World Health Organization

Vaccination Coverage Cluster Surveys:

Reference Manual

Version 3

WORKING DRAFT Updated July 2015

**Table of Contents**

List of boxes, figures and tables v

Abbreviations vii

Preface 1

1. Introduction 3

1.1. Why vaccination coverage is assessed 3

1.2. Methods for measuring vaccination coverage 3

1.3. Cluster surveys: a practical survey method for reliable results if designed appropriately and excellent quality control is done 4

1.4. Changes to previous methods and materials 6

2. Design the sample structure of the survey 9

2.1. Convene a survey steering group 10

2.2. Discuss the purpose of the survey 11

2.3. Identify primary questions that affect survey design and sample size 12

2.4. Define the target population 13

2.5. Set inferential goals 14

2.6. Select a survey design 15

2.7. Calculate the required sample size 16

2.8. Draft an analysis plan, table shells and report figures 24

2.9. Budget for the survey and estimate the timeline 24

2.10. Evaluate affordability and timeliness 27

2.11. Implications of adding routine immunization questions to a post-SIA survey 29

2.12. Designing for multiple outcomes 30

2.13. Designing for multiple geographic areas 30

2.14. Designing for multiple levels of administrative or geographic hierarchy 31

2.15. Reporting results by subgroups 33

3. Make concrete plans 34

3.1. Set survey schedule 34

3.2. Decide who will conduct the survey and create a project plan 34

3.3. Obtain ethical clearance 35

3.4. Design data collection tools and methods 35

3.5. Choose data analysis tools 40

3.6. Select a sample 40

3.7. Obtain access to health registers for vaccination records 48

3.8. Select and hire staff 49

3.9. Train staff 53

4. Conduct field work 56

4.1. Collect data from households 56

4.2. Check health registers from the health centre 58

4.3. Monitor the quality of field data collection 58

4.4. Check questionnaire forms and transmit 60

4.5. Clusters that become suddenly inaccessible 61

5. Data entry, cleaning and management 62

5.1. Database design 62

5.2. Data entry 62

5.3. Clean the dataset 63

5.4. Merge datasets and construct derived variables 64

5.5. Generate a codebook 66

6. Tabulation and analysis 68

6.1. Conduct descriptive analyses to characterise the sample and assess its quality 69

6.2. Calculate weights for analysis 73

6.3. Conduct standard analyses 73

6.4. Conduct additional analyses 78

6.5. Classifying coverage 86

7. Interpret, format, and share results 91

7.1. Draft the background section 92

7.2. Draft the survey methods and limitations 92

7.3. Draft the results section 93

7.4. Draft the discussion section 95

7.5. Develop implications and recommendations 95

7.6. Revise the report and obtain clearance on final draft 98

7.7. Share the results 98

References 101

List of boxes, figures and tables

**Boxes**

Box 1. Features of a good probability sample survey 41

Box 2. Using Google Earth to segment large enumeration areas 46

Box 3. Essential components of a report 91

**Figures**

Figure 1. Early steps in survey design 10

Figure 2. Precise estimation uses larger sample sizes than classification 16

Figure 3. The name “Organ Pipe Plots” is inspired by pipes like these 70

Figure 4. Organ pipe plots for four hypothetical strata, each with coverage of 50% 70

Figure 5. Organ pipe plots for four real strata from a self-weighted measles SIA 71

Figure 6. DTP3 coverage estimated at two different times with surveys of different sizes 84

Figure 7. Point estimate, 95% confidence interval and 95% lower confidence bound for coverage in hypothetical province #7 87

Figure 8. Point estimate, 95% confidence interval and 95% upper confidence bound for coverage in hypothetical province #7 87

Figure 9. Measles SIA coverage and confidence interval and bounds for 24 fictional districts and the province that they comprise; districts are sorted by estimated coverage 90

**Tables**

Table 1: Main potential sources of error and strategies to minimize them in immunization coverage surveys 8

Table 2. Timeframe for a national coverage survey 26

Table 3. Results of the household visits and interviews 69

Table 4. Crude vaccination coverage by source of information, by age at the time of the survey, among (N=\*) children aged 12–23 months 76

Table 5. Crude and valid vaccination coverage by age 12 months 77

Table 6. Survey-weighted dropout rates between different vaccine-dose combinations, by source of information 77

Table 7. Potential coverage achievable by time of survey among (n=\*\*) children with a documented source of information (card or clinic register), if all doses had been valid and all opportunities taken 81

Abbreviations

**BCG**: Bacillus Calmette-Guérin vaccine against severe forms of tuberculosis

**CI**: confidence interval

**DEFF**: design effect

**DHS**: Demographic and Health Survey

**DTPCV**: diphtheria–tetanus–pertussis- containing vaccine. DTPCV1 refers to first dose, DTPCV2 refers to the second, etc.

**EA**: enumeration area

**EPI**: Expanded Programme on Immunization

**GIS**: geographic information system

**GPS**: global Positioning System

**HBR**: home-based record

**HepB**: hepatitis B (vaccine)

**Hib**: *Haemophilus influenzae* type b (vaccine)

**HPV**: Human Papilloma Virus

**ICC**: intracluster correlation coefficient, or sometimes intraclass correlation coefficient

**ICT**: information and communication technology

**IPV**: inactivated polio vaccine

**LCB**: lower confidence bound

**LQAS**: Lot Quality Assurance Sampling

**MICS**: Multiple Indicator Cluster Survey

**MCV**: measles-containing vaccine; MCV1 refers to the first dose, MCV2 refers to the second dose

**MMR**: measles-mumps-rubella vaccine

**MOV**: missed opportunity for vaccination

**MR**: measles-rubella vaccine

**OPV**: oral polio vaccine

**PCV**: pneumococcal conjugate vaccine

**PPES**: probability proportional to estimated size

**PSU**: primary sampling unit

**RFP**: Request for proposals

**RI**: routine immunization

**RV**: rotavirus vaccine

**SIA**: supplementary immunization activity

**SOPs**: standard operating procedures

**Td**: tetanus and diphtheria toxoid – adult dose (vaccine)

**TT**: tetanus toxoid (vaccine)

**UCB**: upper confidence bound

**UNICEF**: The United Nations Children's Fund

**YF**: yellow fever

**WHO**: World Health Organization

Preface

The World Health Organization’s (WHO) Department of Immunization, Vaccines, and Biologicals has long provided guidance on assessing vaccination coverage using both cluster and Lot Quality Assurance Sampling (LQAS) survey methods.

Over time, Expanded Programme on Immunization (EPI) coverage surveys have increased in complexity, matching the evolution of the EPI since its inception in 1974. Although many of the previous surveys were likely done well, their implementation was often not thoroughly documented and the methods used were open to criticism. This document updates previous versions of the EPI coverage survey manual, focusing on methods to reduce bias, and improve the accuracy and precision of survey results.

This manual is for ministries of health (such as immunization programme managers, communicable disease epidemiologists and surveillance officers) and their partners who are considering an immunization coverage survey. The survey itself may be contracted out to a research, or other, institution via a request for proposals (RFP), in which case this manual should help groups who are writing the survey proposal to respond to the RFP as well as the team or committee who judges the responses, awards the contract and monitors its implementation.

Much of the document is written in technical language appropriate for readers with a university degree or equivalent in statistics or epidemiology, although the chapters on field implementation and use of results will be understood by those without such expertise. At a minimum, readers who will be tasked with designing the survey and analysing the data need to be very familiar with complex survey sampling, calculating sample sizes and conducting weighted analyses. Those who will be involved in implementing the survey must understand the principles of ensuring data quality, in particular how to ensure that fieldwork follows protocol and standard operating procedures. To make the document easier to read, an informal tone is used to say directly to the reader what should be done, even if the reader is not the person acting on all aspects of the survey.

The WHO recommends that immunization coverage surveys use probability sampling methods and, in general, use census data with lists of enumeration areas for the sampling frame. Therefore, excellent links with the central statistical office, or equivalent, will be needed, and surveys should to be planned well enough in advance to allow time to obtain census data and maps. A multi-disciplinary team or steering committee is recommended to oversee the survey, as detailed in Chapter 2, and should include statistical expertise and individuals familiar with using census data, geographic information systems (GIS) and maps.

Many countries obtain survey data on vaccination coverage every 3–5 years from large-scale multi-purpose survey programmes that meet most programme needs. Additional surveys may nonetheless be needed from time to time, for example, to evaluate coverage achieved by vaccination campaigns, or after major changes have occurred in the vaccination programme. Surveys should use rigorous statistical principles and prescriptive field protocols, which will require a substantial investment in time, expertise and resources. The role of vaccination coverage surveys in programme monitoring must be carefully defined to make the best use of resources. For example, it will rarely be a cost-effective use of resources to attempt to conduct surveys in every district of a country. At the most peripheral health system levels, practical field methods such as health facility-based assessments can evaluate multiple aspects of service provision, coverage and timeliness of each vaccine among clinic attendees, and can stimulate improvement of vaccination as well as recording practices.

This document is one of several current and forthcoming tools to help countries conduct high-quality immunization surveys. Other tools under development to complement this manual include software with standard code for analysing immunization survey data, training materials and methods, a step-by-step guide to survey implementation, and a discussion paper on defining the role of coverage surveys. The contents of this manual are as follows:

**Chapter 1, Introduction,** summarizes the purposes and common methods of measuring coverage together with key points for obtaining high quality data from surveys.

**Chapter 2, Design the sample structure of the survey**, discusses how to establish the objectives and inferential goals of a survey and how to select an appropriate design to meet these objectives. Guidance for estimating the cost and time of different design options is given, together with guidance on how to modify the design if certain options appear too costly, or are so large that there may be doubts about the ability to obtain high quality data in a timeframe that will be helpful to the end users of the information.

**Chapter 3, Make concrete plans**, explains how to prepare for fieldwork by planning the schedule, designing and pilot testing the data collection tools, obtaining ethical clearance for the survey, and assembling a field staff.

**Chapter 4, Conduct field work**, provides information on how to organize the survey in the field, with particular attention to methods to ensure good data quality. This chapter includes tips on the recruitment, selection, and training of field teams and supervisors, descriptions of the supervisor’s role and responsibilities, and examples of checks that should be done in the field.

**Chapter 5, Data entry, cleaning*,* and management**,explains how to design the database, enter the data, clean the data, merge datasets, and create a codebook (data dictionary).

**Chapter 6, Tabulations and analyses**, provides guidance on standard analyses to answer primary questions (such as coverage by given age) and secondary questions (such as missed opportunities for vaccination), including table shells.

**Chapter 7, Interpret, format*,* and share results**,offers guidance on how to interpret the estimates of coverage and how precise they are, to classify coverage at sub-national levels, and aggregate data to estimate coverage at higher levels. This chapter also offers guidance on what to include in the report, and importantly, how to communicate the results of the survey to stakeholders and stimulate appropriate action in response to the results.

WHO trusts that this working draft manual will facilitate the conduct of high-quality surveys and the use of data to improve immunization programme performance. It will be updated according to feedback from the field.

# Introduction

## Why vaccination coverage is assessed

Vaccination[[1]](#footnote-2) coverage is defined as the proportion of a given population that has been vaccinated in a given time period. It is estimated for each vaccine and, for multi-dose vaccines, for each dose received (e.g., diphtheria-tetanus-pertussis-containing vaccine (DTPCV1, DTPCV2)). It is usually presented as a percentage.

Measurements of vaccination coverage levels and trends are used to:

* monitor the performance of routine vaccination services at subnational and national levels, especially if administrative reports are thought to be unreliable;
* measure the effectiveness of interventions to increase coverage;
* evaluate how well a supplementary immunization activity (SIA) has reached the target population;
* provide insights into areas of programme weakness, for example, by showing the proportion of children receiving no vaccines at all (often an indicator of access to health services), estimating the rate of dropout between starting and completing the vaccination series (high dropout potentially indicating health system barriers to re-attendance or weakness of tracking activities), and estimating the frequency of missed immunization opportunities due to non-simultaneous vaccination;
* measure the coverage of vaccines recently introduced into the national immunization programme and compare this to coverage of traditional vaccines (if coverage of the newly introduced vaccine is lower, it may suggest vaccine supply problems and/or suboptimal information, education and communications activities around the new vaccine introduction);
* contribute data to models of the impact of vaccination on disease burden, including risk assessment of outbreak potential; and
* act as an indicator of programme readiness to introduce new vaccines, in particular for receiving support from the Gavi, the Vaccine Alliance for new vaccine introduction.

## Methods for measuring vaccination coverage

Vaccination coverage can be measured by administrative reports or by several types of surveys. Unfortunately, in many countries, administrative coverage estimates are inaccurate due to errors in the denominator (total target population), errors in recording vaccinations at health facilities, and errors in compiling the data on vaccinations to report to higher levels (Cutts, Izurieta & Rhoda, 2013). Substantial efforts are ongoing to improve administrative coverage estimates, including regular data quality self-assessments and development of appropriate action plans, development and rollout of registry-based systems, increased use of digital technology for the vaccine supply chain and for vaccination reporting, and renewed efforts to disseminate best practices in vaccination recording both on home-based and health facility records. Administrative data have the advantage of being available at all levels of the health system with very little delays, which enables programme managers to do real-time monitoring, investigate potential problems and take remedial action. Improving the accuracy of administrative data is a high priority. By improving recording practices and encouraging the retention of home-based records, investment in better administrative data will also improve the quality of survey data.

Surveys can be helpful to monitor coverage while efforts to improve administrative reporting systems are ongoing. In coverage surveys, evidence is collected from vaccination records, usually home-based records (HBRs), as well as from a vaccination history as recalled by the individual or, for a child, the child’s caretakers.

Some surveys supplement evidence from records and recall by collecting biological samples (usually blood, but sometimes oral fluid samples) and measuring the presence of antibodies. Serosurveys use methods for collecting and testing specimens from a defined population over a specified period of time to estimate the prevalence of antibodies against a given aetiologic agent as a direct measure of immunity.

There are, however, several difficulties in trying to correlate seroprevalence with vaccine coverage. First, for most vaccines, the presence of antibody following vaccination cannot be distinguished from that following natural infection. Exceptions are the presence of tetanus antibody (which indicates vaccination because infection does not generate lasting immunity) and hepatitis B vaccine (which induces antibody only to surface antigen whereas infection also induces antibody to other antigens such as core antigen). Second, for multi-dose vaccines, detection of antibodies does not indicate reliably how many doses have been received. Third, absence of detectable antibody does not necessarily mean that the individual was never vaccinated; the individual may not have responded to vaccination (for example, due to cold chain failure), or antibody levels may have waned to low levels that were not detected by the laboratory assay.

Biomarkers are therefore potentially useful to estimate population-level protection but not necessarily to validate coverage measurements or vaccination programme performance (Cutts, Izurieta & Rhoda, 2013; MacNeil, Lee & Dietz, 2014). The development of antibody assays on oral fluid samples for tetanus and measles may make surveys with repeated sample collection more acceptable, and facilitate evaluation of vaccination campaigns. Separate WHO guidelines for hepatitis B serosurveys have been published (WHO, 2011), and are under development for measles-rubella serosurveys. The measles-rubella guidelines will build on the general issues of survey design, sample selection, and field implementation described in this document. Serosurveys are not considered further in this document.

## Cluster surveys: a practical survey method for reliable results if designed appropriately and excellent quality control is done

Cluster surveys can overcome the shortcomings of administrative reports, and are more feasible to implement than surveys that use a simple random sample because fieldwork is concentrated in a given number of clusters (see Chapter 2 and Annexes B1, B2, and B3). Cluster survey methods can be used either to *measure* coverage achieved by the routine vaccination programme (providing a percentage coverage result with its 95% confidence interval for each vaccine-dose) or to *classify* coverage using qualitative labels like *probably adequate*, *probably inadequate*, or *intermediate*. Previously, lot quality assurance sampling (LQAS) was used to classify coverage, but this manual shows how cluster surveys may be used instead of LQAS for this purpose.

Probability samples are recommended at all stages of sampling and weighted statistical analyses. Probability samples allow you to:

* reduce the potential for selection bias due to fieldworker practices;
* increase the comparability of survey data with those from ongoing large multi-purpose surveys such as the Demographic and Health Surveys (DHS) [www.dhsprogram.com] and Multiple Indicator Cluster Surveys (MICS) [www.unicef.org/statistics/index\_24302.html]; and
* allow the calculation of meaningful confidence intervals and confidence bounds.

The advantages of a probability sample are that every eligible respondent has a chance of being selected for the sample, and the probability of the respondent’s selection can be calculated. This survey design yields an estimate of coverage with a calculated confidence interval for estimating coverage, or with one-sided upper or lower confidence bounds for classifying coverage.

The DHS and MICS use highly standardized probability sampling methods, and their sponsoring agencies provide substantial technical assistance and quality control for the design, implementation, analysis, and reporting of results (Hancioglu & Arnold, 2013). By contrast, the EPI coverage survey has historically been less standardized in its implementation and reporting. Although it has played an important role in monitoring programme performance over the past 30 years and in encouraging health workers to understand the status of vaccination of the communities they serve, the method has had certain disadvantages (Brogan, Flagg, Deming & Waldman, 1994; Cutts, Izurieta & Rhoda, 2013; Grais, Rose & Gurthmann, 2007), including:

* Non-probability sample: In the original *Immunization Coverage Survey: Reference Manual* (WHO/EPI/MLM/91.10), interviewers were instructed to go house to house from a starting point until they enrolled a quota, usually of 7 children per cluster. Although the starting point was identified using a random selection process, different households had unequal and unquantified probabilities of being selected as the starting point. This was not a true probability sample.
* Selection of households by fieldworkers: This practice could introduce bias if fieldworkers were tempted to prefer easily accessible households.
* Single design regardless of sample size or goals: There has been a tendency to use a single design (most often 30 clusters of 7 individuals per cluster) without appropriate adaptation of sample size and survey design according to survey goals, although the 2005 reference manual (WHO, 2005)gave guidance on how to adapt the design.
* Limited revisits: There was often a failure to conduct or document revisits to households where the respondent was not available at the first visit.
* No weight calculation: Assumptions about a self-weighting design were usually not valid because the sampling frame was out of date, inaccurate or incomplete, and non-probability sampling was used. No data were collected to allow calculation of appropriate weights.
* Limited ability to assess quality: It was difficult for external reviewers or policymakers to assess the quality and reliability of surveys because there was little or no documentation of quality control of fieldwork or of data management. Also, survey meta-data were rarely made available internationally.

Globally, immunization programmes have made remarkable progress since the EPI coverage survey was introduced. Most countries now have high average coverage of an increasing number of vaccines delivered to several different age groups. Newer vaccines are much more expensive than older vaccines, and strategies such as SIAs are resource-intensive, providing vaccines to wide age groups. Hence, it is ever more important to have high-quality data for programme monitoring and evaluation. When coverage surveys are done, results must be credible to national and international policymakers. This document offers updated guidance on EPI coverage surveys to address the changing context of the EPI.

## Changes to previous methods and materials

Improvements to the EPI survey method in this revision of the manual include the following changes:

**Use a probability-based sample.** Perhaps the most significant change is that WHO now strongly recommends creating a true probability-based sample, in which the probability of each child being selected is quantifiable and non-zero. A single-stage or two-stage probability sample may be used; see section 3.6 for guidance on how to choose between these options. *A probability sample will require the use of maps or satellite images of clusters*; see Annexes E and F for guidance on how to create and use these.

**Have households selected by a central group of planners rather than interviewers in the field.** The survey coordinator or statistician, and not field teams, must select the households regardless of whether a single-stage or two-stage design is used. Experience has shown that when field data collectors have the responsibility for selecting the households in a survey, they may tend to make decisions based on convenience, compromising the representativeness of results and probably biasing coverage estimates upwards. (For example, families missed by interviewers because they live in areas difficult to access may also be less likely to attend vaccination clinics.) The field data collectors should have no choice in which houses they visit. This will improve representativeness, as well as facilitate supervision and external monitoring of adherence to the survey protocol. See section 3.6.4.

**Eliminate the residency requirement.** The 2005 EPI manual proposed that only persons who had been residing in the area for at least six months be included in the sample. The updated guidance removes this requirement because it can lead to potential bias: migrant populations, including seasonal workers, would not be located in their usual residences and so would not be eligible to enter the survey at their temporary living site. They would thus not have the opportunity to be included in either sample. Given that highly mobile population groups may be less likely to be fully vaccinated, their exclusion could bias vaccination estimates upwards. Instead, WHO recommends including both residents and all other persons who slept in the household the previous night, as is done in DHS and MICS. Likewise, the document proposes adding a question to the individual questionnaire to document how long each surveyed individual has lived in that household. (For SIAs, the question could be expanded to determine whether they were living in the areas included in the SIA at the time of the SIA). Including all persons irrespective of residence will help immunization programmes assess their ability to enlist and provide services to any new arrival and track those who have moved into and out of an area. It will also allow the programme to assess an SIA’s success in reaching mobile as well as more settled populations.

**Interview every eligible child in the household.** Earlier protocols had interviewers select a single respondent when a household contained more than one eligible individual. This manual recommends collecting data for every eligible individual in every household surveyed. This will require careful recording of the household ID on survey forms, and appropriate accounting in analysis software to reflect an additional level of correlation between children in the same household. But it will facilitate estimation of total numbers of children, and eliminate a potential source of bias in which fieldworkers may have otherwise influenced survey results. This change will have the largest consequences in surveys with wide windows of age eligibility, such as measles SIAs where age eligibility may range from 9 months up to 15 years or older. See section 4.1.3.

**Conduct a weighted analysis.** Under the process set forth in this new manual, the probability of an individual being selected will vary from cluster to cluster, as will the number of completed questionnaires. Therefore, it is essential to conduct a weighted analysis that accounts properly for the complex sampling design, to avoid a biased estimate of coverage and confidence intervals. See section 6.2.

**Select an appropriate sample size for the survey goals.** The traditional EPI cluster survey chose a fixed sample of 7 children in 30 clusters (7 x 30) to guarantee a maximum absolute confidence interval width of ± 10% at an assumed coverage level of 50%, and design effect of 2. A maximum precision of ± 10% was acceptable at that time because vaccination coverage was expected to be fairly low, and programmatic decisions at such levels did not require greater precision. Nowadays, there is great variation between countries in terms of immunization schedules and programme strategies, and within countries in terms of coverage. There is a range of potential goals for immunization coverage surveys. This document offers updated guidance on estimating the appropriate sample size for a variety of goals, including detecting differences in coverage between administrative areas, detecting changes over time in the same administrative area, or confirming coverage levels in SIAs or other activities that require high levels of coverage. See section 2.7.

**Take account of multiple potential survey goals and determine the most feasible combination of goals to address in the survey.** One increasingly common scenario is that a survey is done to evaluate coverage in a SIA that targeted a wide age group (for example, up to age 15 years for measles-containing vaccine (MCV) or up to age 30 years for meningococcal vaccine), and programme planners and partners want to investigate variation in province or district coverage. A stratified cluster design may be used which has a sample size adequate for classification at peripheral levels and for estimation of coverage at higher levels, as long as probability sampling and strict quality control are used at all levels. We give guidance on how to calculate sample sizes for multiple objectives, how to review the priorities of each objective, and how to compromise where necessary. See section 2.12.

**Visit health facilities to find vaccination records.** Traditionally a child’s vaccination status has been inferred from home-based records or the caretaker’s memory. Given the number of vaccinations now offered and the potential to confuse vaccinations received during SIAs with those received through the routine programme, it is increasingly difficult for caretakers to know and remember all the vaccinations a child had received. When the home-based record is not available, or is poorly filled (illegible or incomplete), WHO recommends that vaccination documentation be sought at the child’s usual health care facility(s) in addition to asking for and recording the caretaker’s recall about the child’s vaccination history.

The caretaker’s recall is still useful because it may be difficult to obtain complete vaccination data from health facilities for several reasons. The individual may have been vaccinated at multiple health facilities (including some in other geographic areas), or given vaccinations during outreach sessions that were not recorded in the health facility register. Vaccinations that are recorded are often done by date of visit rather than by registering each individual on only one page of the register, making it difficult to search for the relevant data. Another challenge is that registers may not be available for all age cohorts included in the survey.

When feasible, using health facility records is an important additional component of credible coverage surveys, until an effective method is implemented to improve the availability, use, and retention of home-based records. See section 3.7. This requires extra time and expense, but should increase the accuracy of coverage estimates. It has the added benefit of reinforcing the importance of good record keeping at health facilities.

**Photograph vaccination cards and health facility registers.** It is essential to record the dates from health records accurately, in order to draw strong conclusions about the timeliness and validity of vaccination. Data entry typing errors are more common for entering dates than for other types of survey responses. Digital cameras are inexpensive now, and smartphones are increasingly available, having the added advantage of geographical positional systems (GPS) capability, and we recommend that protocols for new surveys include a step of photographing cards and registers so dates can be verified during data cleaning. This will require some data management to track photo file names and associate them with the appropriate survey records. See section 3.4.5.

In summary, this manual aims to reduce the main sources of error in coverage surveys using methods shown in the table and detailed in the following chapters.

