." The Lancet **372**(9655): 2031-2046.

Little, R. J. A. and D. B. Rubin (2002). Statistical analysis with missing data. Hoboken, NJ, John Wiley & Sons.

Lowther, S. A., F. C. Curriero, et al. (2009). "Population immunity to measles virus and the effect of HIV-1 infection after a mass measles vaccination campaign in Lusaka, Zambia: a cross-sectional survey." The Lancet **373**(9668): 1025-1032.

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.

Luman, E. T., T. K. Ryman, et al. (2009). "Estimating vaccination coverage: Validity of household-retained vaccination cards and parental recall." Vaccine **27**(19): 2534-2539.

Luman, E. T., A. Worku, et al. (2007). "Comparison of two survey methodologies to assess vaccination coverage." International Journal of Epidemiology **36**(3): 633-641.

Measure DHS. (2012). "What We Do." Retrieved 6/1/2012, 2012, from www.measuredhs.com/What-We-Do/Survey-Types/DHS.cfm.

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 August 29, 2012.

Milligan, P., N. Alpha, et al. (2004). "Comparison of two cluster sampling methods for health surveys in developing countries." International Journal of Epidemiology **33**(3): 469-476.

Murray, C. J. L., B. Shengelia, et al. (2003). "Validity of reported vaccination coverage in 45 countries." The Lancet **362**(9389): 1022-1027.

Myatt, M., T. Feleke, et al. (2005). "A field trial of a survey method for estimating the coverage of selective feeding programmes." Bulletin of the World Health Organization **83**(1): 20-26.

Nigatu, W., D. Nokes, et al. (1999). "Detection of measles specific IgG in oral fluid using an FITC/anti-FITC IgG capture enzyme linked immunosorbent assay (GACELISA)." Journal of Virological Methods **83**(1): 135-144.

Nigatu, W., D. Samuel, et al. (2008). "Evaluation of a measles vaccine campaign in Ethiopia using oral-fluid antibody surveys." Vaccine **26**(37): 4769-4774.

Orenstein, W. A., R. H. Bernier, et al. (1985). "Field evaluation of vaccine efficacy." Bulletin of the World Health Organization **63**(6): 1055-1068.

PATH. (2012). "Rethinking the vaccine supply chain." Retrieved 6/29/2012, 2012, from http://www.path.org/projects/project-optimize.

Pezzoli, L., S. Pineda, et al. (2009). "Cluster-sample surveys and lot quality assurance sampling to evaluate yellow fever immunisation coverage following a national campaign, Bolivia, 2007." Tropical Medicine & International Health **14**(3): 355-361.

Rheingans, R., D. Atherly, et al. (2012). "Distributional impact of rotavirus vaccination in 25 GAVI countries: estimating disparities in benefits and cost-effectiveness." Vaccine **30 Suppl 1**: A15-23.

Rhoda, D. A., S. A. Fernandez, et al. (2010). "LQAS: User Beware." International Journal of Epidemiology **39**(1): 60-68.

Robertson, S. E., M. Anker, et al. (1997). "The lot quality technique: a global review of applications in the assessment of health services and disease surveillance." World Health Statistics Quarterly **50**: 199-209.

Robertson, S. E. and J. J. Valadez (2006). "Global review of health care surveys using lot quality assurance sampling (LQAS), 1984-2004." Social Science & Medicine **63**(6): 1648-1660.

Ronveaux, O., F. Arrieta, et al. (2009). "Assessment of the quality of immunization data produced by the national individual registration system in Uruguay, 2006." Revista Panamericana de Salud Pública **26**(2): 153-160.

Ronveaux, O., D. Rickert, et al. (2005). "The immunization data quality audit: verifying the quality and consistency of immunization monitoring systems." Bulletin of the World Health Organization **83**(7): 503-510.

Rutstein, S. O. and G. Rojas (2006). Guide to DHS Statistics. Calverton, Maryland, Demographic and Health Surveys ORC Macro.

Schafer, J. L. (1997). Analysis of incomplete multivariate data. Boca Raton, FL, Chapman & Hall/CRC.