Table 1: Main potential sources of error and strategies to minimize them in immunization coverage surveys

|  |  |  |
| --- | --- | --- |
| Source of error | Effect of error on results | Strategies to minimise 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  If large populations have been excluded (e.g., security constraints at time of census), consider special efforts to include them  Be clear when writing report which populations may have been excluded and what the likely effect is on coverage |
| Selection bias  - sampling procedures | Non-probability sampling may lead to bias in either direction | Use probability sampling method  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 caretakers and HBRs  Do not substitute households |
| Information bias  - Lack of HBR or poorly filled HBR | May under- or over-estimate coverage depending on how missing data are handled and how HBRs are read by enumerators | Consider publicising reminders about HBRs prior to survey  Allow time for mothers to look for HBR, revisit if necessary  Include questions as to condition of HBR and checks for errors  Seek health facility-based records on children without HBR or with poorly filled HBR |
| Information bias  - Inaccurate verbal history | Caretakers may forget how many doses have been received or may over-report if feel pressure to say they have been vaccinated | 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 tetanus toxoid, ask careful questions about all doses received in previous and current pregnancies and in campaigns |
| Data transcription and data entry errors | May increase data classed as missing  Can bias coverage results | Conduct close supervision  Photograph vaccination records  Conduct range and consistency checks while enumerators can revisit household if necessary to correct data |
| Missing data | If non-random, biases result, often upwards | Conduct high-quality planning, training and supervision  Include appropriate statistical adjustment for missing data |

Table published in: Cutts FT, Izurieta HS, 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. Table doi:10.1371/journal.pmed.1001404.t002

# Design the sample structure of the survey

The purpose of this chapter is to explain how to design a vaccination coverage survey. It includes recommendations and instructions on identifying primary survey questions, setting inferential goals, identifying an appropriate survey design, calculating a sample size, estimating a budget and timeline, and deciding whether the survey is affordable and timely. The steps are illustrated in Figure 1.

Figure 1. Early steps in survey design

Identify primary questions (section 2.3)

Translate questions to inferential goals (section 2.5)

Select a study design to meet the goals and calculate sample size

(sections 2.6 and 2.7)

Estimate budget and timeline (section 2.9)

No? Revise

Affordable & Timely?

Yes

Begin Survey Planning

## Convene a survey steering group

Forming a task force or steering group will help coordinate the complex task of designing and conducting the survey. Representatives may be solicited from the host country’s national ministry of health, national census agency, WHO, UNICEF, the funding agency, and other partners. Ideally, some members should have experience with past vaccination surveys in the area so the group can customize the survey to the local context, and anticipate and address the country’s unique challenges. Because this revised manual relies on more rigorous statistical design and inference than earlier versions did, it will also be helpful for the steering group to secure technical assistance from a sampling statistician in the early stages of the work.

## Discuss the purpose of the survey

The goal of the survey design process is to establish consensus about the primary programmatic questions the survey is designed to answer, and to set realistic goals for and an achievable approach to answering those questions.

Surveys can be expensive and time-consuming, so check existing information and data first to see if a new survey is truly necessary. If you decide to spend the time and money to do a survey, follow the steps in this revised manual to ensure that it your survey is a useful and worthwhile investment.

The survey design process is iterative and often requires revising the primary questions and goals. The estimated sample size required to achieve your goals will inform the final decision on whether these goals can be achieved in an affordable and timely manner. Often, programme managers and donors start with ambitious and expensive survey goals, such as knowing the exact coverage in every district. Once they see the sample size and budget required, however, they may choose to redefine the questions. For example, they may change the goal from estimating coverage to classifying it at the district level, or they may just select a few districts where precise coverage estimates are needed (for example, those where major demographic or programmatic changes have occurred recently). They may decide to do separate surveys in these few districts in addition to a national survey, rather than trying to estimate coverage in all districts.

To illustrate these issues, this chapter focuses mainly on a simple scenario that addresses only one geographic level (stratum) and one outcome. The administrative or geographic levels include national, intermediate (called *province* throughout this guide), and peripheral levels (called *district* throughout this guide). For the purposes of this guide, a district probably has 10,000+ population. The end of the chapter contains recommendations for addressing multiple questions and different levels.

## Identify primary questions that affect survey design and sample size

The first step in designing a survey is to decide which questions the survey results will answer. It is helpful to identify one primary question and use the material in Annexes B1, B2, and B3 to determine the survey sample size. The survey will usually address several other secondary goals such as assessing dropout rates, validity and timeliness of doses, missed opportunities for vaccination, or reasons for not being fully vaccinated, but in most cases you will not use these questions to determine the sample size (see Chapter 6).

There are three major types of primary questions. An *estimation* question is a descriptive question that will result in a quantitative estimate of coverage and related estimates. Comparative or *hypothesis testing* questions compare coverage with an important programmatic threshold or across time, or between populations or geographic strata, or between levels of other characteristics like sex, education, or wealth. Finally, *classification* questions yield qualitative coverage labels (for example, “not high” or “not low”) instead of precise quantitative estimates.

### Descriptive or estimation questions

Here are some common **descriptive or estimation** questions, which lead to a quantitative estimate of vaccination coverage:

* What is the target population coverage by a vaccine-dose combination (for example, DTPCV1, DTPCV2, and DTPCV3) [[2]](#footnote-3)?
* What proportion of the target population is fully vaccinated according to the national schedule[[3]](#footnote-4)?
* What proportion of the target population was vaccinated during an SIA (also known as a *vaccination* *campaign*)?
* What proportion, or how many, of the individuals vaccinated during the SIA had never been vaccinated with those vaccines before?
* What proportion of children born in the last 12 months were protected at birth against tetanus?

### Comparative or hypothesis-testing questions

Comparative or hypothesis-testing questions such as the ones below allow you to compare coverage over time, or between sexes, populations, geographic strata, etc.:

* Has coverage for a vaccine improved since the last survey measurement?
* Is there evidence that coverage (routine and/or SIA) differs between provinces or districts?[[4]](#footnote-5)
* Is there evidence that coverage (routine and/or SIA) in one sub-population is higher than another (for example, boys vs. girls, those with uneducated mothers vs. those with educated mothers, indigenous vs. non-indigenous)?
* Are survey results consistent with the administrative coverage estimate (for example, within ± 5 percentage points of the administrative estimate)?

### Classification Questions

Questions such as the ones below may be used to produce qualitative labels like “high”, “moderate” or “low” to classify coverage for either routine vaccination or post-SIA surveys:

* Which health districts have coverage that is *below* an important programmatic threshold (for example, DTPCV3 coverage below 80%)?
* Which health districts have coverage that is *above* an important threshold?
* Which health districts have estimated coverage so close to the threshold that the survey does not tell us with 95% confidence whether it is above or below the threshold?

## Define the target population

To clarify the primary questions, it is important to specify the eligibility criteria for the population you plan to survey. For evaluations of routine vaccination coverage, target populations are defined in 12-month groups to represent the births in a one-year period – an annual birth cohort.

Use the following criteria to define the population for most routine vaccination coverage surveys:

* children aged 12–23 months, if the final primary vaccination is at 9 months of age – this is the most commonly chosen target population;
* children aged 24–35 months, if the age recommended for the vaccination (for example, MCV2, DTPCV4) is between 12–23 months of age;
* women who gave birth in the last 12 months[[5]](#footnote-6) (whether the child survived or not), if evaluating tetanus (Td or TT) coverage among pregnant women and whether their children were protected against neonatal tetanus at birth; and
* girls aged 14 years (and not yet 15), if evaluating HPV vaccine in a country where HPV vaccine is recommended for girls 9–13 years old. This age range may need to be adapted according to the vaccination schedule in each individual country.

For evaluation of SIA coverage, remember that the age group targeted by the SIA is sometimes stratified to provide precise estimates within subgroups (for example, <5 year-olds, 5–9 year-olds, 10­–14 year-olds, etc. for an measles-rubella (MR) SIA).

## Set inferential goals

Once you have identified the survey’s primary questions, you are ready to set inferential goals. An inferential goal states how much uncertainty is acceptable in the primary outcome.

In general, the more certain you need the outcome of the survey to be, the more respondents you will need (larger sample size), and the more expensive the survey will be. In an extreme case, a census of all eligible children would reveal vaccination coverage at the national, province, and district levels very precisely. A full census would be very expensive and impractical; to reduce the survey costs, we commonly assess vaccination status in a representative sample of children and accept some uncertainty in the results.

Uncertainty and inferential goals are described in different ways depending on the primary survey question.

* When *estimating* coverage, the inferential goal is expressed as a *confidence interval (CI).* Select a sample size that balances precision (typically represented with the 95% confidence interval) with the budget and time required to survey large numbers of respondents. For example, you might estimate the proportion of children who are fully immunized by one year of age, with the 95% CI no wider than ± 5% if the coverage is 70% or higher.
* When *comparing* two coverage estimates using a formal *hypothesis test*, the inferential goal is expressed as *statistical power.* The design and sample size are the result of a compromise between the ability to find a difference of a programmatically relevant magnitude (statistical power) and the available budget or time. Statistical power is usually characterized by three parameters:

1. The minimum detectable difference between two groups, or between a fixed threshold and the survey sample
2. The probability of making a Type I error, usually named α (*alpha)*. This refers to the probability that the hypothesis test will declare the difference to be statistically significant when in truth there is no underlying difference.
3. The power of the test, which is the probability that the hypothesis test will find a statistically significant difference given that the difference exists in the population quantities. Power is often expressed as 1­ – β (*beta).* See Annex B3 for more detail.

For example, to assess whether national coverage has improved since the last survey, you might conduct a 1-sided hypothesis test, setting α to 5% and yielding at least 80% power (β= 20%), to detect an improvement in coverage if the true difference has increased by 10% or more.

* Finally, when *classifying* coverage, the inferential goal is expressed using the probability of *classification error* (often called *misclassification*). The sample sizes usually compromise between the likely rates of misclassification and the available budget and time. In this case, define the thresholds against which the province or district is classified, and then set upper bounds on the probabilities of classification errors. See Annex B2 for more detail. For example, if you want to classify SIA coverage as low or high, and low means under 90%, then you might specify that the probability that any particular district with actual SIA coverage truly above 90% is misclassified as low should be 5% or smaller. That is, there is a less than 5% chance of so-called *failing* a district that has coverage above 90%. Likewise, the probability that any district with actual SIA coverage truly below 80% is misclassified as high should be 10% or smaller.

## Select a survey design

Once you have identified your primary questions, determined eligibility criteria, and specified your inferential goals, you should be able to propose a cluster survey design, sample size, and analysis plan to meet those goals.

If you are planning a survey that requires multiple outcomes, populations, administrative regions, or geographic levels (national, province, district), it is strongly recommended that you consult with a sampling statistician. We provide some guidance for these situations at the end of this chapter, but such designs are complex and are most successful with a statistician’s assistance. In simpler situations, you should be able to use the tables in this document to identify a design and sample size to meet the goals of your survey.

### Survey design for estimating coverage

If the goal of your survey is to *estimate* coverage with a point estimate and confidence interval, even at the lowest level of the health system under study, you will need a fairly large sample size. Figure 2 shows that surveys for precise estimation in each stratum are based on larger samples with more clusters, compared to surveys designed only to classify at the lowest level of geographic stratum. The sample size tables in Annex B1 will help you establish the number of respondents and clusters required.

### Survey design for classifying coverage

When the survey’s goal is to *classify* coverage, you may be able to use smaller sample sizes than you would need for precise estimates or powerful hypothesis tests. This can lead to substantial cost savings, but be sure that classification is all that is required, because this design may not yield a precise quantitative estimate of coverage at the lowest geographic level of the health system under study (districts, for example). Keep in mind that you will still need a minimum of 15 clusters in each stratum (such as a district) for a classification survey.

If designed properly, small surveys that classify coverage at the lowest level under study may combine the data from the lowest levels to *estimate* coverage at the higher levels (such as province and national levels). This design can be cost effective, and the estimates at the aggregated levels are often quite precise.

Figure 2. Precise estimation uses larger sample sizes than classification

Is estimation   
(coverage estimate with CI) required at the lowest level under study   
(e.g. districts)?

Use a relatively large cluster sample (30+ clusters per stratum ) for descriptive or comparative results. Use Table B-1 in Annex B1 to calculate an effective sample size.

Yes

No

Use a smaller cluster survey to classify at the lowest level and aggregate up to estimate at higher levels. Be sure you have at least 15 clusters per stratum. Use Table B-2 or B-3 in Annex B1 to calculate an effective sample size.

## Calculate the required sample size

To budget the survey accurately, you must calculate a sample size that will yield a dataset that meets the inferential goals. Annexes B1, B2, and B3 describe the parameters needed to calculate sample sizes. Work with the annexes or a sampling statistician to select a sample size (number of clusters and target number of respondents per cluster).

If you plan to report precise survey results in several demographic subgroups, you must ensure that there are a sufficient number of respondents in each group. When a subgroup is comparatively small in the population it is sometimes necessary to *oversample* members of that group, purposefully interviewing more members of that group than might have appeared randomly in the sample. The respondents are still selected in a random fashion so their results are representative of the subgroup population, but the sampling plan takes special measures to draw more respondents from areas where that subgroup lives. The precision of subgroup coverage estimates is determined by the subgroup sample size. When a survey oversamples some groups, their survey weights are specifically adjusted so their responses represent the appropriate proportion in calculations that combine subgroups. If it is important to obtain precise coverage estimates for demographic subgroups in your survey, work with a statistician to develop an appropriate sampling plan.

### Sample size for estimating, classifying, or comparing coverage

For surveys of several non-overlapping geographical areas such as provinces or districts, where coverage will be assessed in each stratum, it is traditional to conduct what is essentially a separate survey in each stratum. The stratum-level results are often combined to estimate an aggregated coverage figure. For example, the steering group may wish to estimate coverage in each province in a country to within ± 5%, and also to combine the provincial figures to obtain a national coverage estimate with even more precision. See section 2.13 (near the end of this chapter) for specific advice regarding surveys conducted in numerous geographic areas at once.

Whether the goal is estimation of coverage with a confidence interval, or classification of coverage with respect to a threshold, a certain number of households must be visited to yield enough eligible, cooperative respondents to meet the survey’s inferential goals. This number is calculated by identifying a set of five numbers to multiply together: A x B x C x D x E. These parameters are explained below, with detailed descriptions in Annexes B1, B2, and B3.

1. Identify the number of strata, as defined in a table or required by the steering group, in which you will repeat the survey.
2. Use a table to identify the base sample size per stratum (the *effective sample size)* – this is the sample size that would be needed if a simple random sample were used.
3. Use a table to identify the likely design effect (DEFF), which is a multiplier required because this is a cluster survey and vaccination status is likely to be spatially correlated. Earlier survey guidelines have assumed a design effect of 2 when you lack a recent estimate from a similar survey in your country. Annex B1 shows how to estimate design effect using Table C; it suggests being conservative and selecting a higher value to make it likely to meet the inferential goals in strata where coverage varies substantially from area to area and cluster to cluster.
4. Estimate the average number of households you’ll need to visit to yield the desired sample size. This will depend on the demographics of the survey target population as well as the birth rate and average household size in the country. It may vary between rural and urban areas.
5. Use a table to identify a multiplier that accounts for expected non-response due to persons not being at home after at least two revisits, or eligible persons who refuse to participate.

For classifying coverage, there are additional parameters relating to the thresholds being examined (for example, probably below 90% or probably above 80%) and the probability of classification errors. Annex B2 describes each of these parameters.

Similar calculations are used to calculate sample sizes to for comparing coverage, for example, between provinces, over time, or, for example, a comparison of HPV coverage among girls who do and do not attend school. For surveys comparing coverage, you will also need to specify the parameters for power and statistical significance.

Use Annexes B1, B2, and B3 to guide your selection of figures to multiply together. The next section discusses some of the common parameters used to calculate the sample size required to meet the survey’s inferential goals.

### Common parameters for sample size calculations

The calculations for each inferential goal require certain parameters. Gather these numbers, or estimate them, before you do the calculations. This section briefly describes the main parameters; additional definitions and details are in Annex A and Annexes B1, B2, and B3.

* **Target population size:** If the sample size turns out to be >10% of the target population then it will be worthwhile to apply a finite population correction to the sample size calculation and to the estimation equations. The details are not described here. Contact a sampling statistician for assistance.
* **Anticipated vaccination coverage (p):** The steering group will often have an idea of what coverage levels the survey will find, and those expectations can affect sample size. For a fixed level of precision or statistical power, larger sample sizes are required if the expected coverage is near 50%, while smaller sample sizes will suffice if the coverage is expected to be near 0% or 100%. This parameter may vary for different strata if the steering committee has sufficient information about the expected coverage in each stratum.
* **Intracluster correlation coefficient (ICC)**: This is a measure of correlation of responses within clusters. This number affects the design effect (DEFF) and therefore affects the sample size calculation. Usually, you will not know this number in the planning stage, so you can use an observed figure from a recent survey in the study area. Alternately, you can use a conservative value that is slightly larger than what is likely to be observed in the field, to increase the likelihood that the results will have acceptable precision. Annex B1 gives some guidance on selecting ICC values.
* **Confidence level (α):** This is usually 5%. The confidence intervals for estimation will be (100-α) %, or usually 95%.
* **Confidence interval (CI) half width**: This measures the precision of a coverage estimate. If the (100-α) % CI should be no wider than ± 5% (for example, CI = (52%, 62%)), this value will be 5%. The more precise the estimate, the narrower the CI will be, and a larger sample will be required. If less precision is acceptable, the CI will be wider and the required sample size will be smaller.[[6]](#footnote-7)
* **Target number of respondents per cluster (*m*)**: This parameter is usually selected to fall between 5 and 15, and is based on the number of households a data collection team can visit in a day as well as the total number of target respondents expected in an average size cluster, assuming that all eligible respondents in the those households visited are interviewed. We call this figure a *target* because we cannot know precisely how many eligible respondents will be found in each cluster. The number of completed questionnaires will vary from cluster to cluster, and the average number of eligible respondents per cluster will hopefully be ≥ *m*.
* **Target number of clusters per stratum:** The total sample size divided by *m* yields the target number of clusters per stratum. This number is fixed at the time the sample size is selected, and the clusters are selected randomly.
* **Parameters relating to the statistical power of the test and the probability of errors.** Annex B3 describes each of these parameters.

The next section provides a few examples of how to set these parameters.

### Examples of calculating a sample size

**Example 1: National level coverage only**

If the steering group wishes to estimate national-level coverage with confidence intervals no wider than ± 10% when coverage is at 50%, then the tables in Annex B1 indicate that the numbers for A x B x C x D x E should be as follows:

1. Number of strata = 1 (national estimate only)
2. Effective sample size = 103 (Annex B1, Table B-1)
3. Assume we will collect data from an average of *m=*7 respondents per cluster and assume an intracluster correlation coefficient of 1/3, so the design effect will be 3. (Annex B1, Table C)
4. Assume that an eligible child will be found in an average of 20% of the homes visited, based on the estimated number of households with children in the target age, so we must visit an average of 5 homes per eligible child.
5. Assume that 10% of families with eligible children will either not be at home when the survey team visits, or will refuse to participate in the survey, so we inflate the sample size by 11% to account for likely non-response. (Annex B1, Table E)

These values can be combined to calculate several quantities that are important for planning and budgeting purposes:

1. Estimated total target respondents with completed questionnaires: target = A x B x C = (1)(103)(3) = 309. The actual number will vary because different clusters will yield different numbers of eligible respondents.
2. Total households to visit to yield approximately 309 completed questionnaires:   
   (A x B x C) x D x E = (309)(5)(1.11) = 1,715
3. Number of clusters = = = 44.1. Round up to 45.
4. Number of households to visit per cluster = D x E x m = (5)(1.11)(7) = 38.85. Round up to 40.

In this example, the survey calls for 45 clusters—census enumeration areas (EAs)—to be randomly selected across the country. If EAs are likely to hold substantially more than 40 households, then the EA can be divided (using detailed maps) into segments that each hold about 40 households and a single segment can be randomly selected (see Annex E).

This selection is done before the data collectors go to the field. The team planning the survey logistics will either use quality satellite maps or will make a planning trip to each cluster. In either case, they will draw an excellent map of the cluster and its boundaries. After selecting one random segment they will prepare a map for the field data collectors to use, showing the boundaries of the selected segment very clearly. Field data collectors later visit the clusters and visit every household inside the cluster (or segment) boundaries, taking data from all eligible respondents. The number of completed interviews per cluster will vary because the team is not doing a quota sample but instead interviewing every eligible respondent in the pre-selected segment. On average, the survey should yield about seven completed surveys per cluster. Planners can decide whether a team can do all the work in a cluster in a single day, or whether it is more realistic to plan two days of work per cluster, accounting for the need to revisit households where no one is at home during the first interview attempt. The planners can also decide how many people make up a data collection team and how many teams one supervisor can effectively serve. These factors all affect the estimated budget for the survey.

**Example 2: National and provincial coverage**

Now assume that the steering group wishes to estimate routine vaccination coverage *in each province* as well as at the national level. In a country with five provinces, this essentially involves conducting essentially five separate surveys, and then combining the results in a weighted fashion to estimate national level coverage. Suppose the steering group wishes to estimate coverage in each province with confidence intervals that are no wider than ± 5% when coverage is at 50% in each province. The tables in Annex B1 yield the following:

1. Number of strata = 5 (one survey in each province)
2. Effective sample size = 401 (Annex B1, Table B-1)
3. Assume we will collect data from an average of *m=*7 respondents per cluster, and assume an intracluster correlation coefficient of 1/3, so the design effect will be 3. (Annex B1, Table C)
4. Assume that an eligible child will be found in an average of 20% of the homes visited, so we must visit an average of 5 homes per eligible child.
5. Assume that 10% of families with eligible children will either not be at home when the survey team visits or will refuse to participate in the survey, so inflate the sample size by 11% to account for likely non-response. (Annex B1, Table E)

These values can be combined to calculate several quantities that are important for planning and budgeting purposes:

1. Total target respondents with completed questionnaires: target = A x B x C = (5)(401)(3) = 6,015. The actual number will vary because different clusters will yield different numbers of eligible respondents.
2. Total households to visit to yield an average of 6,015 completed questionnaires:   
   (A x B x C) x D x E = (6,015)(5)(1.11) = 33,384
3. Target households to visit in each province: B x C x D x E = (401)(3)(5)(1.11) = 6,677
4. Number of clusters per stratum = = = 172
5. Number of households to visit per cluster = D x E x m = (5)(1.11)(7) = 38.85. Round up to 40.
6. Total clusters in the survey = = = 860

In this example, 172 clusters will be randomly selected per province. In each of those clusters, detailed maps will be used to decide how and whether to segment the cluster to identify a randomly selected contiguous group of 40 households. All 40 households per cluster will be visited and field data collectors will complete a questionnaire for each eligible respondent. The number of completed questionnaires will vary per cluster, but the average should be near 7. Separate weighted coverage figures will be calculated for each province, and then all the results may be combined in a weighted calculation to estimate national level coverage. The national coverage figures will be extremely precise, having a combined effective sample size of (401)(5) = 2,005.

Note that increasing precision from ± 10% in Example 1 to ± 5% in Example 2 increased the effective sample size from 103 to 401. It is costly to have an increased sample size to improve the precision. See Table B-2 in Annex B1 for additional detail on this point.

**Example 3: Imprecise estimation for classification at the province level**

In Example 2, it would be quite expensive to achieve an effective sample size of 401 per province. Upon reflection, the steering group may decide that they do not strictly need ± 5% precision everywhere, but rather they want to clarify which provinces have very high coverage, which ones have very low coverage, and which are likely to have coverage in between.

For example, if an important programmatic threshold for DPTCV3 is 80%, the steering group may wish to identify which provinces have coverage that is clearly higher than 80%, clearly lower than 80%, or likely to be near 80%. This is a classification goal; Tables B-2 and B-3 in Annex B1 are relevant here for calculating the effective sample size (parameter B).

This manual suggests using one-sided confidence bounds to classify coverage. Select a sample size for each stratum, conduct the survey, and calculate confidence bounds. The classification rules are as follows:

1. If the one-sided 95% lower confidence bound is above the threshold, classify coverage as being very likely to fall above the threshold.
2. If the one-sided 95% upper confidence bound is below the threshold, classify coverage as being very likely to fall below the threshold.
3. If the upper and lower one-sided bounds fall on either side of the threshold, one above and one below, conclude that the sample size was too small to classify the coverage as being above or below 80% with 95% confidence.

This last result might be disappointing: you have spent a substantial amount of money and effort to collect data just to find that the classification result is inconclusive. To avoid this situation, you would have to select a sample size large enough to yield conclusive results for your survey’s threshold. To classify which strata are likely to have coverage above or below 80%, the study designer selects a distance from the threshold, called *delta*, and uses Tables B-2 or B-3 to look up a sample size that will guarantee a suitably high probability that the one-sided confidence bound will fall on the correct side of the threshold.

This affects the required sample size dramatically. If coverage in a stratum is very high (for example, 95%), then a survey with an effective sample size as low as 45 will yield a sample where the one-sided 95% lower confidence bound is very likely to fall above the important threshold of 80%. However, the closer you get to 80%, the bigger the effective sample size will need to be. If the true coverage is 85%, you will need an effective sample size of about 250. If the true coverage is 81%, you will need an effective sample size of nearly 10,000 respondents to draw a confident conclusion that coverage is above the 80% threshold!

This process of classification is illustrated graphically in Annex N and in Figure 9 in section 6.5.2. There are sample sizes in Annex B3 to help draw strong conclusions for delta values of 1%, 5%, 10%, and 15%. Smaller delta values require much larger sample sizes to yield conclusive classification results.

The important point in this example is that programmatically useful classification can sometimes be achieved using smaller sample sizes than needed for precise estimation if the study designer is willing to accept classification #3 above (sample size not large enough to classify with 95% confidence) when the true coverage falls is within delta points of the programmatic threshold. Although it can be disappointing to have inconclusive classification results in some strata, there are three features that make the results programmatically valuable:

1. The graphic portrayal of the coverage results, as illustrated in Annex N, will sometimes make it clear that coverage is very likely to fall above or below the threshold, even when a conclusion may not be assigned 95% confidence. In other words, if one of the one-sided bounds is quite near the threshold, you may be able to confidently classify coverage, albeit with a confidence level slightly lower than 95%.
2. You will interpret the inconclusive results in the context of strata with conclusive results, so if some strata are classified as above the threshold, some below, and some inconclusive, then you know where the inconclusive strata fall compared with the others.
3. Finally, if the sample uses nested strata, like sampling from all provinces in a nation, the results from conclusive and inconclusive strata alike will be aggregated together to estimate and classify coverage quite precisely at the national level.

**Example 4: HPV coverage among 12-year-old girls**

In this example the steering group is evaluating coverage with Human Papilloma Virus (HPV) vaccine among girls aged 12 in a single province. If the vaccine is administered through location-based methods, possibly at schools, then the survey might have several goals:

1. Estimate coverage among those most likely to benefit from the vaccine administration strategy – girls who are enrolled in school and regularly attend.
2. Estimate coverage among the overall population of girls who need the vaccine – girls who are a particular age (for example, 12), regardless of whether they attend school.

The first goal might evaluate the success of the delivery strategy while the second goal evaluates the likely population protection in a cohort defined by age.

If the inferential goal is to estimate coverage of 2 or more HPV doses among girls age 12 with precision no worse than ± 5% if coverage is 75%, then the tables in Annex B1 yield the following:

1. Number of strata = 1 (a single survey in a single province)
2. Effective sample size = 340 (Annex B1, Table B-1)
3. Assume we will collect data from an average of *m=*10 respondents per cluster and assume an intracluster correlation coefficient of 1/6, so the design effect will be 2.5. (Annex B1, Table C)
4. Assume that an eligible child will be found in an average of 10% of the homes visited, so we must visit an average of 10 homes per eligible girl.
5. Assume that 10% of families with eligible children will either not be at home when the survey team visits or will refuse to participate in the survey, so inflate the sample size by 11% to account for likely non-response. (Annex B1, Table E)

These values can be combined to calculate several quantities that are important for planning and budgeting purposes:

1. Total target respondents with completed questionnaires: target = A x B x C = (1)(340)(2.5)= 850. The actual number will vary because different clusters will yield different numbers of eligible respondents.
2. Total households to visit to yield an average of 6,015 completed questionnaires:   
   (A x B x C) x D x E = (850)(10)(1.11) = 9,435
3. Number of clusters = = = 85
4. Number of households to visit per cluster = D x E x *m* = (10)(1.11)(10) = 111

In this example, 85 clusters will be randomly selected in the province. In each cluster, detailed maps will be used to decide how and whether to segment the cluster to identify a randomly selected contiguous group of 111 households. All 111 households per cluster will be visited, and data collectors will complete a questionnaire for each girl who is 12 years old. The number of completed questionnaires will vary per cluster, but the average should be near ten. If the estimated coverage is not lower than 75% and if the ICC is not higher than the assumed 1/6, then the confidence interval should be no wider than ± 5%.

Note again here that the relatively large sample size is driven by the requirement for narrow precision. If the steering committee were willing to accept precision of ± 7%, then Table B-1 indicates that the effective sample size would drop from 340 to 182 and the number of clusters would drop from 85 to 46.

If the vaccine delivery strategy was school-based and if school attendance was not 100% among 12-year-old girls, then a portion of the sampled girls would be unschooled and less likely to have been vaccinated. So the estimated coverage among the survey population would be a mix of the coverage among schoolgirls and coverage among unschooled girls. This would tend to yield a lower coverage estimate, which might be appropriate for evaluating population level protection, but would likely underestimate coverage among schoolgirls only. To guarantee a high precision estimate of coverage among schoolgirls it would be necessary to either restrict the survey sample to schoolgirls, or to oversample schoolgirls. If these approaches were pursued in a household survey then it would likely be necessary to visit more than 10 households to find each eligible respondent and so the planning would need to account for additional effort.

## Draft an analysis plan, table shells and report figures

At this stage in the planning process, it is helpful to draft an analysis plan and lay out table shells for the final survey report. This will help you budget realistically for the analysis portion of the project, and will also confirm whether the survey design will meet the programmatic goals of the survey stakeholders. See Chapter 6 and Annex Q for examples.

## Budget for the survey and estimate the timeline

Next, create a budget for the survey design you’ve selected. As you budget money and time, consider all aspects listed in this manual. Consult Table 2 for a list of activities and items to include in the budget. In addition to budgeting the monetary cost of the survey, make an estimate of the project timeline, accounting realistically for likely delays. Remember that your top priority is to ensure high data quality. To do this, you should have only as many field teams as can realistically be well supervised. It is better to use a small number of field teams and take longer to implement the survey than to have so many field teams that their training and supervision suffers, and data quality is compromised.

In addition to fixed costs, the cost of cluster surveys is proportional to the number of strata, the number of clusters per stratum, and the total number of respondents. Be sure to include all items with a cost that depends upon the questions, goals, sampling design, or sample size. See the DHS and MICS budget templates at <http://dhsprogram.com/publications/publication-dhsm10-dhs-questionnaires-and-manuals.cfm> and <http://mics.unicef.org/tools> for examples.

The timetable should likewise be adjusted according to the specific needs of the survey and, especially, the local administrative procedures required. Often, it takes additional time to access funds, choose a contractor to do the survey (if one is used) and gain ethical clearance from the relevant organizations.

For post-SIA surveys, it is best to conduct fieldwork very soon after the campaign in order to have a chance of seeing finger marks indicating vaccination or retrieving any SIA-specific cards that were given to caretakers. Thus, it is important to prepare for the survey well in advance, ideally at the same time you prepare for SIA implementation. The training and fieldwork for a post-SIA survey can be shorter than the timeline in Table 2 suggests, if data are only needed at the national level and only vaccines administered in the SIA are assessed. If the steering group requests data at the province or district level, and especially if they also request these data on all routine vaccinations, the survey becomes much larger, and it will probably not be completed quickly after the SIA ends. In order to ensure high-quality data, if results are needed quickly, it is better to compromise on the goals of the survey than to add too many field teams.

Table 2. Timeframe for a national coverage survey[[7]](#footnote-8)

|  |  |  |
| --- | --- | --- |
| Stage | Activity | Timeline |
| Planning and survey preparation | Form a steering group and technical subcommittees; identify the implementing agency; agree on methods to recruit field coordinators, supervisors, and interviewers; agree on whether data will be recorded using paper forms or digital technology; identify technical assistance if required; set up liaison with census office; order and obtain supplies; and identify transport. | Months 1–4  (may take longer if an RFP is issued for selection of an implementing agency, or if the survey has a complex survey design with multiple indicators, depending on ethics committee procedures and timetable, and depending on time needed to make funding available). |
| Design survey and modify/compromise design and modification/compromise to fit resource availability |
| Obtain funding for the survey |
| Obtain ethical approval as required |
| Select a sample (including obtaining enumeration area maps) |
| Visit health authorities in the areas selected for the survey, to explain survey and obtain co-operation |
| Design, pretest and translate the questionnaire |
| Prepare digital entry procedures, if used |
| Pretest household sampling procedures (use of enumeration area maps, identification of boundaries, segmentation, one- or two-step process of listing and interviewing), |
| Prepare manuals/ standard operating procedures (SOPs) |
| Prepare training site(s) and materials |
| Prepare database |
| Training | Train field workers and supervisors on household listing, collection of GPS coordinates, conducting interviews, getting data from health facilities, checking completed questionnaires, digital data entry where relevant, ensuring SOPs are followed and taking photos of vaccination records | Month 5 (longer for large surveys; allow two weeks for every 40 field staff being recruited) |
| Train data entry staff if paper forms are used |
| Data collection | Create maps and household lists  Collect data from eligible persons (listing and interviewing may be a one- or two-step process, depending on survey design)  Do quality control in the field  Resolve queries | Months 6 (if small survey), or 6–8 (for survey with multiple domains or strata); length depends on size of survey, travel time, ability to ensure high quality data collection) |
| Data management and analysis | Data entry and editing (if paper forms used) | Months 6–7 (small survey) or 6–9 (large survey); data entry begins concurrently with data collection and continues after last data comes from field) |
| Final data checking and cleaning |
| Data analysis, produce tables and graphs |
| Report generation and dissemination | Prepare/review preliminary report | Months 10–12 |
| Prepare final report, with summary of key findings |
| Conduct national feedback seminar, review final report, and develop action plan based on findings |
| Prepare reports/fact sheets for health workers |
| Workshops with health workers at subnational levels |

## Evaluate affordability and timeliness

If the proposed design is affordable and the results are likely to be available in the timeframe needed, you can begin to do more specific planning, as described in Chapter 3.

If the design is not affordable or if it would take too long, either appropriate more money for the survey or modify some combination of questions, strata, and inferential goals to find a lower-cost design that still addresses the steering group’s primary questions, with an acceptable level of uncertainty and in an acceptable timeframe. See below for examples of compromise strategies.

If the designs that are affordable do not adequately address the primary programmatic questions, the steering group should seriously consider **not** doing a survey at this time, and instead use other methods to assess and strengthen vaccination services.

If it is not possible to appropriate more money to conduct a large survey that meets the initial goals of the survey steering group, but some sort of survey is still desirable, the design team must compromise on one or more parameters to find a less expensive survey that still yields helpful results. These parameters may be varied to reduce the cost of the survey.

1. Adjust the number of geographic strata in which conclusions will be reported. If the steering group wants results in all districts but the cost is too high, it might be affordable to do a survey in each province instead, and give up the goal, for example, of having precise district-level results.
2. Adjust the survey goals in different strata. For example, you might estimate SIA coverage at the province level but assess routine vaccination coverage at the national level only. Since the target age group for SIA coverage is much wider than for routine immunization, sample sizes are reached by visiting a smaller number of households for SIA than for RI coverage.
3. Adjust the desired precision of the coverage estimates in each stratum.
4. Classify rather than estimate coverage at the lowest geographic hierarchy level. Rather than calculating a narrow confidence interval in each district, it may often suffice to use a smaller sample to classify coverage in each district, and aggregate data across districts to estimate coverage precisely at the province and national levels. The smaller sample will identify districts that are doing very poorly and those that are doing very well. There is likely to be a middle category of districts that are not clearly doing either poorly or well. In order to identify their current coverage precisely, a larger survey would be needed, but at least the small survey identifies that they are neither at the top nor the bottom of the performance continuum. When three or more strata are aggregated up to the next level of hierarchy, the confidence intervals typically become substantially more narrow and informative.

For example, assume a country has 10 provinces, each having between 15 and 25 districts, for a national total of 203 districts. The steering group may initially wish to estimate coverage in all districts with ± 5% precision. Average national coverage of DTPCV3 is thought to be 85%, varying from 55% to 95% between districts. To estimate coverage to ± 5% when it is only 55%, using a design effect of 4 requires 1,600 completed interviews.[[8]](#footnote-9) At 10 completed questionnaires per cluster, this would require 160 clusters per district. Repeating this in 203 districts would require visiting 32,480 clusters and collecting data from 324,800 respondents! This is prohibitively expensive, and would take a very long time to implement while ensuring high quality. Below are options for revising the survey goals.

1. Estimate coverage at national level and in a small number of key districts (such as those thought to have particularly poor administrative data, those where major recent programmatic or demographic changes occurred, or major metropolitan areas).
2. Classify coverage in all districts and aggregate data to estimate coverage at provincial and national levels. Classification might be achieved using 15 or 20 clusters per district. This would require a total of (203 x 15) = 3,045 clusters, covering all districts of the country. Although the total sample size will be smaller than when coverage is estimated in all districts, there are still important logistical considerations for getting well trained and supervised teams to this many clusters.
3. Estimate coverage precisely at only the provincial and national levels, using, for example, 160 clusters per province. This requires a total of (10 x 160) = 1,600 clusters, which may not necessarily include all districts if some districts have very small populations. The precision of coverage estimates at the provincial level could also be varied, to determine the effect on budget and time.
4. Estimate coverage imprecisely at the provincial level, for example, using 30 clusters per province and aggregating to estimate coverage at the national level. This means visiting only (10 x 30) = 300 clusters – a substantially smaller sample size than the other options. It will yield, however, imprecise estimates at the provincial level. This will be useful for identifying (classifying) provinces that are clearly low or clearly high, but not useful for making fine distinctions between provinces whose coverage levels are nearly equal.

The options here fit a wide range of budgets, ranging from 32,480 clusters down to 300 clusters. The larger options yield precise district level estimates and the smallest option yields precise estimates only at the national level, while providing some insight into which provinces are performing best or worst.

## Implications of adding routine immunization questions to a post-SIA survey

It is increasingly common for survey stakeholders to consider adding questions about routine immunization (RI) to a survey designed to evaluate SIA coverage. Planners may reason that substantial resources are already being devoted to planning and conducting a nationally representative survey, and believe those resources should be leveraged to assess the performance of the RI services while the survey staffs are already in the field to collect data. It seems reasonable, but an RI survey can require a much larger field effort than a post-SIA survey does. Sorting out what is best in each situation will require careful consideration, to strike a balance between a lean and timely SIA coverage estimate and a precise, geographically specific, multi-vaccine assessment of RI services.

Whether it is feasible and affordable to bundle RI questions with an SIA survey will depend on the inferential goals of both surveys. The best time to work through these issues is long before the actual SIA begins.

These are the considerations that may substantially expand the resources required when adding RI questions to an SIA survey.

* The window of age-eligibility is very small for RI surveys (usually a one-year window) compared with that for an SIA (often a 14-year window), so the survey staff must visit more households just to find an eligible respondent. If precise RI coverage estimates are desired, the number of homes to visit in each cluster will be multiplied by a large factor – possibly five or more. This is a substantial increase in cost and logistical complexity.
* The standard RI questionnaire takes much longer to complete than a post-SIA interview, so field staff will be able to complete substantially fewer interviews per day.
* RI coverage figures for important vaccines are often much lower than SIA coverage achieved, thus requiring a larger sample size to achieve the target precision.
* The intracluster correlation coefficient (ICC), which drives the design effect, will be substantially higher for RI vaccines than for that observed in a well-run SIA with consistently high coverage, so the RI design effect will increase the required sample size for precise estimation.
* It is a best practice in RI surveys to visit health facilities and obtain vaccination dates from EPI registers if the child’s caretaker cannot furnish a home-based record. This also represents a substantial commitment of time and resources.
* Finally, stakeholders may wish to estimate RI coverage in many more, smaller strata (such as health districts) than the people evaluating SIA coverage do. As described above, the overall sample size is proportional to the number of strata where you will report results, so this can increase the survey sample size.

If the idea to add an RI component occurs late in the SIA survey planning process, the extra planning and resources required could easily postpone the survey fieldwork for several months. A long delay will likely degrade the quality of SIA coverage responses and estimates, by increasing recall bias.

But if the goals of the SIA campaign are for precise estimation and the goals of the RI survey are less precise, and if the geographic or administrative focus is similar for both surveys, then it may be possible to add an RI component without much extra effort or delays. For example, it may be relatively easy to add an RI component if the RI survey requires results at a higher level of hierarchy (province level) than the SIA survey (district level). The key is to discuss it early, estimate the sample size and timeline realistically, and explore whether there is a design that does indeed leverage the SIA survey resources without compromising its goals.

## Designing for multiple outcomes

Sample size calculations are most straightforward when the survey steering group identifies a single primary goal to size the survey. When agreement cannot be reached on a primary goal, it is possible to do sample size calculations independently for two or more goals, and estimate a budget for the largest of the several various sample sizes.

If that design is affordable, it should be possible to meet several goals. If it is not affordable, some sort of compromise will be necessary. If the steering group intends to draw strong conclusions on several different outcomes simultaneously, it may be helpful to ask a sampling statistician whether some adjustment to the sample size is required to limit the increase in probability of error when conducting multiple simultaneous comparisons.

## Designing for multiple geographic areas

If you are planning to assess coverage in more than one geographic or administrative area, it may be necessary to calculate the sample size required in each area to estimate the budget for the survey. In some cases the sample sizes may vary considerably from one stratum to another, especially if the expected coverage outcomes vary substantially. Strata with coverage near 50% will require larger samples to obtain a given level of precision (for example, ± 5%) than strata with coverage near 0% or 100%. A simple shortcut may be to calculate the required sample size that is likely to be largest, and conduct surveys of that size in each stratum. You may save some money and time, however, by calculating sample sizes for each stratum individually, based on what is known about each stratum’s likely coverage outcome.

For example, using Table B-1 in Annex B1, to estimate coverage with ± 5% precision requires an effective sample size of 401 if coverage is in the range of 50%–70%, but only requires an effective sample size of 216 if coverage is near 90%. Substantial savings are potentially available by doing a smaller survey in locations with higher coverage. Of course, if you knew the coverage before doing the survey, you would not need to do a survey at all, so it is usually a good idea to select a conservative sample size in case coverage is closer to 50% than was originally anticipated.

## Designing for multiple levels of administrative or geographic hierarchy

Coverage surveys often assess coverage for several levels of a geographic or administrative hierarchy. The steering group may wish to estimate coverage within each province, and then aggregate results to estimate national coverage figures. In other situations coverage may be assessed at three levels. For example, the steering group may wish to identify all districts where SIA coverage is very likely to be below 95% and aggregate district surveys to estimate coverage in each province, with a confidence interval no wider than ± 5%, and then aggregate provincial results up to a national coverage figure with a confidence interval that is even more narrow, such as ± 3%.

In these cases, identify the level in the hierarchy with the most important inferential goal, identify a design and sample size that will meet that goal, and check to see whether the goals at other levels will be met as well. It is often the lowest level of the hierarchy (that is, those with the smallest geographic or administrative extent) where the survey results will be used to drive actions. The goals at that level are often the most important, with precision at higher levels being of secondary importance. If a design meets the goals at one level, but does not meet the inferential goals at another level, you will likely need to increase the sample size to move closer to satisfying goals at all levels. Balance this option against the budget and time implications of conducting a larger survey.

The tools in this manual should help survey teams identify designs that will meet goals at the most important level. In situations that are not complicated, it may also help them assess whether a single design will meet goals at multiple levels. For more complex scenarios, it will be helpful to enlist help from a sampling statistician.

**Example: Combining multiple outcomes and multiple levels of hierarchy**

Consider a measles campaign coverage survey in a country with 60 health districts, nested within ten provinces. Possible inferential goals might be:

1. Estimate campaign coverage nationally, without reporting subnational results (1 stratum);
2. Estimate campaign coverage in each province and nationally (10 strata; number of clusters depends on desired precision);
3. Classify coverage in each province and estimate national coverage precisely (10 strata; fewer clusters per province than in the previous design);
4. Estimate coverage in each district, and aggregate for provincial and national results (60 strata; number of clusters depends on desired precision);
5. Classify coverage in each district, and aggregate for provincial and national results (60 strata; fewer clusters per district than in the previous design).

Assume the steering group selects option 5. They will conduct a separate survey in each of 60 districts, using 15 clusters each and a target number of 10 completed interviews per cluster. The target age range group for the campaign is 9 months to 14 years, so they expect to find a cooperative, eligible respondent in every second household they visit, on average. That means visiting 20 randomly selected households per cluster, or 300 per district, or 18,000 nationwide. The number of expected completed interviews is 10 per cluster, 150 per district, 900 per province, and 9,000 nationally. In comparison to the classic 30 x 7 design, this survey is somewhat smaller, at 15 x 10. But it is being conducted in every district, so the overall effort and sample size is very large.

Now consider adding an RI component to the survey. The reasoning is logical: since the post-campaign survey will be nationally representative and survey workers will visit 18,000 homes across the land, why not also estimate RI coverage at the same time? If the sample size is fixed and you can find one child aged 12–23 months with a cooperative caretaker in every five households visited, the expected number of RI respondents per cluster is four, the expected number per district is 60, the expected number per province is 360, and the expected number nationally is 3,600.

Adding the RI component has numerous implications for survey logistics, data collection, data management, analysis and reporting, and cost and schedule. Each cluster’s work will take longer because you are adding an average of four RI interviews per cluster. Four RI interviews could turn into eight interviews if you also ask about tetanus vaccinations among women who gave birth in the last year. The supervision, training and data collection will be more complicated than for a simple post-campaign survey. The additional complexity of conducting three simultaneous surveys (SIA, 12–23 months for RI, and 0–11 months for tetanus) may tempt the survey organizers to collect primary data using handheld electronic devices. Will they photograph vaccination cards? Visit health facilities in search of documented evidence of vaccination? In some cases, the steering group may decide that the added insight is worth the cost of adding the RI component.

Careful consideration should be given at this point to whether adding the RI component will delay the start of the fieldwork and possibly compromise the quality of campaign-related responses. Finger marks from the SIA campaign may no longer be visible by the time the survey begins, or caretakers may lose their campaign-issued vaccination cards and forget or become confused about which of their children were vaccinated in the campaign. Furthermore, in countries with important seasonal migration due to weather, agriculture, and availability of work, a delay will give people time to move; some who were vaccinated will leave and some who were not will return. The survey results will reflect a combination of campaign effectiveness and population movement, which may be challenging to interpret.

Another consideration is the precision of the RI coverage estimates. With such a small number of RI respondents per cluster, the design effect would likely be small (maybe 1.5), so the effective sample size per province would be 240. Table B-1 in Annex B1 indicates that this is sufficient to yield precision of ± 6% at the province level if RI coverage were 75%. The effective sample size in each district would be = 40, which would result in coverage estimates that are quite imprecise. Table B-1 indicates that the confidence intervals would be wider than ± 10% when the effective sample size is 40.

If it is acceptable to classify SIA coverage in each district, and estimate it more precisely at the provincial and national levels, the sample size and data collection effort in each cluster and district may be manageable.

If the steering group in this example wanted precise RI estimates in every district, the sample size must increase, as described in section 2.11. To obtain an average of 10 RI respondents per cluster it would be necessary to visit 50 homes per cluster. This would increase the design effect to 2.5 and increase the effective sample size to 60 RI respondents per district, which Table B-1 indicates will still yield estimates with confidence intervals wider than ± 10%. A larger sample size and more clusters per district are necessary for more precise estimation. This will increase the level of effort for the survey and affect the survey start date, which in turn has consequences for the quality of the SIA survey results.

## Reporting results by subgroups

Survey stakeholders often wish to report coverage results by subgroups, such as sex or age groups (in a post-SIA survey), whether or not the child attends school (in an HPV survey), economic status, religion, or education of the caretaker. These comparisons may be so important that the study designers take steps to ensure a large enough sample size to estimate coverage precisely among those groups.

The subgroups may be listed explicitly as strata in the design phase, or the groups might be over-sampled (with respect to their prevalence in the population) to obtain precise results. If precise estimates are required for important subgroups, it is important to maintain this goal even when other goals are compromised or dropped. In the end, it will be numerically possible to report results by various subgroups, but those estimates will not be precise if the sample size is too small.

Designing the study specifically to report on some subgroups does not prevent you from calculating and reporting results based on other subgroups, but the survey designers and the survey report should be clear about for which groups the survey was intentionally designed to yield precise estimates. Some surveys use the guideline that results should only be reported for estimates or tests where the relative standard error (100 x standard error of the estimate/estimate) is no greater than 30% or where there are at least 12 *statistical degrees of freedom* (the number of clusters containing the subgroup minus the number of strata containing the subgroup) – see the Centers for Disease Control and Prevention’s NHANES tutorial[[9]](#footnote-10). When finalizing the survey design, it will be helpful to have the project statistician look over the analysis plan and identify any subgroups or comparisons that may be in danger of yielding imprecise estimates and to reconsider whether to report them.

# Make concrete plans

## Set survey schedule

Review the survey request and goals to determine the time constraints for the survey. When must the survey findings be available? Work backwards from this date, using Table 2 to determine how long the survey will take to complete. Keep in mind other potential deadlines, such as donor review or national budget sessions. Also keep in mind some of the factors that tend to delay the surveys, such as obtaining ethical clearance for the survey, obtaining access to accurate sampling frames and having all the resources in place.

Below are some other considerations to take into account when preparing a schedule for the survey.

* Avoid seasons with adverse weather conditions. Avoid the rainy season, the winter (in northern or southern countries), the hottest summer months, etc. as they may influence the physical accessibility of the households. The increased hardship on the survey workers may affect the reliability of the data collection. Difficulties in transportation may also translate into increased costs.
* Avoid religious and cultural events. For example, the month of Ramadan with its fasting may be hard on household members and survey workers, who may find it difficult to concentrate on the questionnaires. Also, during religious, political, and cultural events you are likely to find the population absent from their regular households, particularly in urban settings.
* Avoid, if possible, certain agricultural seasonal cycles. Rural people are either very busy or absent from their households during planting, harvest, migrant seasonal work, *jhum* (rotating cultivation on hills) or nomadic migration.
* Determine what time of day to do the survey. The survey should be timed to maximize chances of finding people at home. This may require early or late interviews during the day, to accommodate people who will be out for work during the day (including women, in urban slums or rural areas). Market days may also not be a good time to find parents at home. Conducting fieldwork during the weekend may find respondent at home but may conflict with data collectors’ weekly rest.
* For post-SIA surveys, start fieldwork no later than a month after the campaign to minimize the recall bias.