Schafer, J. L. (2001). Analyzing the NHANES III Multiply Imputed Data Set: Methods and Examples. Hyattsville, Maryland, Prepared for National Center for Health Statistics.

Simons, E., M. Ferrari, et al. (2012). "Assessment of the 2010 global measles mortality reduction goal: results from a model of surveillance data." Lancet.

Tapia, M. D., M. F. Pasetti, et al. (2006). "Measurement of tetanus antitoxin in oral fluid: a tool to conduct serosurveys." The Pediatric Infectious Disease Journal **25**(9): 819-825.

Turner, A. G., R. J. Magnani, et al. (1996). "A Not Quite as Quick but Much Cleaner Alternative to the Expanded Programme on Immunization (EPI) Cluster Survey Design." International Journal of Epidemiology **25**(1): 198-203.

UNICEF. (2012). "Childinfo." Retrieved 6/29/2012, 2012, from http://www.childinfo.org/.

Usman, H. R., M. H. Rahbar, et al. (2011). "Randomized controlled trial to improve childhood immunization adherence in rural Pakistan: redesigned immunization card and maternal education." Tropical Medicine & International Health **16**(3): 334-342.

Valadez, J. J. (1991). Assessing child survival programs in developing countries: testing lot quality assurance sampling, Harvard School of Public Health. Department of Population and International Health.

Valadez, J. J. (2004). E-mail to USAID responding to LQAS questions.

Valadez, J. J. and B. R. Devkota (2002). Decentralized supervision of community health program using LQAS in two districts of southern Nepal. Community-Based Health Care: Lessons from Bangladesh to Boston. J. Rhode and J. Wyon. Boston, Management Sciences for Health**:** 169-200.

Valadez, J. J., W. Weiss, et al. (2002). Assessing Community Health Programs A Participant's Manual and Workbook Using LQAS for Baseline Surveys and Regular Monitoring.

Valadez, J. J., W. Weiss, et al. (2003). Assessing Community Health programs: A Trainer's Guide: Using LQAS for Baseline Surveys and Regular Monitoring. London, Teaching Aids at Low Cost (TALC).

Valadez, J. J. and L. H. Weld (1992). "Maternal recall error of child vaccination status in a developing nation." American Journal of Public Health **82**(1): 120-122.

Weiss, W. M., G. Burnham, et al. (2009). "Evaluating the Experience of GAPS—A Methodology for Improving Quality of Mass Immunization Campaigns in Developing Countries." Journal of Health, Population, and Nutrition **27**(5): 684-695.

World Health Organization (2005). Immunization coverage Cluster survey: reference manual, Immunization, Vaccines and Biologicals.

World Health Organization (2006). "Tetanus vaccine WHO position paper." Weekly Epidemiological Record **85**(20): 197-208.

World Health Organization (2011a). Assessing Immunization Coverage with Clustered Lot Quality Assurance Sampling (Clustered-LQAS) Field Manual.

World Health Organization. (2011b). "India. WHO and UNICEF estimates of immunization coverage: 2010 revision." Retrieved June 29, 2012, 2012, from http://www.who.int/immunization_monitoring/data/ind.pdf.

World Health Organization. (2011c). "Nigeria. WHO and UNICEF estimates of immunization coverage: 2010 revision." Retrieved 6/29/2012, 2012, from http://www.who.int/immunization_monitoring/data/nga.pdf.

World Health Organization (2012a). Assessing Vaccination Coverage Levels Using Clustered Lot Quality Assurance Sampling - Field Manual - VERSION EDITED FOR THE GLOBAL POLIO ERADICATION INITIATIVE (GPEI). Geneva, Switzerland

World Health Organization (2012b). Sample design and procedures for Hepatitis B immunization surveys: A companion to the WHO cluster survey reference manual, Immunization, Vaccines and Biologicals.

World Health Organization. (2012c). "WHO Vaccine Preventable Diseases Monitoring System." Retrieved 6/29/2012, 2012, from http://apps.who.int/immunization_monitoring/en/globalsummary/scheduleselect.cfm.

World Health Organization. (2012d). "World immunization week 2012." Retrieved May 15, 2012, 2012, from http://www.who.int/immunization/newsroom/events/immunization_week/en/index.html.