## Decide who will conduct the survey and create a project plan

Determine which organization is responsible for completing the survey. Academic institutions are often a reliable option. If a contractor is used, draft a detailed Terms of Reference document that clearly indicates the contractor’s responsibility for completing the steps described throughout in this manual. You may choose to contract out some of these tasks and not others. Determine how and when the expenses of the survey will be covered. Clarify with contractors how much money must be paid in advance and how much will be paid only upon receipt of proper deliverables (you may want to include penalties in case of significant delays).

Whether you are contracting out some of the project, or doing it all in-house, create a project plan that includes details detailed roles and responsibilities for the following tasks:

* obtaining ethical clearance
* gathering the necessary preliminary documentation, such as census data, maps, etc.
* designing data collection tools and methods
* choosing data analysis tools
* obtaining a sampling frame and selecting a sample
* obtaining vaccination registers
* hiring and training staff
* conducting fieldwork and ensuring quality data collection
* designing a database
* entering and cleaning data
* analysing data
* writing a report and sharing results.

## Obtain ethical clearance

The survey must be conducted in accordance with the national policies on ethics for surveys involving human subjects. Doing so typically requires an extra round of paperwork to explain and justify the study. Allow adequate time in the planning phase for this necessary – and often time-consuming – step.

If a national body exists to review the ethics of the study design of the study, the survey coordinator must obtain clearance from this body. For a standard survey, clearance should be a simple process. For a surveys using with biological samples, it may take longer to obtain clearance.

Most Institutional Review Boards (IRBs) will accept verbal informed consent for a standard coverage survey, which is relatively non-intrusive and does not seek sensitive information. Verbal informed consent has four elements:

1. a description of the objectives of the survey;
2. basic information on how the survey will be conducted;
3. assurances about the confidentiality of the results; and
4. a specific request for permission to conduct the interview, which can be obtained from each household by explaining, in detail and in the local language, the purposes of the survey.

Avoid making people respondents sign a consent form if at all possible. Many residents, particularly in rural areas, are wary of outsiders asking them to sign documents that might be confused with land deeds or taxes. Insisting on written consent has thus complicated the survey implementation in many communities. If, however, you are planning to collect biological samples, written consent will likely be required.

The Review Board will need to see a concrete description of how the confidentiality of the data will be preserved, how the individual identifying markers will be stripped and who will have access to what type of records.

## Design data collection tools and methods

### Vaccination data to collect

In order to standardize procedures across surveys, we recommend the following hierarchy of evidence of vaccination[[10]](#footnote-11) (see section 5.4.2).

1. **Home-based records** (vaccination cards). The best evidence is a legible date of vaccination on the home-based record (vaccination card) with a day, a month, and a year.
2. **Health centre records**. At times it will be necessary to check a child’s vaccination status in the health centre records (see section 3.7). There may be several obstacles to getting or using the data from the health centres: the record may not be legible; the record may have incomplete information, including date of birth; the child or his/her parents may have several different names; and registers may be only available only during short periods. However, you can overcome such obstacles by getting support from the local health authorities, identifying all relevant registers, photocopying all pages for the relevant time period before the time the household visits takes place, and assigning specific staff to review the records ideally within 24 hours of the household visits.
3. **Recall, or verbal history of vaccination**. If there is no home-based record of vaccination, or if it is incomplete, the next level of evidence is a verbal *history* of vaccination by the caretaker (vaccination recall). Start by asking the caretaker the place of the injection (on the body) for injectable vaccines, or act out putting drops in the mouth to ask about oral polio vaccine or rotavirus vaccines. Ask when the vaccine was received in relation to other documented vaccinations. Plan to use helpful visual aids matching the national vaccination practices when asking this question. Also ask the caretaker where the person went to receive the vaccination (for example, clinic, outreach site, hospital, school, home). A child might have been vaccinated in a health centre different from the nearest one. In such case it will not be possible to look for the record at the closest health centre. If a date is mentioned in the card it should be recorded, otherwise it should be considered as verbal history.

### Design forms

Although the WHO vaccination coverage survey manuals have proposed several standard survey forms over the years, the introduction of new vaccines and the specific needs of each new survey suggest that these templates need to be adjusted and new forms produced for each step of the survey. The forms listed below are the ones most likely to be needed. These different forms will be translated and back-translated as appropriate and finalized after the training and pilot tests.

* **List of Households** – In a single-stage survey every household in each cluster will be interviewed. It is important to make an updated sketch map (See Annex F) and to list every building or structure in the cluster, assigning an ID to each, and to list every household in each structure, identifying whether anyone in the household is eligible for your survey. The map will be important during data collection and it must be accurate and clear in case independent monitors follow along behind the survey workers to check their work in a small number of follow-up interviews. In a two-stage survey, every household will be identified on the sketch map and household list, and then a small number of households will be randomly selected to participate in the survey. Form household (HH) will serve as the sampling frame for household selection, and then interviewers will use Form HH and the sketch map to go back to the selected households. Note that a *household* is considered to be a collection of persons who usually eat food prepared from a single cooking area, or kitchen. In some countries, there will be several households contained within an extended family’s compound. Assign a separate household ID to each cooking area, even if the households are related, and record the appropriate ID on each interview form.
* **Household member listing form** – Form HM in Annex H is used to document who lives in each interviewed household, who is eligible for different components of the survey, whether they consent or refuse to participate, whether the appropriate respondent (the child’s caretaker or a woman who gave birth in last 12 months) was absent despite repeated visits to the household, and how many revisits were made. Several persons in the household may be eligible for different parts of the survey. At any visit, all, some or none of the appropriate respondents might be home, so the form should allow for a clear indication of interview status and of whether the team needs to return to the household again to complete its work.
* **Individual questionnaires** – Forms RI, TT, and SIA in Annex H serve as examples on which to record responses for a routine immunization survey, a tetanus protection-at-birth survey, or a post-vaccination campaign survey, respectively.
* **Health facility register forms** – Forms RIHC and TTHC in Annex H serve as examples on which to record data collected from a registry at the health facility.
* **Cluster forms** – Other forms may be designed and incorporated, as necessary, for summarizing data by cluster, such as total households, total completed interviews, and total completed survey questionnaires for each component of the survey (12–23 month, 0­–11 month, and post-SIA).
* **Forms or checklists** – These forms are for the field supervisors to record problems and progress.

Forms for collection of vaccination data should be designed to simplify data transcription from home-based records and minimize recording errors. For example, the order in which vaccines are listed on the questionnaire should match the order in which they are listed on home-based records. The “date of vaccination” fields should be big enough to allow for legible recording, so data entry operators can easily read the date. Enough space should be provided on the paper questionnaire to include relevant comments.

A note about finger marking as as evidence of vaccination: marking the child’s finger with an indelible pen during SIAs for measles, polio, maternal and neonatal tetanus, etc. is often used by vaccination teams for intracampaign monitoring purposes. These marks should **not** be used as the sole or even primary source of vaccination evidence in coverage surveys because it is rare for post-campaign surveys to be conducted soon enough before the markings fade away, and there are often issues with not enough pens/markers being distributed during the campaign.

### Design digital surveys, if applicable

Mobile devices (portable computers, tablets, PDAs, and smartphones) are ubiquitous and increasingly used for data collection whenever safe. The survey forms must be adapted for mobile devices if the survey will use digital data collection.

Sometimes the data entry into a mobile device is linked directly via data transmission to a central location for storage. The questionnaire templates are put on the telephones devices in advance and the data is entered in the field. Safeguards can be built in to discard obvious mistakes, like out-of-range dates of birth. Such data-based questionnaires require a software application to design the questionnaire templates, and a plan for safe and regular data back-up.

Using devices for direct data entry must allow the interviewer to check the entries for mistakes and correct them before the data is transmitted. The supervisor must be able to review the records each day, even when data collection is done and has been transmitted through mobile devices. In several countries, a list of the data entered during the day is sent back to the field every evening for corrections.

Digital data collection has benefits. Direct data entry eliminates the issue of bad handwriting. Using a smartphone allows access to the GPS coordinates of the house, which will help identify if a household is within the right geographical boundaries. In some cases, it will help the supervisor to identify a house that has to be re-checked. A smartphone can also document the time of entry and exit of each house.

It is likely that the use of computer-assisted data collection will expand rapidly, and survey planners should consult with experienced groups before using it for a given survey. This manual will be updated as computer-assisted data collection becomes more prevalent.

### Put individual IDs on forms

Every surveyed individual must be allocated a unique ID. This unique number links the household questionnaire, the photo of the card, and the photo/scan of the health centre record. The ID is made up of a sequence of numbers related to different type of information:

* cluster number (up to 4 digits)
* household number for that cluster (three digits; each interviewer is assigned in advance 99 numbers in advance, such as 0-99, 100-199, 200-299, etc.)
* child number for the household (usually one digit; maybe two digits in surveys of SIA coverage).

Each survey coordinator will structure the ID digits according to the survey’s specific needs. The cluster number will be known in advance and, based on the sample documentation, will show which administrative area it is in and whether it is urban or rural. Thus, individual IDs can often be pre-printed on the survey forms. If not, the ID should be handwritten legibly on a small white piece of paper to be used for photos.

### Plan to collect photos of evidence

Pictures of individual children are not needed; do not take them. However, taking a picture of the card and/or the health centre record for each child provides a reference document that serves multiple purposes. A paper-based data collection form may include a place for the photo of the vaccination card, showing also the household and child ID, and a photo/scan of the health centre record. If the data is collected on a smart phone or other mobile device, pictures of these documents may be attached to the interview form.

Photograph only the portion of the register or card that you need, focusing the camera close in. If the ethics committee requires it, cover up the child’s name to maintain confidentiality, using for example a self-stick label (Post-it®) on which the child’s unique questionnaire ID from the questionnaire is written. Save the photo using a file name with the same a unique ID number of the child, to help with later work associating digital photos with digital survey records. Record the filename(s) of the photo(s) on the child’s paper interview form.

Taking photos of evidence has several advantages. When dates are available on cards or health centre registers they are sometimes difficult to decipher and the recording might be incorrect. A photo offers the opportunity to re-examine the dates and possibly correct them, or check a date that is out of range in the database. A photo might be also be useful when a calendar other than the Gregorian calendar has been used; the dates are entered in the phone in the local calendar and automatically translated in to the Gregorian calendar. Looking at a photo of the card will show if the date error was actually written on the card (for example, sometimes people continue writing the previous year for the first several weeks of a new year), or if it was a transcription error. Finally, having a photo of a home-based record or vaccination card may help identify a child in the health centre register.

Collecting, storing, and managing these photos requires trained personnel and digital resources. There is some workload associated with managing the photos, possibly rotating them, cropping them, and in some cases manually renaming the photo files to ensure easy matching with survey respondents and records. If interview responses are collected electronically then the data collection software may include a robust and straightforward system for associating photos with suvey records. The protocol should be clear about how photos are managed, and the process should be pre-tested and practiced during staff training to set consistent and workable procedures.

Observe all the relevant national rules and restrictions concerning data privacy. Only authorized persons should have access to the digital photo files, and records. Only authorized persons should have access to the list that indicates which photos are associated with which survey respondents. Keep questionnaires and photos in separate directories to ensure the privacy of health information, and to prevent unauthorized persons from matching questionnaire records and with photos of cards or health centre records that may contain names.

### Pre-test survey forms and cluster maps

Before the survey begins, field supervisors or other senior survey staff should do 5–10 interviews to test the household listing form, to get a sense of whether the households have been listed correctly.

It is also important to test the reliability of the maps showing the clusters or segments. Before the survey begins, plan to visit at least one urban and one rural enumeration area that is **not** a part of the survey, to see if the maps are accurate. If the maps are not good and there are no better maps available, it may be necessary to create sketch maps (see section 3.6.3).

## Choose data analysis tools

Next, decide what program or tools you will use to analyse the survey data. To calculate coverage estimates and confidence intervals, you will need statistical software that accounts for the survey design and the survey weights because the surveys recommended in this manual are not self-weighting. In the absence of dedicated software for vaccination coverage survey analysis (like there once was with COSAS), data analysts have been using Stata, R, SAS, Epi Info, and SPSS or other software programs to analyse the data and produce the needed tables.

These programs work well as long as the parameters of analysis are clearly understood (the *missed opportunities for vaccination* analysis is often the least understood by programme managers—see section 6.4.1). Your chosen tools must also offer flexibility for specific analyses, like distribution of doses of a given vaccine over time or age-at-vaccination distribution by vaccine. The WHO intends to provide statistical programs and a user’s guide for countries and consultants to analyse survey data in a manner that is consistent with recommendations in this manual.

## Select a sample

In Chapter 2, we discussed how to select a sample design and sample size, including the number of clusters. Once these are set, you can select the sample for the survey.

Scientific probability sampling is the only way to achieve unbiased survey results. It also is the only methodology by which to estimate sampling error – the effect of interviewing only a portion instead of the whole population of interest. Sampling error measures how precise an estimate of the whole population the sample is. Features of probability sampling are summarized in Box 1.

Box 1. Features of a good probability sample survey

**Features of a good probability sample survey**

* **Uses a complete and recent sampling frame.** A *sampling frame* is a complete list of all *sampling units* that entirely covers the target population, such as a recent and well-conducted census. Any proposed frame should be evaluated to identify any gaps (for example, nomadic populations or homeless persons). If these gaps cannot be filled by preliminary work to update the census in certain areas, this should be well documented in the survey report as one of the limitations of the survey.
* **Uses accepted probability sampling methods** such as simple random sampling, systematic random sampling, or sampling with probability proportional to estimated size, at every stage of sample selection.
* **Selects a representative sample at the required geographic level(s),** such as national, stratified national, certain districts, etc.
* **If cluster sampling is used, includes an adequate number of clusters.** For a given total sample size, a large number of clusters with a small number of individuals in each is better than a few clusters with large numbers of individuals in each.
* **Ensures that the field implementation is faithful to the sample design.**
* **Ensures that the sample size is sufficient to achieve reliability and precision requirements.**
* **Is well documented to facilitate review and calculation of survey weights and non-response adjustments.**

### Using cluster sampling

For household surveys, cluster sampling is nearly always chosen instead of a simple random sample in order to reduce field costs and time. Clusters are selected from a sampling frame, which is a complete list of all sampling unitsthat entirely covers the target population. For a multi-stage survey, there should be a sampling frame for each stage of selection. The sampling unit for the first stage of selection is called the primary sampling unit (PSU); the sampling unit for the second stage of selection is called the secondary sampling unit(SSU), and so on. Desirable qualities of a PSU sampling frame are:

* + - it covers the entire population (exhaustive)
    - every household is in only one of the units (mutually exclusive)
    - its boundaries are well-defined
    - maps are available for every PSU that is selected
    - there is an estimate of population (preferably the target population or the number of households) for each PSU. This estimate will often need to be made using data on the number of households, the average household size and the birth rate.

The sampling frame may be a list of any geographic unit that has clearly defined boundaries, such as census enumeration areas (EAs), villages, gridded high-resolution satellite maps or urban neighbourhoods. The WHO recommends using EAs for the reasons described below.

* EAs are the smallest defined geographical units. Being small reduces the work of listing and sampling households within clusters.
* EAs are exhaustive and mutually exclusive. Isolated households might be missed if a listing of villages or towns were used, whereas every geographic area in the country should have been assigned to an EA. When all EAs are put together they should cover the whole country like a jigsaw puzzle, thus isolated households are less likely to be missed. During censuses, census officers develop sketch maps of EAs (and often more detailed maps as well) to show the boundaries. Most countries now also include GPS coordinates of EAs in their census data, making it easier to check the boundaries and also potentially allowing EA borders to be overlaid on satellite images such as Google Earth or others for segmentation (see section 3.6.3). By contrast, it is often unclear where the borders of villages, towns and urban neighbourhoods are, especially in regard to outlying homesteads and hamlets.
* In most instances, EAs are more consistent in size than villages, towns, and urban neighbourhoods, leading to a more constant workload per cluster than if a list of villages or towns were used for the sampling frame.
* Towns and urban neighbourhoods are often larger than the sampling interval used for the probability proportional to estimated size (PPES) systematic selection of clusters. Using a listing of EAs as the sampling frame should avoid this problem. Otherwise, you will have to divide all towns and neighbourhoods into separate PSUs that may be larger than the sampling interval before clusters are selected. This can be a lot of work. If you do not use EAs or divide the towns and neighbourhoods into units smaller than the sampling units, these towns and neighbourhoods will become *certainty units* (meaning that they are bound to be selected under PPES sampling), and will need to be treated like separate strata in the analysis.

In previous EPI surveys, it was thought that a self-weighted (unweighted) analysis could be done if clusters were selected using PPES, in which size was usually the (estimated) total population of the cluster. In reality, however, these samples were not self-weighting because the total population (all ages) was used for the estimate of size rather than the target population for the survey (for example, children aged 12–23 months), and because the figures on total population were often out of date. Because a perfectly complete, accurate, and up-to-date sampling frame for the target population is never available, WHO recommends conducting a weighted analysis.

In principle, a random sample of clusters could be selected for the survey. However, there are advantages to using a PPES sample: the larger population groups of a country (such as the capital city) are likely to be included in a PPES sample, whereas by chance they could be excluded in a simple random or uniform probability sample. Therefore, WHO continues to recommend PPES sampling methods, but recommends that data be gathered to allow a weighted analysis instead of assuming that PPES sampling makes self-weighted analyses valid.

### Determine if an existing sample is available

Designing, selecting, and implementing a proper probability sample from beginning to end can be a time-consuming and expensive process. Hence, survey planners often first look to see if there are existing samples that would be appropriate for an EPI survey.

Many countries have well-developed survey programmes through their national statistical offices or health ministries. It may be possible, therefore, to use an existing sample **if it is a valid probability sample and is available.** Often, agreementwill be needed from the survey sponsoring or implementing agency*.* Many countries use master samples developed from master sampling frames, from which subsets are selected for use in particular surveys. Explore that possibility for the EPI survey. There are various ways in which an existing sample may be used.

* Attach your survey questionnaire modules to the questionnaire for another survey. This is an option only if the other survey will be conducted within the prescribed time frame for your survey, and if its sample size is adequate for your needs.
* Work in the same EAs that were selected in a previous survey. You can do this if the survey was recent, was conducted well, and had an adequate number of EAs (see Chapter 2). This can save you having to obtain census data and maps from the census office, and can also help you budget your survey costs in detail ahead of time.
* Use the household lists that were done in the previous survey. You can do this if it meets the same conditions as above, and also had a thorough and well-conducted household listing stage prior to household selection. The drawbacks of this option are that the household listing can quickly become out of date, and household occupancy or composition may change in different seasons.

Existing samples that may be good candidates are the DHS, MICS and similar surveys. These surveys will undoubtedly be designed with a probability sample. You could use a recent sample, or you could work with planners of an upcoming DHS/MICS to determine how to improve the quality of the data on vaccinations (for example, by adding a review of health centre records). Evaluate whether their sample size is large enough for the required number of people in your target age group(s), and whether the number of PSUs and cluster sizes are within the ranges discussed in this manual.

Since DHS bases its sample size calculations on the number of women of reproductive age required for its primary purposes, there are often fewer than four children aged 12–23 months per cluster. A new, larger number of households would therefore be needed to estimate routine immunization coverage, and you will need to assess whether the PSUs in their sample have enough households to give, on average, the desired number of children per cluster (see Annex B1). For SIA coverage evaluation of a broader age group, the number of households in the DHS is likely to be adequate.

### If no suitable sample exists, develop a sample

It is important to work closely with the National Statistics Office to obtain the sampling frame, which is usually the most recent census with population projections where relevant. Also check if there have been DHS or MICS surveys since the census and whether those surveys updated the sampling frame. If so, even if you do not use their actual sample (as in the option above), it may be better to use their updated sampling frame than to use the census, unless there were any areas excluded at the time of the DHS/MICS which have since become accessible.

Some areas that were included in the census or in previous surveys may have to be excluded from the current survey for security reasons or occasionally for climatic reasons (for example, if part of the country had been recently flooded). Be sure to identify these areas as much as possible before taking the sample, and document them carefully both in the survey protocol and when reporting survey findings.

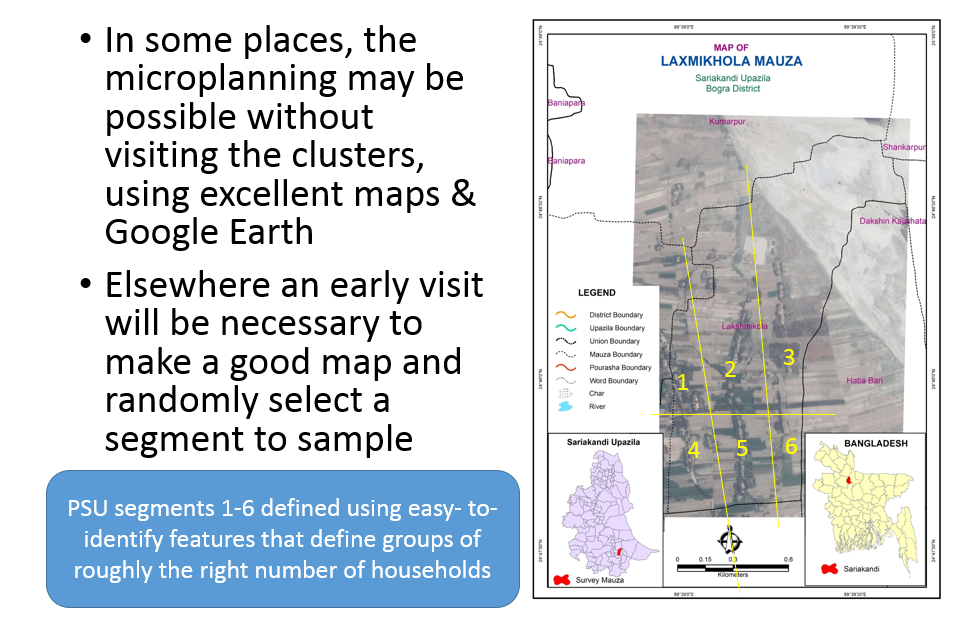
Below are the steps for selecting clusters.

1. **Obtain a sampling frame of EAs for the most recent census, where available**. Invest the time and effort to obtain the cooperation of the census office so you have access to the census spreadsheets. Also, learn how to use their EA sketch maps and GPS coordinates, if available. It is often beneficial to include a fieldworker with census experience in on each survey team. If a list of EAs cannot be obtained from the census bureau or national statistical office after exhaustive efforts, there are alternative sampling frames that can be considered under exceptional circumstances (DHS Sampling and Household Listing Manual, 2012a). It is important that whatever administrative unit is used, its boundaries can be clearly and objectively identified in the field, so you can easily segment the PSU if necessary (see below) and select individuals in each cluster.
2. **Evaluate the sampling frame for population coverage, distribution, identification and coding, as well as size and consistency (DHS Sampling and Household Listing Manual, ICF International, 2012a)**. Carefully document whether any areas were excluded for any reason. Also document any major changes that are thought to have occurred since the census was conducted, such as population movements due to major construction like dams. The WHO recognizes that in most instances, the sampling frames will not be up-to date, or the population estimate will be for the entire population instead of the target population. It is impractical to update sampling frames for a vaccination coverage survey, but using an existing sampling frame is adequate for calculating survey weights based on the probability that the PSU is selected into the survey sample (see section 6.2).
3. **If implicit urban-rural stratification is desired, sort the file of EAs and their respective populations by urban and rural.**
4. **In one column, show the census population count, or the number of households in each of the EAs**. This is the measure of size of each EA.
5. **If any EA is small and likely to have fewer households than the target per PSU (see Annex B1), combine it with a geographically contiguous neighbour so that they form a single entry in the sampling list or frame**. Update the associated population to be the sum of the two individual EA populations. For example, if sample size calculations show that on average, 10 children aged 12–23 months must be included in each cluster and that based on the population demographics (birth rate and average household size), on average eight households must be visited to find one child aged 12–23 months, then PSUs should have at least 80 households. If an EA has fewer than 80 households, it should be combined on the list with its nearest neighbour, and the resulting combination should be treated as one cluster for the subsequent steps.
6. **The sample size calculations will have specified how many clusters (denoted as *n*) must be selected for the survey**. This should be done for each stratum if the survey is stratified.
7. **In each stratum, a sample of *n* EAs is then selected independently using systematic sampling with replacement, with probability proportional to the estimated size**. None of the EAs will be too small because the small ones were aggregated in step 5. The selected EAs that have fewer than two times the target number of households are the clusters for your sample.
8. **Divide any large selected EAs (having many more households than are needed) from the sample list into segments that are estimated to have (a) at least the target number of households per cluster and (b) no more than two times the target number of households per cluster (see Annex E for more details)**. The census office can usually provide a map (aerial photograph, digital map, or hand-drawn) called a *sketch map*. It shows landmarks within the EA, the location of the boundaries, streets within the EA (if there are any), and where households are concentrated within the EA (especially for rural areas). If EAs are geo-referenced, Google Earth images can be used as a substitute, often with more detailed information (see Box 2). If there are no existing maps or the maps are of poor quality or seem incomplete, a mapping team must locate the EA to draw a sketch map to create segments. Sketch maps for segmentation purposes need only to show dwellings and not each individual household, and thus can be completed relatively quickly – as little as half a day (see Annex E). To segment urban areas, it is almost always preferable to use a mapping team rather than aerial or satellite photos. Using a random number table or a computer program, randomly select one segment in each of these large EAs; these segments plus the EAs that were of an appropriate size and did not need segmenting (number 7 above) are the clusters for your survey.

The worked example in Annex D shows you how to combine systematic PPESsampling with geographic arrangement of the sampling frame to achieve *implicit* stratification by urban/rural residence.

For special population such as refugees or internally displaced persons (IDPs), vaccination coverage survey recommendations will vary. In stabilized situations, it will be important to include refugees and IDP populations into the national coverage surveys. In the immediate displacement context of these populations with a dynamic population movement, insecurity issues, and different population size than typically used for standard EPI surveys, other survey guidance may be used. In addition, maps and an up-to-date list of residents may not be available during an emergency setting. For guidance regarding planning of a vaccination coverage survey in immediate emergencies, please refer to the forthcoming 2015 United Nations High Commissioner for Refugees (UNHCR) vaccination survey reference manual.

Box 2. Using Google Earth to segment large enumeration areas



### List households in selected clusters

Depending on the survey goals, target age group(s), length of the individual questionnaires, and local demographics, survey planners may choose a single-stage design or two-stage design. In a single-stage design, all eligible children in the selected clusters are enrolled to participate in the survey. A two-stage design has an initial phase of household listing and random household selection, followed by a repeat visit to interview parents of eligible children.

For evaluation of routine vaccination coverage using a relatively short questionnaire (for example, one that does not have many extra questions on knowledge, attitudes and practice, or indicators related to other health programmes), a single-stage approach in which all eligible children in the selected clusters are enrolled is often more efficient than a two-stage approach. For example, if the survey requires 10 children aged 12–23 months per cluster, and the local birth rate is 30/1000 population and average household size is 5 persons, on average the number of households needed to enrol 10 children would be = 66.7. If an enumeration area (EA) contains on average 100 households, the average number of children expected to be found in that EA would be = 15 children. It is likely to be more efficient to enrol all eligible children in the EA, by visiting all households and enrolling eligible children immediately, than to have a first step of listing households, selecting 67 households randomly, revisiting those 67 households, and enrolling eligible children at the revisit. On the other hand, if a long questionnaire is to be administered to eligible individuals (that is, one that takes an hour or more to complete), it may take less time overall to use a two-stage approach so you are not interviewing more individuals than necessary. In surveys such as post-MR campaign surveys, which enrol persons of a wide age range and an eligible person is found in every one or two households visited, a single-stage approach would result in a sample size much higher than needed, and a two-stage approach may be preferred.

Whether a single-stage or two-stage design is used, a household listing step is essential (see Annex F). In each cluster, survey teams list each structure on a listing form, noting which are inhabited and which are not (for example, schools and offices). See Form HH in Annex H for an example. Interviews are conducted at each household to record the names of the heads of the households and the household composition on this form. In a single-stage design, this is done concurrently with enrolling eligible persons, and it facilitates quality control of the completeness of the fieldwork and provides the data needed for the weighted analysis. In a two-stage design, the listing enables the coordinator to select a random sample of households for field teams to visit at the second stage to collect vaccination data. In both designs, up to two revisits should be conducted, if needed, to obtain all the information for all eligible persons.

If no one is at home, it may be possible to ask neighbours whether any eligible respondents live in the household. Form HH lists a field to indicate whether the information about the household comes from a resident or a neighbour.

Regardless of whether the survey uses a single-stage or two-stage design, the outcome of all visits to households in the survey sample must be carefully documented. Children in households with an available and willing respondent may be more likely to have been vaccinated than those in households with unavailable or uncooperative respondents, so careful accounting for missing data is needed to reduce bias in coverage estimates. Form HM in Annex H lists a space for a disposition code (interview outcome code) for three visits to every eligible respondent in the household.

### Collect data on vaccination from eligible persons in each household selected for the sample

In single-stage designs, all households in the cluster are screened and all eligible persons are included in the sample. It is essential to visit all dwellings, list all households, and enrol all eligible individuals in the cluster, no matter what the estimate target number of respondents per cluster originally was in sample size calculations. In two-stage designs, a random sample or systematic random sample of households within the cluster is pre-selected and the list given to field teams. All households on this list are visited, and if an eligible person lives there or has spent the previous night there, a vaccination questionnaire is completed.

Up to two revisits should be done as necessary to complete vaccination questionnaires as fully and accurately as possible. If a respondent is not present at the first visit, do up to two more visits to meet them. If a respondent (for example, the caretaker of a 12–23 month-old child) is present at the first visit but the home-based record is not available, then complete as much of the questionnaire as possible at the first visit but do up to two more visits to review the home-based record and complete the relevant section of the questionnaire.

To have enough time for high-quality household listing, enrolment of eligible persons and collection of complete and high-quality data, it is likely that more than one day will be needed in each cluster, whether a single-stage or two-stage design is used. Exceptions may be for post-SIA coverage evaluations for national-level coverage estimates of an SIA that targeted a wide age range (such as MR campaign of children aged 9 months to 14 years, or meningitis A campaigns of persons up to age 30 years). As few as five households may be needed in each cluster, and the questionnaires are short, so it may be possible to do the household listing and mapping in the morning and revisit the few selected households in the afternoon (for an example, see Meyer et al. 2015). Revisits of households where a respondent was absent at the first visit are still required, however, which might still require a second day of fieldwork in the cluster.

## Obtain access to health registers for vaccination records

You will likely need access to health facility records to check the vaccination status of some of the children, so it is wise to budget the additional time and resources necessary to do this. Plan to visit all health facilities that vaccinate children in the survey clusters to establish collaboration, gather early documentation (photocopies of the records), and assess the health register quality (legibility of the records).

Before fieldwork begins, obtain lists of the names of the EPI providers, health facilities, and outreach posts with their geographical catchment areas. It is best to obtain these lists from the district director of health or the EPI manager, whom the survey coordinator should visit anyway as a courtesy before teams go into the field. You should also ask local guides for the names and locations of vaccination places the local population uses. If it is common in your country for children to receive vaccinations from private providers, the steering committee may decide that data collectors should visit the private providers’ offices to obtain records for children whose home-based records are not available.

It will be necessary to search for evidence of vaccination status in health facility records if one or more children in the cluster have a caretaker who says that they received some routine vaccinations locally, and if:

* the caretaker does not show interviewers the vaccination card, or
* the card indicates some doses with a tick mark, but no date, or
* the caretaker says that the child received some routine doses that are not recorded on the card.

It is not cost-effective to run after each EPI provider and wait until the provider has finished the day’s work to access the registers. A more efficient strategy, where appropriate, may be to borrow the relevant registers for a couple of hours and photocopy them. To do this, it is best to request that the EPI manager bring the EPI providers together for a day in one place with all the registers. If you can obtain these photocopies in advance, the extractors can begin their work immediately, the day after the questionnaires have been filled.

Be aware that even the original health records could be hard to decipher and may require clarification from the original writer.

## Select and hire staff

Over time there has been a growing trend to subcontract vaccination coverage surveys to private or research institutions. If the survey is subcontracted, all requirements of the survey should be spelled out in detail in the request for proposal (RFP), and the submissions scrutinized for their exact adequacy adherence to the terms of reference.

One key staff person is the **survey coordinator**. The coordinator has authority over everyone involved in the survey, and works directly with whomever requested the survey.

The coordinator is responsible for:

* overseeing the implementation of the vaccination coverage survey
* ensuring the cooperation of other relevant government agencies
* making budget estimates for the survey before potential funding sources are identified for the survey
* selecting field teams
* overseeing the fieldwork
* reporting survey results
* overseeing training and pilot testing
* overseeing data entry and data management.

Whether directly hired by the survey coordinator or indirectly by a contractor, all types of workers who will be involved with the data collection and analysis must be identified and selected. The coordinator or the contractor must select people capable of working as members of a team and qualified to undertake their respective roles, as defined by the job description. The coordinator or the contractor should establish the required profile of each type of staff for the tasks they have to perform.

The data collection and analysis process have several consecutive steps, each involving its own team of skilled workers:

* data collection, with a team of field interviewers and supervisors (in two-stage cluster surveys, the data collection team may be subdivided into a household mapping/ and listing team and an interviewing team);
* data entry, with a team of computer operators, data cleaners, and supervisors;
* data analysis, with data analysts producing the tables already defined by the senior officials and partners requesting the survey.

Each step of the process requires a thorough data-checking process:

* in the field by the interviewers (checking each other’s forms for completeness and accuracy) and supervisors (checking forms, observing interviews, conducting repeat interviews);
* by the data entry staff at the time of the dataset is created (creating the data set by using a double entry process and by inserting limits for each field); and
* by the data analysts (checking the consistency and range of the data).

### Field staff

#### Regional coordinators

Regional coordinators are responsible for the fieldwork in one or more strata of the survey. They check the quality of maps and microplans. Similarly, they assist supervisors and interviewers to be able to find the appropriate clusters, communicate with each supervisor daily, and make others aware of progress and changes in plans. Regional coordinators also work to ensure consistent responses to unforeseen developments. Supervisors report daily progress to the regional coordinators, and in turn the regional coordinators report the progress up to the survey coordinator.

Regional coordinators also conduct quality checks by revisiting a portion of households already surveyed to verify that the household listing and interviews were conducted properly, that all eligible respondents in those households completed questionnaires, and that vaccination dates (and possibly other responses) were recorded correctly in homes where cards are available.

Interviewers should know ahead of time that a proportion of households will be revisited by regional coordinators, or by other independent monitors, but should not know which ones.

#### Field supervisors

Field supervisors have several roles. They must make sure that the fieldwork of their teams is performed according to standards. Although the supervisor cannot be with every team every moment, this person is expected to be in the field, observing the teams as much as possible. Field supervisors are also the first-line reference for clarification in case the interviewer has doubts. They must also flag inconsistencies in the questionnaires, and must fill out the activity tables at the end of each day and pass them to the survey coordinator.

Too often, supervisors are selected at the end of the training from among the brightest trainees. This may be inadequate for several reasons. Supervisors do not only need technical skills, but also the capacity to lead a fieldwork team, and to monitor and constructively correct poor practices of field interviewers. Also, specific training on these skills may have to be organized at the same time that interviewers are trained to interview.

Field supervisors are responsible for:

* ensuring the welfare and safety of the team
* ensuring each member of their teams is fluent with the questionnaires and techniques of the interview
* ensuring each member of the team has the necessary materials for their his or her daily activities
* overseeing the activity in the field
  + confirming that a verbal history of vaccination is obtained using the standardized approach, agreed upon during training, so that the language that does not bias the responses
  + confirming that adequate time is allowed for the respondent to look for all available home-based records
  + checking that field staff do not make transcription errors when copying down dates from the cards or health facility records
  + visiting every home in a sub-sample of clusters to confirm that each was visited and revisited if necessary.
* reviewing all forms before leaving the cluster (perhaps at the end of each day) for legibility, completeness, and errors accuracy, and the use of photos of cards when possible
* ensuring that completed data collection forms are given to those responsible for data processing in a timely fashion
* checking the quality of photos.

#### Interviewers

Interviewers work under the supervision and guidance of the field supervisor, and are responsible for collecting the data according to the instructions given in the data collection forms. They are accountable for the data they collect and the way they collect it.

Below are some things to consider when selecting interviewers.

* Interviewers should have a sufficient level of education (defined nationally), a pleasant personality tuned to local social customs, enough physical stamina to walk long distances under rain and sun under through sometimes difficult terrain at times, and a fluency in the language(s) spoken by the interviewees.
* Depending on the local customs, it may be necessary to have the correct mix of male and female members of the field teams. In some cultures interviews can only be conducted between people of the same sex, and in some cultures female interviewers must be accompanied by a male staff person.
* It may or may not be an advantage to hire field interviewers who have worked in on other surveys. Although they have demonstrated they can perform under field conditions, their previous experience may give them a false sense of confidence and weaken their capacity to pay attention to the specific requirements of the new survey.It can be beneficial to include an interviewer with census experience in each survey team, in case there is a need to do sketch maps.
* Interviewers must be able to write carefully and clearly, especially numbers.
* The survey coordinator or the contractor should avoid using health or EPI staff as interviewers when possible.
  + People associated with the vaccination services (local EPI staff) may unwittingly influence the way respondents reply to some questions, particularly those relating to reasons for not being vaccinated. However, people unfamiliar with the vaccination services may not naturally probe for important information on vaccination age, dates, and reasons for failure, and may also confuse dates that they see on a card (for example, the proposed return date with the actual date of vaccination). This is why it is important to train interviewers thoroughly on vaccination practices and on the rationale of the survey questions. In case no local candidates are found, the coordinator may consider hiring a health or vaccination staff member from another area, if the candidate can speak the local language.

#### Drivers and local guides

Drivers are responsible for the proper timing of the daily activities, and for the reliability and safety of the teams’ transportation to and from sites. This is a very important role, so drivers and guides must be made to feel part of the team and feel accountable for the timely conduct of the survey.

The selection of a local guide is not usually the responsibility of the coordinator. Usually the coordinator makes arrangements with authorities in the areas to be covered by the survey cluster or health facilities, to assign guides to the field teams.

The role of local guides is to:

* help field teams familiarize themselves with the clusters they are to survey
* introduce them to the cluster’s administrative and social authorities
* advise survey staff on when it is best to visit households
* introduce field teams at houses if requested by the interviewers.

#### Local guides should not be involved in deciding which dwellings to visit, or in interviewing and collecting data.

#### Observers

The coordinator may decide to include international or national participants or observers to enhance the confidence and objectivity in the results of the survey.

### Data management and analysis

#### Information communication technology specialist

Where data is collected using mobile devices, the information communication technology (ICT) specialist is a full-time position based at the central office. This person is responsible for receiving the daily data collection on the server, checking the coherence of the data, and returning any problematic data to the field that night to be checked and corrected the next day.

#### Data manager

The data manager is a full-time position worker responsible for designing the database structure and the data entry interface. This person verifies that all data (GPS coordinates, questionnaires from households and from health centre registers, photos of cards, etc.) have been sent daily, monitors the data checking process, and verifies that the monitoring tools have been filled out daily by the supervisors and given to the survey coordinator. Monitoring tools include the numbers of household visited, percentage of questionaires completed, percentage of children whose vaccine records were extracted from the health centre registers, etc.

The data manageralsomerges files from the data entry operators, and ensures that the paper forms are correctly archived and stored, copies of the data file are free of viruses, and the data file has been copied for backup purposes. Finally, the data manager is responsible for training and supervising the data entry operators.

#### Data entry operators

Depending on the number of computers available for data entry, more than one shift of data entry operators may be employed to complete data entry. When using double shifts, avoid inconsistencies by training all data entry operators and their managers uniformly, so all managers give the same answers to the same procedural questions. Data entry operators should be identified and trained shortly before data entry begins.

#### Statistician

The statistician contributes at several stages of the project, first working closely with the steering group and later working closely with the data manager. In early conversations about the survey goals, the statistician calculates sample sizes to meet the objectives identified by the steering group. Later, the statistician reviews the proposed questionnaires and works with the data manager to define the database design, design a codebook, and specify appropriate checks on valid ranges of values for survey responses.

The statistician also helps to evaluate candidate sampling frames for clusters, may conduct or help with cluster selection, and contributes to the microplanning protocol to be sure that microplanners save information that will be needed to calculate survey weights. Likewise, he or she works with the steering group to draft table shells and identify graphs needed for in the survey report.

When a sample dataset is available, the statistician also writes well-documented statistical code to check the dataset, identify unexpected data values, calculate derived variables, populate table shells, and generate graphical figures.

After the data is analysed, the statistician helps draft the methods, results, and strengths and limitations sections of the report, and works with other authors to be sure that results are interpreted clearly and correctly. The statistician populates individual variable summary tables in the final codebook and, when appropriate, makes both the dataset and analysis code available for checking by independent parties.

When digital ICTs are used to collect data in the field and upload it to a server, the statistician works with the data manager to create tools to summarize the data collected thus far, and to identify problems based on whether data are missing or have strange values, or whether the latitude and longitude of each team’s data are in the expected location of the clusters.

## Train staff

A good survey requires dedicated interviewers who have mastered the use of good tools. The acquisition of the needed skills is the result of the quality of the individual candidates and their training.

Final interviewer selection should take place at the end of the training session. Candidates with poor handwriting or those who still have an incomplete understanding of the forms should not be selected. Train more people than needed so you can select the best, and also have additional trained workers available in reserve in case several selected workers default (for sickness or other reason).

### Training time and number of trainees

Training should be given considerable attention and time. Do not rush the training, and be sure to confirm that the information presented is clearly understood by all trainees. In addition to training on the survey process and tools, supervisors need training in supervisory skills and in how to do field checks for data quality.

It may require multiple checks to ensure that the staff has acquired the necessary skills. Not doing so will jeopardize the quality of the results. This is why each instructor should be limited to 20 trainees, so the instructor can devote sufficient time and attention to each trainee. Having even fewer trainees per instructor may be even better.

A minimum of five days is generally required for the initial training, the field pilot test, the analysis of the pilot test data (to identify individual mistakes or the flaws in the instruments), direct feedback and potentially revising the tools. Enough time should be allocated to ensure that field staff understand how to identify the boundaries of the selected clusters or segments, how to do the household listing, and how to complete the individual questionnaires correctly. If there are several variations in vaccination cards or EPI register books in circulation, interviewers should learn to recognize and extract data from each type.

### Training topics and methods

Provide training on how to handle common problems with household-level data collection. Useful areas to address include:

* what to do with several households in a common dwelling
* how to define the date of birth if it is not clearly written in the card
* how to deal with incomplete or illegible dates or errors in the chronology of dates of birth and vaccinations
* how to document a vaccination history from the caretaker, and what to do if there are incomplete forms or absent cards, or if the caretakers are not present.

For training on using health centre records, focus on the most common problems:

* inability to access the records (staff out of station, records in the field, records already archived elsewhere, etc.)
* inability to locate the child from the records due to misspelling of the child’s or parent’s name
* inability to locate the child in the records due to registrations organized by day instead of alphabetically or by card serial sequence.

Training methods should be as practical as possible. Include instruction on a standard way to write numbers clearly (with handwriting exercises if paper forms are used), and on how to review incorrectly filled forms. Include role plays on how to do introductions, ascertain dates and assess likely vaccinations from caretakers. Close observations during training and the pilot field test will allow the trainer to give immediate feedback and corrective action. Such training is necessary even when field staff will use digital data entry.

Consider doing a video recording of the pilot field practices of the trainees (budget for this in advance). The day after the field pilot test, the parts of the videos documenting shortcomings or errors should be shown and discussed. A good technique is to let the trainees discuss what is wrong or could be improved in that section of the video.

# Conduct field work

## Collect data from households

### Visit all households selected for interviews

High quality data collection (including the revisit of households initially found vacant) will probably require field staff to spend more than one day in each cluster. Sometimes evening or early morning visits will be required. The interviewers may spend the night in the cluster if it is logistically possible and safe. In any case, all logistical arrangements should allow them to start early enough to find the children and their caretakers at home.

Sometimes evening (or early next morning) visits will be required. When clusters are located where most households have both parents working outside the home during the day, interviews may have to take place in the evening after caretakers have returned home. In these cases, local guides may be even more important to obtain access to houses. Evening visits may have to be done by male interviewers if security is a concern, or if cultural considerations require it.

The child does not need to be physically present at the time of the interview, but a caretaker or knowledgeable guardian, ideally someone who can supply the child’s vaccination card, must be present in order to proceed with the survey. If no knowledgeable caretaker or card is available during the first visit, a second and third visit in the evening or on the next day should take place before the interviewer leaves the cluster. Record the interview outcome for each visit using a disposition code on Form HM.

The total counts of eligible children in each cluster will be used to calculate survey weights. Putting zero or indicating a missing value for households would lead to an underestimation of the total. So, if no one is at home during the initial visit, some survey protocols will allow interviewers to ask neighbours how many survey-eligible children live in a household where no one is at home at the time of the initial visit, and to record this information on the household listing form (Form HH in Annex H).

### Conduct the interview

After introducing themselves and explaining the purpose of the survey, the interviewers should establish whether anyone in the home is eligible (spent the previous night in that household and is of the appropriate age group), and if so, obtain informed consent to administer the questionnaire. In most cases, ethical review boards will allow a protocol to use verbal informed consent only if the survey does not include taking biological samples. If the coverage survey is combined with a serosurvey, then the protocol may involve having an adult sign a consent form.

To ascertain the eligibility of a child it is necessary to identify his/her age and therefore the date of birth. This can be done from the vaccination card, or a birth certificate, if available. If a card is not available, the date of birth should be reconstructed from a calendar of local events (prepared during training): religious festivals, political events like elections, climatic events (monsoons, cold weather), etc. It might be time consuming but essential.

After obtaining consent to proceed, begin the interview, following the logical flow of the questionnaire. The availability of a vaccination card should be immediately assessed using the specific questions in the questionnaire (Have you ever been given a card? Do you still have it? Can you bring it?). It is vital to give the caretaker time to find the card, and to offer to return at a later time if necessary (for example, if the card is in a locked cupboard and the father has the key and will return later that day).

If a card is available, the interviewer should check the date of birth and available dates of vaccinations for legibility and consistency. The card should be interpreted according to the format used in the area. For example, sometimes a date written in pencil means the date the child should return for the next dose. The protocol should be clear about any local or national practices that could be confusing to survey staff.

If there are no written dates for a vaccination the child is eligible for, the caretaker should be asked for a history of that vaccination, using the national EPI body site for each injection, such as right arm or left leg, as a reference. The interviewer should also ask about the name of the place (health facility, outreach site, etc.) where each vaccination was received to facilitate with getting the health record from the health register.

### Refer unvaccinated children to the health centre

If the interviewer learns that a child in the household is overdue for a vaccine, he or she should recommend that the caretaker take the child to the health centre to receive the vaccine. Before the survey begins, ask the ministry of health to create a referral letter for this purpose, and give a copy to the child’s caretaker. Give the health centre a copy in advance so they are aware that the survey team may refer a small number of unvaccinated children over the age of 12 months. Also give a copy of the letter to the caretaker of any child that should be referred to the health centre for a vaccine.

### Check the completed questionnaire

Every completed questionnaire should be checked by the interviewer first, and later by the supervisor. Every question of the form should be filled in clearly and legibly. If one interview team member writes the dates, then before leaving the home, the other team member should check the form to verify the correctness and legibility of the dates. The dates on the questionnaire must match what was recorded on the vaccination card, even if the vaccination card has invalid dates. The data manager or field coordinator, not the interviews, must decide how to handle such dates. If the protocol includes taking a photo of the vaccination card, the photo should be checked for clarity. Additional photos should be taken, if necessary, to eliminate a bright glare, dark shadow, or blurriness.

The supervisor must verify every questionnaire for completeness, consistency, and legibility, and also evaluate photos for clarity and completeness. If there are errors in completing the questionnaire, the interviewer must correct them before leaving the cluster.

## Check health registers from the health centre

See section 3.7 for guidance on when it is appropriate to check health registers. If it is necessary to check the health registers, the first task is to find the child in the health register:

* Narrow the time period to match the month and year of birth with the month and year of the record pages;
* In case of a health record issued serially with the vaccination card, try to match them;
* Try to match the name of the village, hamlet, or administrative unit from the questionnaire with that on the register; and
* Try to match the name of the child as well as the names of the father and mother. Often people have two names (their official administrative name and their usual name), which makes matching difficult. In some cultures, very young infants do not receive a name until several weeks after birth.

After the child’s record is found in the health register, the team should look for a record of the vaccinations each surveyed child was *eligible* for, and record that information on a separate health centre form (Form RIHC in Annex H).

Eventually the survey will include up to three types of vaccination information for each child:

* vaccination history according to the card
* vaccination history according to the caretaker’s recall, for any vaccination not recorded on the card
* vaccination history according to the health registers.

Sometimes these sources will have discrepancies. The data collection field teams do not need to make any decisions in case of discrepancies, but simply to record verbatim what has been found. At a later stage, data analysts will address the discrepancies, carefully documenting each decision they make about editing data in the database and why.

## Monitor the quality of field data collection

A quality survey depends on the work done in the field. There are several potential sources for error in the data, and the interviewers, supervisors, and field coordinators have the primary responsibility for identifying and correcting errors in the initial collection and recording of the data.

### Re-check households with no eligible children

If the household listing form says that there is no eligible child in the household, check the household again to be sure.

### Check completed questionnaires

Responses from the child’s caretaker, the home-based record, or the health register may be missing, illegible, or in error. The interviewer may have misunderstood the child’s caretaker or misread the home-based record or health register. The interviewer may also have forgotten to enter the information or may have entered it incorrectly.

Each form should contain the following information, and supervisors in the field should check each questionnaire for these items:

* **Form number**: Each questionnaire should be assigned a unique form number to facilitate checking the data entered with a paper form.
* **Cluster number**: A cluster number should be entered for each form, because the data cannot be properly analysed if there is no cluster number. Ideally, clusters should be numbered from 1 to the total number of clusters. For example, if data are being collected for 30 clusters, clusters should be numbered from 1 through 30. Although census bureaus usually assign a much longer and more complicated identification (ID) to each EA to indicate province and district location as well as a rural/urban distinction, these long sequences of ID characters are subject to transcription errors and should be avoided in field paperwork. The survey data coordinator can maintain a list that matches the simple survey cluster number from each stratum (for example, 1–30) to the complete and specific EA number provided by the census bureau.
* **Household number**: The household number is a combination of structure ID and household serial number, as recorded on Form HH (see Annex H). The household number for each household in which an eligible child has been interviewed should be recorded to facilitate data checking. Since several interviewers are likely to be working in the same cluster, each interviewer should be assigned a set of structure numbers in advance (for example, 100–199, 300–399).
* **Household resident number**: A resident number must likewise be entered for each form so that the data can be properly analysed. Household resident numbers are assigned within households and range from 1 to the total number of residents in the household. Forms RI, TT, and PC in Annex H include a place to record the resident number for each child and caretaker from Form HM. It can be helpful to record the child’s first name as well. (Note: During data cleaning and management, each child will be assigned a unique ID in for the survey, consisting of a combination of stratum ID, cluster ID, household ID, and household resident number. It is not necessary to construct this unique number in the field.)
* **Child’s date of birth**: The child’s date of birth should be entered on the questionnaire and checked to ensure the date of birth is between the eligible dates of birth for the age cohort.
* **Date of interview**: The date the interview was conducted should be recorded and checked.
* **Dates of vaccinations**: Vaccination dates should be between the date of birth and the date of the interview. The answers to questions on dates of vaccination should be consistent with the response to the answers about the presence of a home-based record. If there is no home-based record, there should be no dates of vaccination, and instead, there should be answers on the caretaker’s verbal history of vaccination. For such children, vaccination dates will be sought in the health facility and recorded on a health facility form.
* **Home-based record (vaccination card)**: If the completed questionnaire indicates that there is no home-based record for an eligible child, check to make sure this is actually true.
* Other fields should have an entry within the range of acceptable entries.
* Finally, a field for comments about the interview is often useful, even when data is collected using mobile devices.

Each data collection form should have an entry for each field (unless some questions cause others to be skipped) and the responses should be legible. In general, in survey forms, text in lowercase represents what is to be read as part of the interview and text in uppercase represents text that is not to be read, such as instructions to interviewers. Each form should have a correct cluster and household resident number entered. Only those who meet the eligibility criteria should be included in the sample.

There are several levels of quality monitoring expected in the field:

1. Each *interviewer* is expected to submit only completed, legible, and accurate questionnaires. When there are teams of two interviewers, it is useful to have each worker check the other’s questionnaires after completion.
2. Every day, the *supervisor* must check every questionnaire for completeness, legibility, and accuracy. The supervisor checks that the household list indicates that questionnaires have been completed for all eligible children, and if not, there are reasons recorded for missing questionnaires (for example, caretaker not available after two visits or refused to participate). All forms must be checked and corrected **before** leaving the cluster area. The supervisor’s signature on the questionnaire confirms that this was done.
3. The *survey coordinator* or *contractor* is expected to organize a revisit of 10% (as an ideal) of all eligible children a day or two after they have been visited, to be sure that maps were followed correctly, cluster or segment boundaries were correctly identified, and that fieldworkers did not skip (either intentionally or by mistake) interviews for eligible children. Because the coordinator’s priority is to support the ongoing survey activities, it is not practical for him/her to do all of these revisits alone. Instead, the survey team should budget for a dedicated supervisor or two to be assigned to that task. Contractors may resist this provision, but it is a recommended practice. A 10% sample of clusters should also be revisited for repeat household listing, to check that the household lists have been done correctly and tally eligible respondents in each home. The children to be revisited can be selected randomly or not, as the coordinator may have doubts on specific questionnaires. When revisiting the households, the supervisor should ask the caretaker to repeat the interview for the sake of quality control, and compare the resulting questionnaire’s results with those of the interviewers.

Supervisors should give feedback immediately to interviewers about any discrepancies, correct the discrepancies, and discuss steps to improve the next day’s work. Any discrepancy or missing data should be resolved by discussions with the interviewers, a review of photographs of the vaccination card (if available), or revisits to households if necessary.

## Check questionnaire forms and transmit

The data collection team should count all questionnaire forms and verify them against the household lists. Once the questionnaires and health register forms have been checked by the supervisor, the supervisor should send them to the survey coordinator through safe channels as soon as possible, to be entered into a database. When data is collected digitally through smartphones or other portable devices, it is easier and faster to transmit the data to the survey coordinator than when paper forms are used.

## Clusters that become suddenly inaccessible

Early in the survey design process, the steering group may have excluded certain portions of the country from the sampling frame due to concerns about the safety of survey workers. Those parts of the country are not sampled, and the survey results will not be representative of vaccination coverage there. The portions of the country in the sampling frame will have a reasonable expectation of being safe and safely accessible at the time of the survey. As the survey commences, however, situations can change and some clusters may become unsafe due to nearby fighting or flared-up hostility toward vaccination and vaccination workers. Clusters may also become inaccessible due to problems like wildfires or flooding.

If the problem is temporary (for example, a flooded river that is expected to recede) and there is a reasonable expectation that safe access will be restored during the period of field data collection, every reasonable effort should be made to retain the originally selected cluster in the sample. This may require postponing data collection there and coming back toward the end of the survey. This is the most desirable outcome from the perspective of data integrity and representativeness. If the problem persists and there is no reasonable expectation of being able to collect data in that cluster as planned, then the survey steering group will need to determine whether to select a replacement cluster, and how the data analysis should account for the missing data from the originally selected cluster.

If the factor that made the cluster inaccessible to the survey team might also periodically make the same cluster inaccessible for vaccine delivery, that cluster might have especially low vaccination coverage and leaving it out of the survey might bias results upward. Some sensitivity analysis might be required to understand what the effect of finding low coverage in that cluster would have been. On the other hand, if the inaccessibility during the survey was clearly not related to anything that might have affected vaccination coverage there (for example, these were the first wildfires in the region in over five years), the steering group may decide simply to substitute another randomly selected cluster for the inaccessible cluster and skip the sensitivity analysis.

The safety of the survey personnel is of primary importance, of course, and decisions about survey operations should ensure as safe a working environment as possible. If some originally selected clusters are omitted or replaced during the fieldwork, then the survey report should document clearly what was done and indicate the reasons for omission, speculate about whether these causes might also affect vaccination coverage there, and document any appropriate sensitivity analyses.

# Data entry, cleaning and management

This chapter describes the steps necessary to prepare the data for analysis and summary table production. These steps assume that data have been recorded on paper forms and are being entered in a computer for consolidation, cleaning, and subsequent analysis. Some surveys have used digital data collection devices to record, store, and transmit data rather than paper forms.

## Database design

Design and test a database in advance of the survey completion. Develop the database structure, create data entry routines and entry range checks and complete consistency checks. The database structure should be complete and accurate, and tested with pilot data so that the development of the statistical analysis programs can begin as soon as possible. The data manager is responsible for designing the database structure and the data entry interface.

Construct a complete list of survey variables, known as a *data dictionary* or *codebook*, at the same time the database structure is established. Each variable will have a type (string or number), a label, and a set of valid values. Categorical variables should have clear, concise labels for each category. Responses like “Do not know” or “Refused to answer” should have well-defined values in the codebook and in the data entry software. After data has been collected, the codebook can be updated to include a brief summary of each variable in the dataset. Section 5.5 describes the components of a useful codebook.

In most cases, each child will be represented with one data record. If the survey collected data on more than one cohort of subjects (for example, a cohort of children 12–23 months of age surveyed for routine vaccination, and a second cohort of women who have given birth in the last year surveyed for tetanus toxoid coverage), it is advisable to have a separate database for each cohort. The data entry form should look as much like the paper data collection form as possible.

Data entry operatorsmay make errors when entering the data, such as entering the data inaccurately, not entering records (or entire forms) completely, or entering forms multiple times. The database should be designed to catch or prevent as many of these errors as possible, using appropriate filters and error checking. The software should accept only valid values for categorical variables and should provide a warning when data appear illogical (for example, the date of the second dose of the oral polio vaccine (OPV2) is earlier than that for OPV1).

## Data entry

Depending on the number of computers available for data entry, more than one shift of data entry operators may be employed to complete data entry. When using double shifts, care should be taken to avoid inconsistencies by training all data entry operators and supervisors uniformly, so all supervisors give the same answers to the same procedural questions. Data entry should take place in a separate room from other survey activities, where the staff is not disturbed and the questionnaires are secure. Each data entry operator should be assigned a unique staff ID number that they enter with every record so feedback can be given to the right people if data quality audits reveal too many data entry errors. To reduce data entry errors, have each data form entered independently by a second data entry operator, and then compare the two entries using computer software (see section 5.3).

Once all the data is entered, the data managermust merge files from the different data entry operators, and ensure that the paper forms are correctly archived and securely stored in a fireproof location that also ensures confidentiality. Only a limited number of survey staff members should have access to forms or photos that contain personally identifying information and they should be well trained on how to do their work without revealing the identify of participants to other people who do not need to know that information. The data manager should also check that copies of the data files are free of viruses, and should backup data files regularly. In some instances, it may be important to document and manage different versions of the master data file to ensure the correct version is being used, and many available software packages have methods to do this automatically.

## Clean the dataset

The data manager should work with the statistician to clean the dataset and create a series of checks for every variable in the dataset. The data cleaning step, when performed over all variables and all records, is time-consuming, but it is important to devote adequate resources for it. It is not sufficient to spot-check a subset of variables or a subset of records. Computer software should compare every variable and every record in the dataset, and all inconsistencies should be resolved before the data are summarized and analysed.

The data manager must have a plan for what to do when there are errors, and must follow the plan consistently. If the data management team changes any values in the dataset, the change should be documented in a data cleaning log. The change should be made using software, not by changing the value in the original dataset. This makes the changes reproducible and makes it possible to reverse the changes if they are later overruled. The software should include either comments or variables that capture the reasoning behind the decision to change a variable’s value. The sections below provide suggestions for handling different types of errors.

### Duplicate, missing, or conflicting data

The data manager should check for duplicate entries or forms that were not entered. When entries for one or more fields differ between the two versions entered, the data manager should refer to the original data collection forms (and, where relevant, photographs of home-based records or health facility registers) to determine which entries are correct.

### Implausible or illogical responses

The values should be checked to be sure they are plausible, and any logical relation should be checked to be sure the relation holds. Some examples: every vaccination date for a particular child should fall between that child’s birth date and the date of the interview, every record from a particular cluster should have been collected on the dates that the team visited that cluster, and every record from a single geographic stratum should have latitude and longitude values that fall within the boundaries of that stratum. The data manager should document any checks done for plausible values or logical relations.

The data manager should correct any implausible values found. Consult the original paper form, photograph of the vaccination card, or health facility record in case the problem occurred during data entry. If the unlikely or invalid value occurs on the original document as well, then the problem should be noted. When it is obvious what the correct value should be (for example, when dates fall in early January, it is common for people to continue to write the previous year), the value can be re-coded, but the decision to re-code responses from the value given to another valid value should be considered soberly. This serious action must be justified, documented clearly, applied consistently, and noted in the final report. If there is any ambiguity at all about the correct value, the safest course of action is often to set improbable values to “missing” and document that decision.

### Skip patterns

Some complicated forms use skip patterns, where one response on an earlier question causes the interviewer to skip later questions. For instance, in the verbal history portion of the questionnaire, if a caretaker says that a child has never received any vaccinations at all, the interviewer would skip the specific questions about BCG, OPV, etc. When data is collected digitally with a smartphone or other device, the skip logic is usually programmed and tested, and is performed automatically. But when data are collected using paper forms, it is common for interviewers to inadvertently ask questions they should have skipped, or fail to ask questions they should have asked.

Data checking should include a step to evaluate whether skip patterns were correctly observed. If a question should have been skipped but data was recorded and entered, change the response to “missing” and document the change.

## Merge datasets and construct derived variables

A forthcoming supplement to this manual will make very specific recommendations regarding how to code and name variables, in order to prepare them to be analysed in an open-source statistical software. This manual gives broad guidance, which may be made more specific by consulting with the statistician who will analyse the survey data.

### Merge datasets

After the data have been entered, cleaned, and checked, there may be some work necessary to merge data from different sources. Data collected in the respondents’ household may be held in a different dataset than that collected in health facilities, and these datasets may need to be merged to construct the master dataset for analysis. Photo file names may need to be merged or associated with individual survey records.

### Construct derived variables

The statistician will need to calculate a set of *derived variables*, new variables created using information about the sample design and the data collected in the survey. These variables help populate the table shells identified in the analysis plan that was developed during the survey planning stage. Table shells appear in Annex Q. Derived variables include indicator variables and the survey weights.

A set of derived variables will combine information from the home-based questionnaire and the records from a health facility to indicate whether a child received a particular dose. Different derived variables can code, as described below, whether the child’s coverage status is documented by (a) any source of evidence (card or register), or (b) documented source OR verbal history. These derived coverage variables will be summarized to estimate vaccination coverage in the survey target population.

If the survey collected data on questions with open-ended answers, in which the respondents’ words are recorded on the data collection form (for example, “Other, please specify:” or “If not, why not?”), it may be useful to have someone evaluate all of the responses to identify common themes or answers. These themes can be coded for later summary, using a small number of categories in a derived categorical variable.

The dataset should include variables to calculate the survey weights and to identify which cluster and household the respondent comes from. If the survey design was repeated across numerous strata (for example, a cluster survey was conducted in each region of the country), there should be a variable to indicate which stratum the respondent belongs to.

#### Derived variables showing evidence of vaccination

The analysis will summarize vaccination in several ways (crude doses, valid doses, doses given before age 1, etc.). It will be helpful to construct indicator variables for many of these conditions for later summary in the tables.

One helpful convention is to code the variable using a “1” if the respondent meets the category, using a “0” if he or she does not, and using a “missing value” if the respondent cannot be assessed for the variable in question. Some examples of helpful vaccination indicator variables include the following, for each vaccine/dose combination (for example, BCG, OPV0, OPV1, OPV2, OPV3, DTPCV1, DTPCV2, DTPCV3, MCV, etc.):

* got\_DTPCV3\_by\_card
* got\_DTPCV3\_by\_register
* got\_DTPCV3\_by\_history
* got\_DTPCV3\_by\_any\_source
* got\_crude\_DTPCV3
* got\_valid\_DTPCV3
* got\_DTPCV3\_by\_12months
* got\_DTPCV3\_resolved\_for\_coverage (this last indicator is the one used for official coverage estimates; it applies logic like that listed below to resolve disagreements between card, register, and history).

#### Resolve data conflicts consistently

Some children in the dataset will have vaccination information from a single source (card, register, or history), but depending on the questionnaire and protocol, there may be more than one source of information for many children. If the sources disagree on whether the child received a particular vaccine or dose, then the analysis plan will need to specify a protocol or hierarchy to decide which source of information to use. It is best to specify this hierarchy early in the process, and to use it to construct the derived variables that indicate whether or not the child received a particular vaccine and dose (or whether they received it before the age of 12 months).

Although it is not possible to know which of the sources of information (card, register or history) most closely represents what happened, for the purpose of a standardized procedure for the analysis **we propose the following method of determining whether a child received a certain vaccine/dose combination**:

1. If health facility records were sought for **every child**:

* If both home-based (card) and health facility-based records (register) are available and there is evidence of vaccination (with a particular vaccine/dose) on **either** the card **or** the register, that vaccine/dose is considered received. If that vaccine/dose is not recorded on either document, then the child is considered unvaccinated for that particular vaccine/dose, even if there is a verbal history of vaccination.
* If a card is available but the child’s record was not located in the health facility records, the vaccination is classified according to the information on the card.
* If no card is available but the child was located in health facility records, the vaccination is classified according to the health facility record.
* If no card is available and the child was not located in the health facility records, vaccination is classified according to the verbal history given by the caretaker.

1. If health facility records were only sought for **children who did not have a home-based record**:

* If a card is available, the vaccination is classified according to the information on the card.
* If no card is available but the child was located in health facility records, the vaccination is classified according to the health facility record.
* If no card is available and the child was not located in the health facility records, vaccination is classified according to the verbal history given by the caretaker.

1. If health facility records were **not sought at all**:

* If a card is available, the vaccination is classified according to the information on the card.
* If no card is available, vaccination is classified according to the verbal history given by the caretaker.

## Generate a codebook

When the dataset is nearly ready, it is helpful to update the codebook (also called a *data dictionary*). The data manager and statistician should review it carefully to identify any remaining implausible values. An excellent codebook includes the following:

* **Overall Summary**. Briefly describes the source of the data, the time period and manner in which it was collected, and contact information for the organization responsible for the survey, in case eventual codebook readers have detailed questions.
* **List of variables**. A simple, uncluttered list of the variable names and labels for quick reading and electronic parsing.
* **Full Dataset Summary**. Summarizes each variable in the dataset, documenting variable name, label, type, and length, and then summarizing the variable in one of several fixed formats:
  + For categorical variables: a frequency table with data values, formatted labels, and a count of the number and percent of observations that take on that value in the dataset.
  + For continuous variables: a univariate summary including minimum, maximum, median, mean, standard deviation, standard error, and the number of observations that are missing, or that use special missing values (for example, Refused, Don’t Know, Questionnaire Item Skipped Appropriately).
  + For dates: an indication of the first and last dates in the dataset (to detect outliers).
  + For open-ended questions: the codebook can either list the variable and the number of missing and non-missing responses, or it can document every unique verbatim answer in the dataset (often in a separate section for each open-ended response).
* **Stratum-Specific Summaries**. In some cases where there are well-defined subgroups in the dataset, the responses from each subgroup are documented in a separate section. These data summaries are usually constructed, calculated, and formatted using automated tools that can easily produce periodic updates to codebooks, and can serve as a basis for conversations about project progress or difficult data-related issues.
* **Notes**. This part of the codebook provides any helpful information about the dataset, including special documentation of data quality flags, problematic periods of data collection, formulae for calculating derived variables, known problems with individual variables, citations to literature that describe derived variables, and validated scales or scores calculated from raw survey responses.

# Tabulation and analysis

This chapter describes standard and optional vaccination coverage analyses, and provides table shells and example figures for how to show the results. The WHO is planning to furnish open-source software starting in 2015, to analyse standard coverage surveys and populate tables and figures like the ones shown in this chapter.

It is essential to specify the desired analyses, table shells, and figures at an early stage of the project, to ensure that the survey sample will be adequate to meet the survey goals, and to ensure that there is adequate budget and time to do the analyses.

In the past, reference manuals have given guidance and formulas for calculating coverage estimates by hand. Now that the survey uses a probability sample and conducts a weighted analysis that accounts properly for the complex sampling design, we recommend always using survey software to do the analysis. Therefore, this manual does not provide formulas for calculating coverage estimates, confidence intervals, or confidence bounds. These should all be calculated using software and syntax appropriate for stratified cluster surveys. Appropriate software might include Stata, R, Epi Info, SAS or SPSS.

The survey report should describe clearly what software you used and, in many cases, what options you used within the software. How were standard errors and confidence intervals calculated? Did you use the Taylor-series linearization method or some other method? What confidence intervals were calculated for the coverage proportions? What statistical methods and what software procedures were used to test hypotheses? The report should be very clear on all these points. Accordingly, the software programs and syntax used to conduct analyses should be saved, not run once and deleted. They should be made available for auditing or for editing in case mistakes are found, or if the analysis needs to be re-run at a later date to incorporate some corrections.

Because this manual recommends collecting data from every eligible respondent in every household interviewed, the statistical software should account for the multi-level nature of the data, and for correlated responses from respondents nested within households nested within clusters. It should thus use appropriate syntax and techniques to incorporate the stratum ID, the cluster ID, the household ID, and where appropriate, the household resident number in the estimation.

Analysis of routine vaccination data takes more time than it once did, because the increasing numbers of vaccines and doses in national EPI schedules make the analysis more complicated. In addition, the new recommendation to seek documented evidence of vaccination by visiting health facilities creates an additional data set that prolongs the analysis process. Even after data have been collected well, managed well, and cleaned well, the summary and analysis of a coverage survey requires a substantial amount of statistical programming to generate clear results that are well-documented and reproducible.

## Conduct descriptive analyses to characterise the sample and assess its quality

### Describe the sample

Use a table such as Table 3 to describe the general characteristics of the sample and show how it compared to what was predicted in the planning phase. If the survey was stratified for example, urban/rural stratification, the table should show the results for each stratum so it is easy to identify any differences (such as in participation rates, card availability rates, sex distribution, age of participants). These differences may raise the possibility of data quality issues that need further investigation. It can be useful to populate this table by cluster during survey implementation to look for any outliers or missing data (for example, households lacking information on composition).

Table 3. Results of the household visits and interviews

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Urban | Rural | Total |
| Total households in sample (or stratum) | | ( ) | ( ) | ( ) |
| Households with information on whether or not an eligible individual resides there  - According to information from household member  - According to information obtained from neighbours[[11]](#footnote-12) | |  |  |  |
| Households with missing information | |  |  |  |
| Number of eligible individuals (by age group, if applicable) | | ( ) | ( ) | ( ) |
| Number of children for whom information on vaccination was obtained | | ( ) | ( ) | ( ) |
| Number of children for whom no information was available:   * + - * Caretaker unavailable       * Refused       * Other | |  |  |  |
| Sex of children: | * + - * Male       * Female |  |  |  |
| Note: Numbers listed in parentheses would be expected counts, based on pre-survey plans and demographic expectations, listed here for comparison purposes. | | | | |

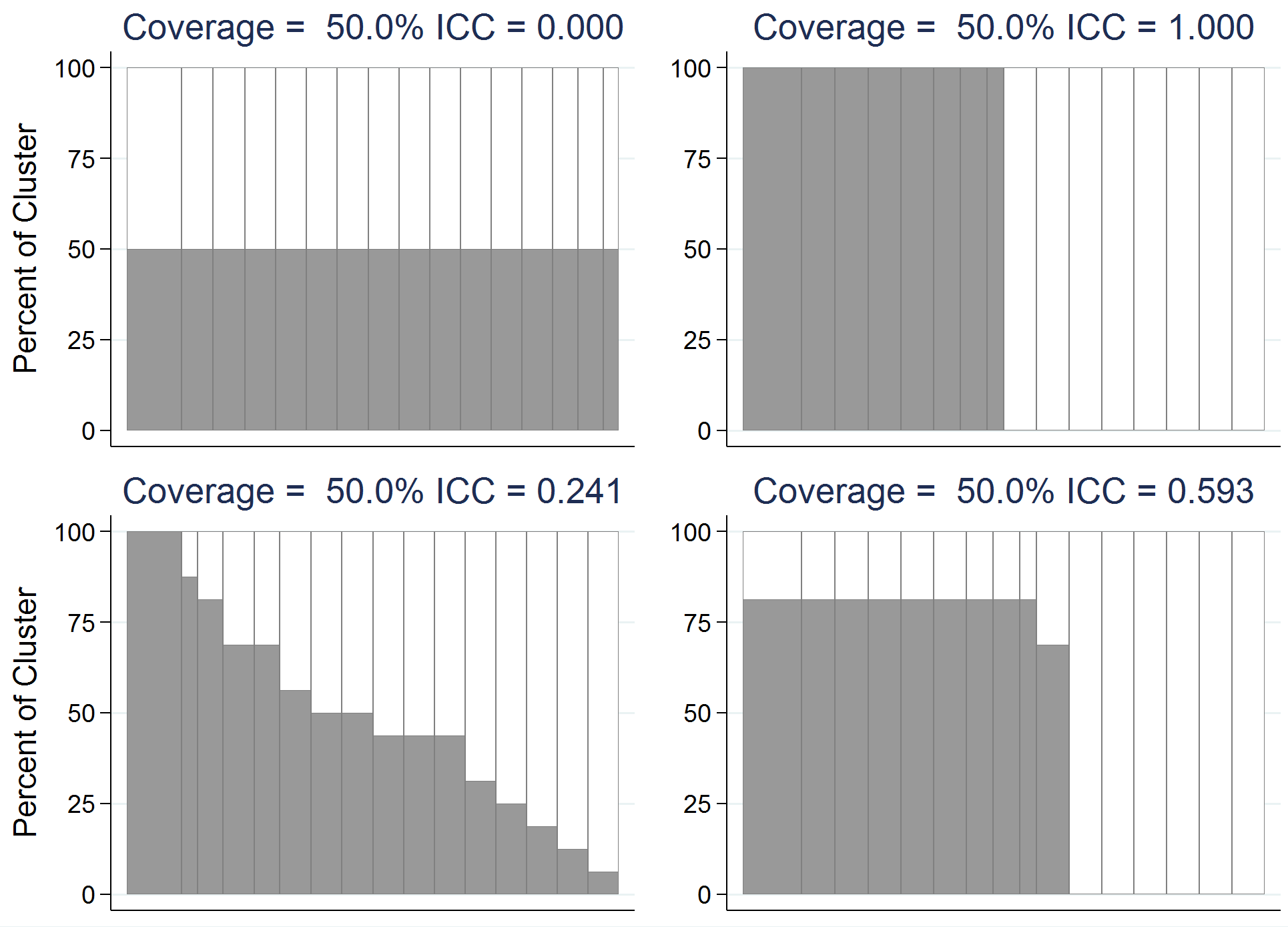
### Summarize coverage data graphically

A helpful way to visualize coverage survey results is with a simple bar graph called an *organ pipe plot,* in which each vertical bar represents a cluster, and the colored portion of the bar represents the weighted proportion of survey respondents in the cluster who were found to be vaccinated. The width of each cluster’s bar is proportional to the sum of its survey weights, and the bars are sorted, left to right, in descending order of cluster-level coverage. See Figure 4 and Figure 5. The plots derive their name from the stepwise decreasing shape of the shaded region, like a section of organ pipes in a concert hall. WHO is preparing downloadable software templates for constructing these figures.

Figure 3. The name “Organ Pipe Plots” is inspired by pipes like these



Figure 4. Organ pipe plots for four hypothetical strata, each with coverage of 50%

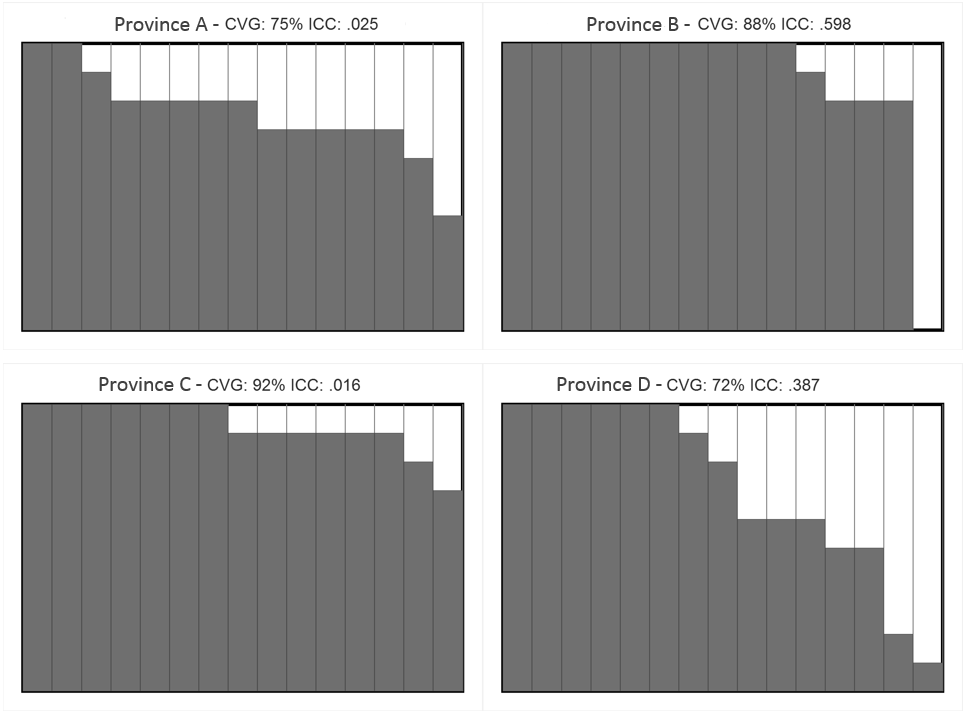


ICC: Intracluster correlation coefficient

The plot provides an intuitive representation of what the survey found. If all survey respondents were vaccinated, the entire chart would be shaded. If none were vaccinated, it would be empty. When bar width is proportional to the sum of survey weights in each cluster, the proportion of the chart that is shaded is equal to the survey-weighted coverage estimate. Any variability of bar heights reflects heterogeneity in the cluster level coverage estimates, and dramatic variability may reflect differences in vaccination programme performance.

The variability in bar heights is a visual representation of intracluster correlation (ICC) and is related to the design effect (DEFF). A stratum with homogeneous coverage will have a design effect very near 1. If some clusters have 100% coverage and all others have 0% coverage, the design effect will take on its maximum possible value. Other patterns of coverage will result in design effects that range between 1 and the average number of responents per cluster.

Figure 5. Organ pipe plots for four real strata from a self-weighted measles SIA



CVG = estimated coverage; ICC = intracluster correlation coefficient

Construct organ pipe plots for each vaccine and each stratum in the survey. They can be very effective and intuitive with very few labels—just the name of the stratum, and the vaccine and dose. It will not always be necessary to label the clusters, although you may wish to subtly indicate the number of completed interviews in each cluster or add some other detail to put the data into context.

### Identify clusters with alarmingly few vaccinated respondents

In most cases, we do not recommend interpreting cluster-level coverage results, because they are usually based on a very small sample and do not provide a precise estimate of local coverage. As a matter of fact, the small sample size in each cluster results in the estimated coverage changing a substantial amount with each person vaccinated. These results are meant to serve as a sample that is aggregated at the stratum level where a meaningfully precise estimate is expected. We do, however, recommend that special attention be paid to clusters that yield remarkably few vaccinated respondents. For instance, if a cluster yields zero children, for instance, who were vaccinated in the most recent SIA (for example, Province B in Figure 5), this is an important result that should be communicated to health officials right away. It does not necessarily indicate that campaign workers failed to vaccinate that cluster, but given a well-organized campaign it would be very unlikely to find that every eligible child surveyed was not vaccinated. Either way, some investigation and follow-up is warranted. Similarly, it would be notable in a routine immunization survey to find a cluster where zero survey respondents had received BCG (or any other first-dose vaccine); this is an important result that should also be communicated to health officials and investigated further as this may indicate a problem with access to vaccination services.

The organ pipe plot will give a quick visual indication of whether there are clusters with alarmingly few vaccinated children in the survey sample. The threshold for what to consider alarmingly few might vary…certainly zero is alarmingly few. In some contexts, one or two or three might also be considered alarmingly few vaccinated children in the survey sample. It can be helpful to provide a separate report on this issue. In fact, this finding does not depend on survey weights, so it would be possible to generate unweighted plots and run this report as soon as the dataset is cleaned, even before the survey weights are available. This would provide immediate actionable information from the survey.

Hopefully most strata will not yield any clusters with low coverage, but when one does, consider providing the following information in a brief report:

1. For each vaccine/dose of interest, list the clusters where alarmingly few respondents were vaccinated. List the stratum, cluster number and name, number of completed interviews, and number of respondents who were vaccinated, possibly breaking out results according to card, register, and caretaker history.
2. If the survey asked caretakers for reasons for non-vaccination, tabulate those reasons by cluster – compare the reasons for non-vaccination in the clusters with higher and lower coverage values. Any striking differences in those reasons may provide a clue as to why the coverage in the sample was so low. Also tabulate any comments that accompanied the survey forms. These responses from caretakers may shed some light on what is happening there; perhaps the neighborhood clinic is usually closed or maybe there is a prevalent anti-vaccination attitude in that neighborhood.
3. Where applicable, provide a map showing the clusters of interest, possibly overlaying health district boundaries, to show health officials precisely where the data of interest were collected.

These materials should be used to follow-up for in each identified cluster, to understand the reasons for the low coverage among respondents.

## Calculate weights for analysis

Each completed survey response will be accompanied by one or more *weights*, calculated by a statistician or by the census agency. When a survey calculation is *weighted*, it means that each person selected for the sample represents a certain number of similar eligible persons from the population. The analysis gives additional weight to respondents who represent more people than to those who represent relatively fewer people. Ideally, the sum of the weights will equal the total target population for the survey.

The first weight is a *sampling weight* that represents the probability that the respondent was selected to participate in the survey (see Annex J):

* In a single-stage cluster sample, where every eligible person in the cluster is sampled, the sampling weight is simply one divided by the probability that the cluster was selected into the survey. This probability is calculated using the numbers in the list used for PPES sampling.
* In a two-stage sample, the sampling weight incorporates the probability that the cluster was selected *and* the probability that the household was selected, given that the cluster was selected.

A second set of weights may be *adjusted for non-response* after the data have been collected and cleaned. These weights are developed after it becomes clear how many households had no one at home, despite high-quality fieldwork with interviewers revisiting those homes at least twice, and also after it becomes clear how many eligible respondents declined to participate.

The first or second set of weights will be sufficient for estimating population proportions, like coverage estimates within each stratum. But in most cases the analysis plan also calls for pooling the estimates across strata to calculate a national coverage estimate. And sometimes the analysis is intended to estimate population totals: What is the estimated number of children in the country who are unvaccinated? What is the estimated number of children born in the last year who were not protected at birth from neonatal tetanus? In order to aggregate coverage estimates across strata or estimate totals, it will be necessary to calculate yet another set of weights: *post-stratified* weights.

Post-stratified weights are adjusted to make them sum to the known eligible population in each stratum, if such population totals are known to be accurate. To post-stratify, each weight is multiplied by a stratum-specific factor equal to the known population of the stratum divided by the sum of (first set or second set of) weights in that stratum. If the weights need to be post-stratified to fit population totals for several demographics, seek the help of a sampling statistician.

## Conduct standard analyses

A standard survey provides results on coverage for each stratum and each vaccine in the survey. Include the following survey-weighted analyses in every coverage survey report:

* Crude coverage (includes all doses, whether valid or not) for each respective vaccine by document (home-based record (card) and/or register) plus history, by the time of the survey (12–23 months of age). This is the most liberal (highest) estimate of coverage.
* Crude coverage for each respective vaccine by age 12 months (or at birth for Td or TT), based on document plus history. Doses received after age 12 months are not counted in this analysis. You will need to make some assumptions about the dates of vaccination for children without documentation in order to calculate coverage levels by age 12 months. Annex L gives an example of how to do this calculation.
* Valid coverage for each respective vaccine and of fully vaccinated children at age 12 months, classifying children without a document as unvaccinated. If both the home-based record and health register data are available, but each has a different date of vaccination, then if **either** of the sources show that the dose was valid, it is accepted in this analysis. The analysis for valid coverage by 12 months of age:
  + Excludes vaccinations given after 12 months.
  + Is based on documented information (home-based record or health centre register).
  + Includes only those DTPCV, OPV, RV, and PCV doses with a minimum of 28 days between doses, and at a minimum age of 6 weeks (36 days[[12]](#footnote-13)) for the first dose and a minimum age of 9 months (266 days of age) for measles-containing vaccination. If the document indicates that one of the earlier doses in a sequence was invalid but followed later by valid doses, then for the purpose of this calculation invalid doses are dropped and later valid doses are shifted down, and counted as if they had been the earlier dose.
    - For example, consider a child who received DTP at 7 weeks, 10 weeks, and 14 weeks. The dose administered at 10 weeks of age is not valid because it was given before four weeks elapsed after the first dose. So that dose would be ignored, and the dose given at 14 weeks would be counted as the second valid dose. In the valid dose analysis, this child is counted as having had DTP1 and DTP2, but not DTP3.
* The dropout rate (proportion) between the first and third doses of multi-dose vaccines and between BCG or first dose DTPCV and measles-containing vaccines, with and without exclusion of invalid doses.
  + For example, for crude dropout rates between DTPCV1 and DTPCV3, if the weighted sum of children who received DTPCV1 is 200 and the weighted sum of children who received DTPCV3 is 150, the dropout rate is = = 25%.

Since the results of a survey are based on a sample rather than a census, they have an element of uncertainty. Confidence intervals of estimates are important to convey the range of values likely to include the true population coverage value with a given probability (usually 95%).

Whenever a population level parameter is estimated with the survey data, confidence intervals should be included in tables, as shown in the following shell tables and in the worked examples in the annex.

On the other hand, it is not necessary to calculate confidence intervals when tables for the report are simply summarizing descriptive statistics about the sample dataset. This distinction is important: If the report says that 24% of the survey respondents were found to be illiterate, then there is no confidence interval needed; you are describing the sample and not the population. The analysis is not weighted; each respondent counts as much as the next. The figure 24% of the sample is not subject to uncertainty. Nevertheless, if you use the survey data to estimate the proportion of caretakers of children 12­–23 months in the entire population who are illiterate, then it is appropriate for the calculation to be weighted, to take the complex design into account, and to include a confidence interval with the point estimate.

Both the analysis plan and the survey report should be very clear about which results are describing the sample only (these will be unweighted and will not have confidence intervals) and which results are describing the eligible population of respondents.

There are different philosophies about the best methods for calculating confidence intervals for proportions using survey data. See Brown, Cai & DasGupta (2001) along with responses to it in the same journal and subsequent literature that cites this article. In this manual, we recommend the modified Clopper-Pearson intervals suggested by Korn & Graubard (1998) because they are conservative. Conclusions drawn from them are likely to be stronger and require fewer caveats than those based on other methods.[[13]](#footnote-14)

### Summarise coverage estimates graphically using inchworm plots

In addition to tabular summaries of vaccination coverage, it is helpful to display coverage results graphically for key vaccines. This manual recommends a new representation of estimated coverage results called *inchworm plots*. See Figure 6, Figure 7, Figure 8 and especially Figure 9 for examples. See the material in Annexes M and N for detailed descriptions and examples.

Inchworm plots portray point estimates along with two-dimensional representations of the 95% confidence intervals, and tick marks at the 95% lower and upper confidence bounds. They can be used to show estimated coverage for one (or more) vaccine(s) per plot, and each plot can convey results for many strata at once. In a single plot, each two-dimensional distribution is drawn using the same total area, so survey estimates with narrow confidence intervals are tall and look like an inchworm that is bunched up, ready to stretch. Estimates with comparatively wide confidence intervals are less tall or bunched up, and look more like an inchworm that is stretched out.

Within a province, the plots sort districts by coverage, with the lowest at the bottom and the highest at the top. Similarly, within the nation the plots can sort provinces by coverage, again with the lowest at the bottom and the highest at the top. These figures are intended to provide survey stakeholders with an intuitive visual summary of estimated coverage across all strata in the survey. They represent the precision of the estimate such that tall narrow inchworms result when sample sizes are large or coverage within a stratum is homogeneous. And long slender inchworms result when there is more uncertainty due to small sample sizes or high heterogeneity in the sample. Inchworm plots sometimes include tabular summaries at the right side of the graphics, listing the point estimate, and one or more of the three confidence intervals described in Annex M. (By the time of this manual’s final revision, a WHO website will provide Stata and R programs to construct inchworm plots from users’ data.)

Table 4. Crude vaccination coverage by source of information, by age at the time of the survey, among (N=\*) children aged 12–23 months

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Vaccine, dose[[14]](#footnote-15) | Documented from home-based card\* | Documented, from card OR register  (b) | If no card or register, according to verbal history  (c) | Total (b+c) |
|  | n1 %  (95%CI) | n2 %  (95%CI) | n3 %  (95%CI) | n2+n3 %  (95%CI) |
| BCG |  |  |  |  |
| HBV0 |  |  |  |  |
| OPV0 |  |  |  |  |
| DTPCV1 |  |  |  |  |
| OPV1 |  |  |  |  |
| PCV1 |  |  |  |  |
| RV1 |  |  |  |  |
| DTPCV2 |  |  |  |  |
| OPV2 |  |  |  |  |
| PCV2 |  |  |  |  |
| RV2 |  |  |  |  |
| DTPCV3 |  |  |  |  |
| OPV3 |  |  |  |  |
| IPV |  |  |  |  |
| PCV3 |  |  |  |  |
| RV3 |  |  |  |  |
| MR 1 |  |  |  |  |
| YF 1 |  |  |  |  |
| Fully vaccinated[[15]](#footnote-16) |  |  |  |  |

\* Column (a) is a subset of Column (b), but is listed separately to make it easier to compare results with other surveys that do not look for health centre records

**N = total number of individuals in the survey. n = number of individuals who received each vaccine according to each source of information.** Note: the % vaccinated is not simply n/N because we do a weighted analysis to take into account the sample design, and not all individuals in the population had the same chance of being selected into the survey (see section 6.2).

Table 5. Crude and valid vaccination coverage by age 12 months

|  |  |  |
| --- | --- | --- |
| Vaccine, dose | Crude Coverage – documented evidence or caretaker recall of vaccination, (includes invalid doses and verbal history)  Estimated % 95% CI 95% LCB 95% UCB | Valid Dose Coverage – documented evidence of vaccination at correct ages and with correct intervals (includes only valid doses)  Estimated % 95% CI 95% LCB 95% UCB |
|  |  |  |
| BCG |  |  |
| HBV0 |  |  |
| OPV0 |  |  |
| DTPCV1 |  |  |
| OPV1 |  |  |
| PCV1 |  |  |
| RV1 |  |  |
| DTPCV2 |  |  |
| OPV2 |  |  |
| PCV2 |  |  |
| RV2 |  |  |
| DTPCV3 |  |  |
| OPV3 |  |  |
| IPV |  |  |
| PCV3 |  |  |
| RV3 (if in schedule) |  |  |
| MR 1 |  |  |
| YF 1 |  |  |
| Fully vaccinated[[16]](#footnote-17) |  |  |

CI: confidence interval: LCB: lower confidence bound; UPC: upper confidence bound

Table 6. Survey-weighted dropout rates between different vaccine-dose combinations, by source of information

|  |  |  |
| --- | --- | --- |
| Dropout between[[17]](#footnote-18) | Any dose, documented or history | Valid doses only, documented source of information |
|  | Coverage difference between earlier and later doses divided by earlier dose Estimated % 95% CI 95% LCB 95% UCB | Coverage difference between earlier and later doses divided by earlier dose  Estimated % 95% CI 95% LCB 95% UCB |
| BCG - MCV1 |  |  |
| DTPCV1 - DTPCV3 |  |  |
| DTPCV1 - MCV1 |  |  |
| DPTCV3 – MCV1 |  |  |
| OPV1 - OPV3 |  |  |
| RV1 - RV3 \* |  |  |
| PCV1 - PCV3 |  |  |

\* (or RV2 if 2-dose schedule) CI: confidence interval: LCB: lower confidence bound; UPC: upper confidence bound

## Conduct additional analyses

This section describes additional analyses that can give very useful information to programme managers. Some rely on having a dataset with vaccination dates, thus, are restricted to children with documented vaccinationand may be advisable only where carad availability is high.

Additional analysis options include:

* Missed opportunities analysis
* Vaccination by calendar month
* Assessment of the age at receipt of each dose (that is, validity and timeliness)
* Coverage by subgroups
* Comparing coverage between different locations in the same survey
* Comparing coverage over time
* Concordance across sources
* Co-administration or simultaneous vaccination.

### Missed opportunities[[18]](#footnote-19)

In the context of a coverage survey, a missed opportunity for vaccination (MOV) is the failure to administer all vaccines for which the child was eligible (according to the national vaccination schedule) on the date of a clinic visit. For these analyses, only children having at least one documented date of vaccination are included. This analysis gives an idea of the MOV, as it is not possible to know whether a real conraindication existed.

For example, a child who received a first dose of DTPCV at age 6 weeks but did not receive pneumococcal conjugate vaccine (PCV) on the same date, when the national schedule recommended both at age 6 weeks and no true contraindication existed, has a MOV for PCV. A child may have multiple MOVs for a given vaccine.

Two types of analyses are recommended: (1) visit-based analysis and (2) child-based analysis. As their names suggest, the visit-based analysis analyses the number of health facility visits of the children where there was 1+ MOV, whereas the child-based analysis analyses the number of children who experienced 1+ MOVs.

The steps to accomplish an MOV analysis are described briefly here, and in more detail in Annex O.

#### Visit-Based Analyses

The visit-based (VB) analysis consists of three calculations: the proportion of visits resulting in MOV for each vaccine (VB1), the proportion of visits resulting in at least one MOV across all vaccines (VB2), and the rate of MOVs per visit across all vaccines (VB3).

*(VB1) Proportion of visits resulting in an MOV for a given vaccine:*

Numerator: Number of visits where a child received another vaccine (documented by card or register) and was eligible for the considered dose, but did not receive the considered dose

Denominator: Number of visits where a child was eligible to receive the considered dose

*(VB2) Proportion of visits with at least one MOV (across all vaccines)*

Numerator: Number of visits with at least one MOV (for any vaccine)

Denominator: Number of visits where a child was eligible to receive at least one vaccine

*(VB3) Rate of MOVs per visit (across all vaccines)*

Numerator: Number of MOVs summed across all vaccines (that is, sum of VB1 numerator across all vaccines)

Denominator: Number of visits where a child was eligible to receive at least one vaccine

Note: This calcuation is a rate, and so results greater than one are plausible.

#### Child-Based Analyses

The child-based (CB) analysis consists of two calculations: the proportion of children who had at least one MOV for a given vaccine (CB1), and the proportion of children with at least one MOV across all vaccines (CB2). CB1 can be further subdivided into the proportion of children who never received the particular vaccine (an uncorrected MOV), and those who did receive it by the time of the survey (a corrected MOV). Similarly, CB2 can be subdivided into the proportion of children where none, all, or some of the MOVs for the child were corrected by the time of the survey.

*(CB1) Proportion of children who had at least one missed opportunity for a given vaccine:*

Numerator: Number of children with at least one vaccination date recorded, who were eligible to receive the considered dose but did not receive the considered dose

Denominator: Number of children with at least one vaccination date recorded, who were eligible to receive the considered dose

Subdividing (CB1):

*(CB1a) Proportion of children with uncorrected MOVs*

Numerator: Children in (CB1) numerator who had not received the given vaccine by the time of the survey

Denominator: Same denominator as (CB1)

*(CB1b) Proportion of children with corrected MOVs*

Numerator: Children in (CB1) numerator who had received the given vaccine at a later visit as documented by the vaccination card

Denominator: Same denominator as (CB1)

*(CB2) Proportion of children who had at least one missed opportunity for any vaccine:*

Numerator: Number of children with at least one vaccination date recorded who did not receive a vaccine/dose when they were eligible for it

Denominator: Number of children with at least one vaccination date recorded who were eligible to receive at least one vaccine/dose

Subdividing (CB2):

*(CB2a) Proportion of children with no corrected MOVs corrected*

Numerator: Children in (CB2) numerator who had not received the vaccine(s) by the time of the survey

Denominator: Same denominator as (CB2)

*(CB2b) Proportion of children with all corrected MOVs corrected*

Numerator: Children in (CB2) numerator who had received the vaccine(s) at a later visit as documented on the vaccination card

Denominator: Same denominator as (CB2)

*(CB2c) Proportion of children with some corrected MOVs corrected*

Numerator: Children in (CB2) numerator who had received some, but not all, of the vaccine(s) at a later visit, as documented by the vaccination card

Denominator: Same denominator as (CB2)

After the visit-based and child-based MOV analyses are conducted, it is possible to calculate the potential coverage that could have been achieved if there had been no missed opportunities. This is done by re-estimating coverage while counting the children who had an *uncorrected* MOV for a given vaccine as if they had received the vaccine. This essentially moves these children from the “did not receive vaccine” group in the original coverage estimate calculation to the “documented from card” group. The coverage estimate is then recalculated, as shown in this shell table.

Table 7. Potential coverage achievable by time of survey among (n=\*\*) children with a documented source of information (card or clinic register), if all doses had been valid and all opportunities taken

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Documented vaccination at correct ages and with correct intervals (only including valid doses\*) | | | % coverage possible if no MOVs  (only including valid doses) | | |
| Vaccine/dose | N  (unweighted) | % | 95% CI | N  (unweighted) | % | 95% CI |
| BCG |  |  |  |  |  |  |
| OPV0 |  |  |  |  |  |  |
| DTPCV1 |  |  |  |  |  |  |
| OPV1 |  |  |  |  |  |  |
| RV1 |  |  |  |  |  |  |
| DTPCV2 |  |  |  |  |  |  |
| OPV2 |  |  |  |  |  |  |
| RV2 |  |  |  |  |  |  |
| DTPCV3 |  |  |  |  |  |  |
| OPV3 |  |  |  |  |  |  |
| IPV |  |  |  |  |  |  |
| RV3 |  |  |  |  |  |  |
| MCV1 |  |  |  |  |  |  |

The steps to go through to arrive at this table are described in detail in Annex O, and illustrated there using data from a recent DHS. The annex also describes how MOV analyses can address potential opportunities to compensate for doses given too early or with too short an interval.

Finally, the survey report should emphasize that if the survey dataset includes only dates from vaccination records then it is likely to underestimate the number of MOVs because some of those same children will have visited the clinics on other occasions (sick visits or well visits) and experienced an MOV, but the dates for those visits are not recorded on the vaccination card.

### Vaccination by calendar month

You can chart the month and year of each vaccine dose administered to children in the survey, to show if there were any time periods when little or no vaccination activities happened. This will provide useful information for discussion with programme managers—for example, discussing if stockouts or seasonal inaccessibility had occurred, or other reasons for lack of vaccination during certain periods.

### Assessment of the child’s age at receipt of each dose

Bar charts showing the age at which children received each vaccine are helpful to show health workers how closely they are following the schedule, and how early (or late) children are likely to be fully protected against vaccine-preventable diseases. This additional information can guide programme performance. It may also be helpful to report mean age at vaccination, median age at vaccination, and an interquartile range.

You can report results in a table, assessing the mean or median number of extra days or weeks (past recommended vaccination dates) that children remain under-vaccinated and at risk of disease, and risk factors due to the delay in vaccination. If statistical expertise is available, the statistician can use a reverse Kaplan-Meier curve (in which the y-axis is the probability of being vaccinated) to show the increase in coverage by age and the benefit of continuing to vaccinate children over one year of age.

### Coverage by subgroups

Calculating coverage by demographic categories such as sex, maternal education, and urban/rural residence can provide useful insight into potential risk factors for under-vaccination.

If you are planning to report survey results by subgroups, you will need a large enough sample size to report precise results within these groups. Alternatively, if detailed data are available from a recent census, you could adjust (*post-stratify*) survey weights to yield representative results for these groups, but the results may not be very precise, especially in districts. Formal statistical tests such as chi-squared tests are needed to determine if differences are statistically significant. The Rao-Scott chi-squared tests are appropriate for data from weighted complex surveys[[19]](#footnote-20) (Rao & Scott, 1979, 1981, 1984, 1987).

If the sample size is not large enough or if the weights have not been adjusted, it is recommended that you do report estimated population-level parameters by subgroup.

Note also that it is not appropriate to simply break the dataset into subgroups to calculate and report coverage separately in each. Because coverage is a ratio, both the numerator (number of vaccinated children) and the denominator (number of eligible children) are random variables that are being estimated with the survey data. Subgroup estimates should be calculated with the appropriate software syntax to incorporate the uncertainty in both the numerator and the denominator. This is sometimes described as *domain analysis*.

### Comparing coverage between different locations in the same survey

It may be desirable to make a formal statistical assessment of whether coverage in one region is likely to be higher than that in another region, using data from a single (cross-sectional) survey. This hypothesis test can be performed using statistical software that takes the complex sample design and survey weights into account, with the report listing the statistical test used along with the test statistic and resulting p-value and conclusion.

These tests are sometimes conducted informally by examining the 95% confidence intervals for the two regions. If the intervals do not overlap, the formal statistical test will clearly find a difference that is statistically significant at α=5%. But we cannot use this so-called eyeball test when the intervals do overlap somewhat – the formal test may or may not conclude that there is a statistically significant difference. If the intervals overlap, calculate using a statistical test (Payton, Greenstone & Schenker, 2003; Schenker & Gentleman, 2001).

Some results may not be statistically significant but are still worth exploring. For example, zero-dose clusters flag problems that need to be investigated further later, even if the result does not show statistical significance.

### Comparing coverage over time

It may be desirable to test the statistical hypothesis that coverage is improving over time in a certain region. There may be relevant data from an earlier survey, and the steering group may wish to use a new survey to confidently conclude that coverage has improved over time. Annex B3 includes instructions for selecting a sample size for the new survey, to ensure adequate power to detect such a difference if it truly exists.

A comparison like this will be problematic if previous surveys were different from the current one in important ways. If the earlier survey was not based on a probability sample or was not analysed using survey weights and software that accounts properly for sampling design and weighted data, the results may not have been representative of the population in question, and so a comparison would be ill-advised. Also, if the earlier survey used different eligibility criteria, covered a different geographical region, or accepted different sources of evidence for vaccination than the current survey, then the two measurements may not be comparable.

However, if the earlier survey was based on a probability sample, was well conducted and well analysed, and had similar eligibility criteria and evidence of vaccination, a comparison may be feasible. If the survey-weighted 95% confidence intervals for the old and new coverage estimates do not overlap, one might conclude that the coverage has indeed changed over time and that the difference is statistically significant, with the probability that the conclusion is an error below 5%. If the confidence intervals overlap somewhat, a more formal test will be required.

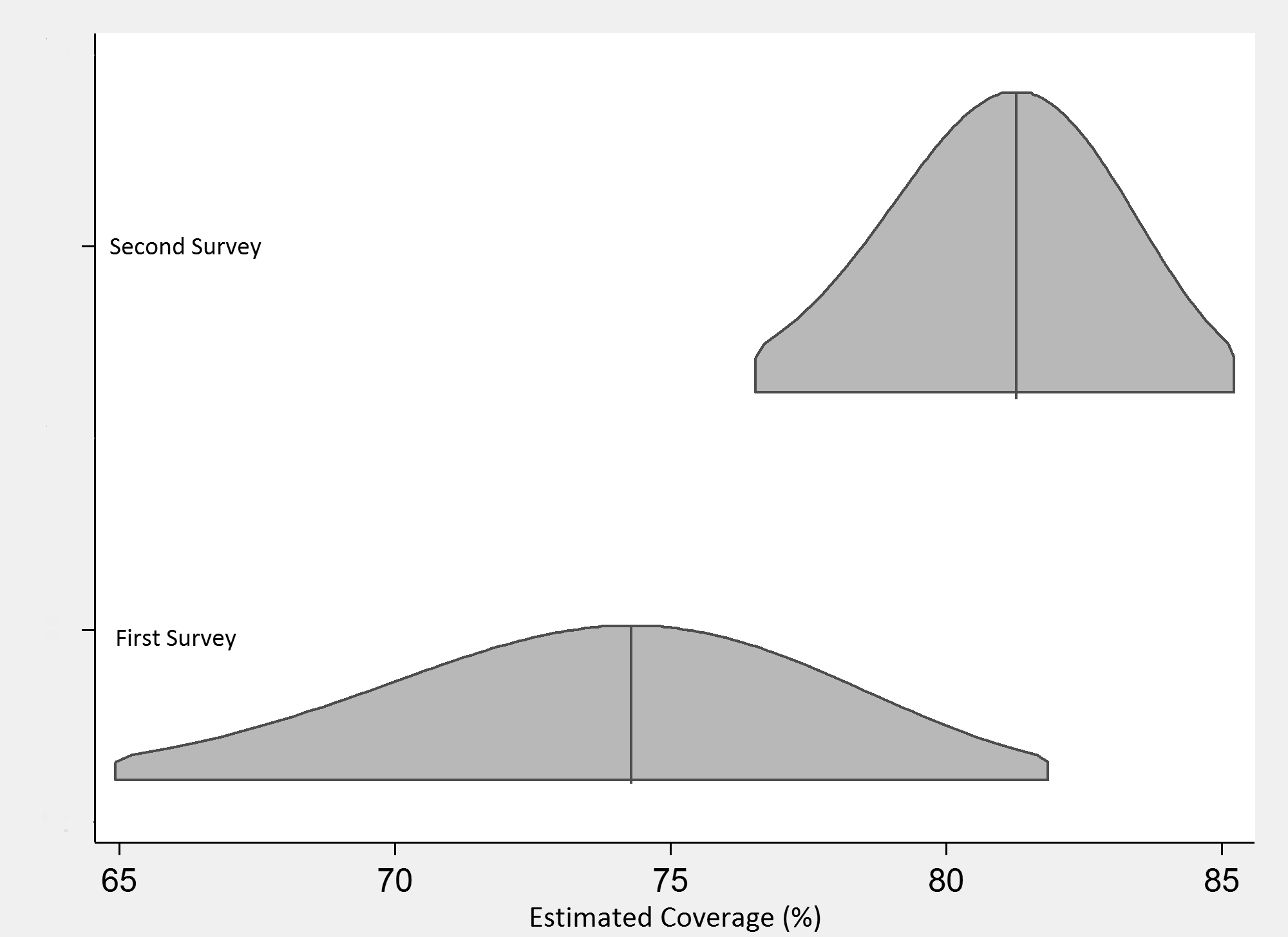
If the dataset from the previous survey is still available, it may be possible to bring both the old and new datasets together in the statistical software and conduct the statistical test. If the older dataset is not available, one way forward is to calculate the effective sample size and coverage estimates from each survey and construct a faux dataset consisting of two simple random samples, with sizes equal to the effective sample sizes of the survey datasets and coverage equal to the point estimates from the survey datasets. Then it is possible to use the faux data to conduct a formal test of difference in proportions.

#### Example of comparing coverage over time

An earlier, well-conducted EPI cluster survey used a probability sample in all stages of the design and reported DTP3 coverage of 74.3% using a sample size of 263 and a design effect of 2.5. Dividing 263 by 2.5 indicates that the effective sample size of the earlier survey was 105 respondents. The binomial exact 95% confidence interval for coverage is (64.8% – 82.3%). Later, a larger well-conducted EPI cluster survey using a probability sample in all stages estimated DTP3 coverage of 81.3% with a sample size of 725 and a design effect of 2.3. The effective sample size of this later survey is 725/2.3 = 315. The exact binomial 95% confidence interval is (76.5% – 85.4%). Estimated coverage has increased by 6 percentage points, from 74.3 to 81.3%.

Figure 6 summarizes the evidence and uncertainty regarding DTP3 coverage from these two surveys, showing the survey point estimates and 95% confidence intervals. Note that although the area under the two curves is the same, the distribution representing the CI from the first survey is much wider, due to its slightly lower coverage estimate and much smaller effective sample size. Note also that both confidence intervals are asymmetrical, with slightly longer tails on the left side (the side facing 50% coverage); this is appropriate for an estimated binomial proportion. The asymmetry would be more substantial if the estimated coverage were closer to 100%.

Figure 6. DTP3 coverage estimated at two different times with surveys of different sizes



We use a formal hypothesis test to address the question of whether the difference is statistically significant with a p-value below 0.05. The null hypothesis for this test is that the underlying population coverage at the earlier and later times is the same. A 2-sided alternative hypothesis would be that the population coverage has changed. The 2-sided test is more conservative; a 1-sided alternative might state that coverage has increased over time. A 1-sided alternative should be stated in the analysis plan before the second set of data are collected, and is only advisable if there is strong reason to believe, because of improvements to the vaccination programme, that coverage has increased. In this case, both a 2-sided and a 1-sided hypothesis test yields p-values higher than 0.05 (2-sided p = 0.127; 1-sided p=0.083; Fisher’s Exact Test).

This means that if these surveys were repeated over and over again in populations with the same underlying coverage for DTP3, we would expect 12.7% of those pairs of surveys to yield sample proportions at least as far apart as the two in these surveys by chance alone. Formally speaking, we fail to reject the null hypothesis. **The difference is suggestive of a change, but does not yield extremely strong evidence that the underlying coverage improved in the period between the two surveys.** Obtaining a p-value smaller than 0.05 for small changes in coverage requires extremely large surveys.

### Reporting results for comparisons

For comparisons conducted with hypothesis tests, the power of the survey to detect statistically significant differences of varying magnitude between different populations or times depends on the sample size and design. It is usually represented by tests of statistical significance.

When you report an estimated difference in coverage between places or times, or between coverage and a threshold, include the magnitude of the difference and its 95% confidence interval. Report the results of formal comparisons between coverage figures with a clear description of the statistical test that was done, the value of the test statistic, and the p-value of the test. The results should also include the size of the sample and an indication that the software took into account the complex sample design, which will often include stratification. For accurate interpretation, it will also be helpful to report the confidence intervals and sample sizes for the two quantities being compared.

It is not enough to report only that a difference is statistically significant. The *magnitude* of the difference is what matters for public health action. A difference of only 1 percentage point between sexes, for example, may be statistically significant if there is a large enough sample, but it may have minimal public health importance. A difference of 10 percentage points (for example, 70% in girls and 80% in boys) is much more likely to make policymakers take action to address gender inequity. So it is always important to report the estimated difference, along with its 95% confidence interval.

In other words, while the p-value informs us that the results have statistical significance, the magnitude of the difference matters for public health practice. Similarly, even when results are not statistically significant, they may be important to the programme and interesting to examine.

When hypothesis tests are one of the design goals of the survey, describe the parameters used to select the sample size. What magnitude of coverage difference was the survey powered to detect? What were the anticipated and observed values of the ICC or the design effect, and the anticipated statistical power? It will be helpful to compare the design parameters with those achieved in the dataset to help interpret hypothesis test results.

Each hypothesis test will have a certain number of so-called *degrees of freedom* that will be reported by the statistical software. Usually the degrees of freedom are equal to the number of clusters involved in the test minus the number of strata involved in the test. One suggestion for survey data analysis is to only report results from subgroup comparisons that have 12 or more degrees of freedom[[20]](#footnote-21). This guidance is intended to protect survey analysts against drawing inferential conclusions from datasets that are too small. We endorse this guidance and suggest that you examine the degrees of freedom for the comparisons in the analysis plan, and refrain from reporting those with fewer than 12.

### Assessment of quality of primary data recording

Surveys might be an opportunity to explore further specific operational aspects, although such additional analysis may increase the survey’s costs, duration, and complexity.

Many countries are conducting regular data quality assessments that compare information in registers with the information provided in reports to higher levels of the health system. Coverage surveys can provide an opportunity to assess the quality of primary data recording in registers and on vaccination cards. For example, if health facility register data is sought and entered for all available respondents, and not only the ones who did not have home-based records, it may be interesting to compare the card record with the register record on whether the child was vaccinated and when.

It may be also be useful to compare the concordance of facility records with caretaker recall. There can be several valid reasons why a caretaker might report that the child received a dose that is not in the register. The dose may have been received elsewhere or during a campaign. But it is interesting to note what proportion of caretaker reports agree with the documented doses. This information can give future survey designers information about how and whether to use caretaker recall of vaccination history as data.

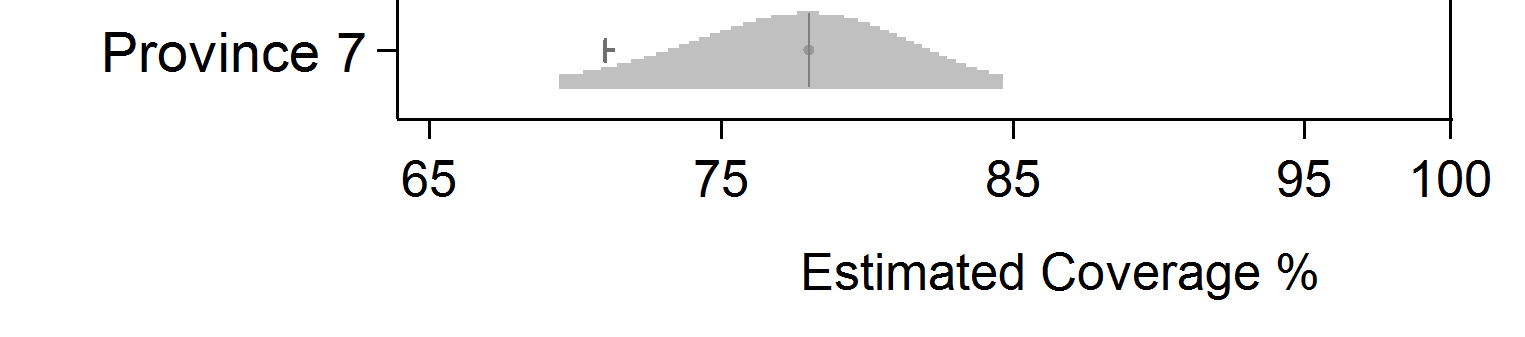
## Classifying coverage

### Overview

This section describes the process of classifying coverage at the lowest level of strata.

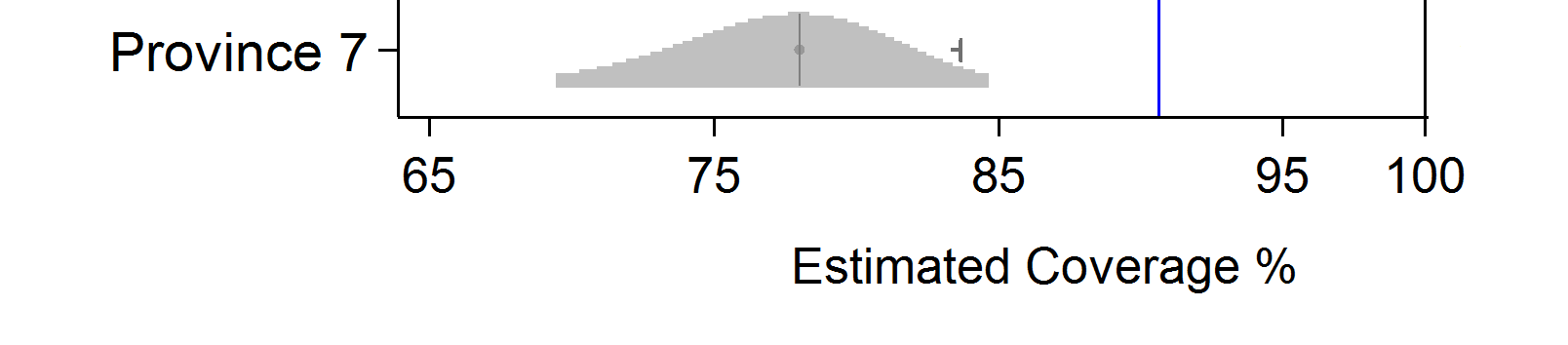
To classify coverage, we calculate a point estimate, a 95% confidence interval, and two 95% 1-sided confidence bounds: upper and lower confidence bounds (UCB and LCB, respectively). These figures are reported in tables and plotted on a graph. We can then make the very simple observation that because coverage **is likely** to fall on one side of the 1-sided bounds, then conversely it **is not likely** to fall on the other side of the bound.

Figure 7. Point estimate, 95% confidence interval and 95% lower confidence bound for coverage in hypothetical province #7



In Figure 7, the shaded distribution for Province 7 shows the 95% confidence interval for estimated coverage. The point estimate, at the highest point of the distribution, is at 78.0%. The 95% lower confidence bound is indicated with a small tick mark above the distribution at 71.0%. We might say, “We are 95% confident that the true population coverage lies above 71%.” If an important programmatic goal for this antigen in this province was 71% or lower, we could confidently classify coverage as falling above the goal. Using the language of hypothesis testing, a 1-sided test would reject the null hypothesis that coverage is < 71%. We might thus classify (label) Province 7 as one that passes, or has coverage that is adequate.

Figure 8. Point estimate, 95% confidence interval and 95% upper confidence bound for coverage in hypothetical province #7



In Figure 8, the shaded 95% confidence interval is the same as in Figure 7, but now we indicate the upper 95% confidence bound with a tick mark at 83.7%. Note that the programmatic goal of 90% coverage is indicated with a blue vertical line. Although the confidence interval for Province 7 is quite wide (69.5% to 84.7%), we can confidently classify the coverage as being 95% likely to fall below 83.7%. So this province clearly fails to meet the goal of 90% coverage. When the programmatic goal lies above the 95% upper confidence bound, then we can confidently classify coverage as falling below the goal. Here, coverage fails, or is inadequate.

In the intermediate situation, where the programmatic goal falls between the upper and lower confidence bounds, we cannot classify coverage as above or below the threshold with 95% confidence. We would have needed to conduct a larger survey to do that. But looking at the graphic confidence intervals for all strata, especially if they are sorted in order of estimated coverage, will show where each stratum falls in the pattern and should provide actionable insight, especially regarding the strata with the lowest and highest levels of coverage.

It is not strictly necessary to portray what you learn from the survey graphically, but it is strongly recommended. You can present point estimates, confidence intervals, and upper and lower confidence bounds in a table only, but the results may not be clear to stakeholders who do not have a clear understanding of confidence intervals and limits. Portraying the two-dimensional distributions of estimated coverage, and showing them for all the strata in the survey at once, is a powerful and intuitive way to communicate what you have learned about coverage from the survey. It is also a powerful way of communicating what you have NOT learned, such as when true coverage is very near a programmatic threshold and the sample size is small. In this case, you cannot use the survey to confidently conclude whether that particular stratum is above or below the threshold of interest.

To sum up:

1. Classification and estimation use the same underlying processes: calculate a point estimate and a confidence interval, and portray them. When classifying, also portray the 1-sided confidence bounds and use those bounds (rather than the ends of the confidence intervals) to make strong statements about whether coverage is above or below an important threshold.
2. This can be done using only tables, but adding graphics may help some audiences understand what you have learned more easily than tables alone.
3. Rather than sort the strata in alphabetic or administrative order, it is helpful to sort them in order of estimated coverage, or in order of the upper or lower confidence bounds. See Figure 9 below.
4. This approach to classification may be used with either small or large sample sizes. As the sample size gets larger, the upper and lower confidence bounds will fall nearer and nearer to the coverage point estimate. Conversely, if the sample sizes are small, the confidence bounds will fall farther from the point estimate. However, the principle of using the bound to confidently characterize whether coverage is above or below a threshold of interest is the same, regardless of sample size.
5. It is permissible to both estimate and classify coverage using a single survey. When describing estimation results, we usually focus on saying that the coverage is likely to fall **within a 2-sided** **confidence interval**. When classifying, we focus on saying that coverage is likely to fall **on one side of a confidence bound**. We recommend using at least 15 clusters per stratum for classification and at least 30 clusters per stratum for precise estimation.

### Examples of classification

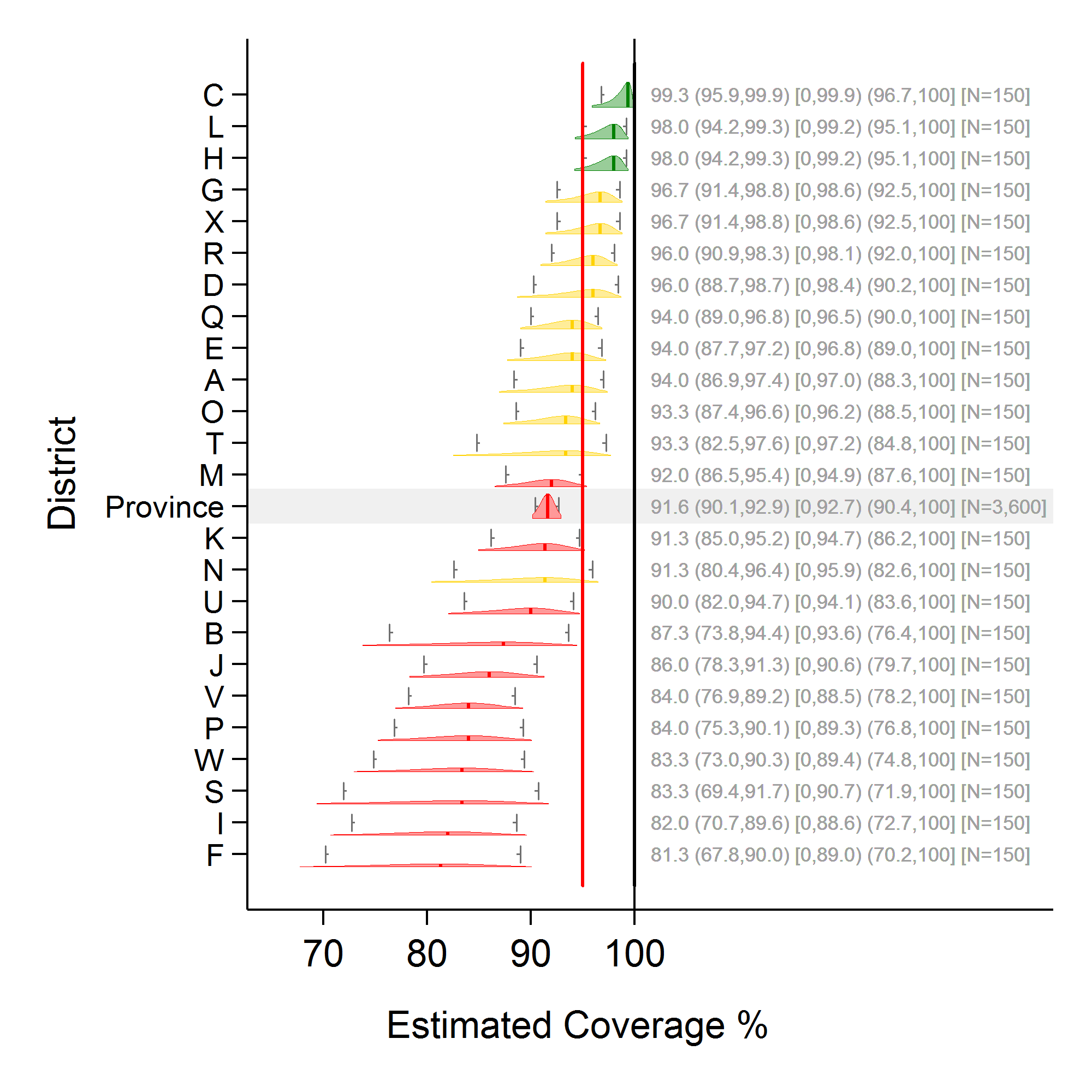
To classify coverage, calculate and plot the point estimate, the 95% CI, and the upper and lower 95% confidence bounds[[21]](#footnote-22). Recall that the 1-sided confidence bound is different than the endpoint of a 95% confidence interval. The 95% lower confidence bound can be calculated using the lower end of a 90% confidence interval. The 95% upper confidence bound can be calculated using the upper end of a 90% confidence interval. These bounds will fall inside the 95% confidence interval.

Figure 9 shows estimated measles SIA coverage for 24 fictional districts, based on samples of 15 clusters and 10 respondents per cluster in each district. For each district, the 95% confidence interval is indicated in light gray and the 95% upper and lower confidence bounds are indicated with small black tick marks. Three intervals are listed at the right side of each distribution. The first is the classic 2-sided 95% confidence interval. The second is the interval that extends from 0% coverage up to the 95% upper confidence bound. The third is the interval that extends from the 95% lower confidence bound up to coverage of 100%. All three intervals are equally valid for drawing conclusions with 95% confidence. The regions are plotted in increasing order of coverage point estimate, from bottom to top. The red vertical line marks the spot where coverage is 95%, an important programmatic threshold for measles. The district data are aggregated to estimate province coverage (shaded with a light gray bar) very precisely.

Although all the districts had samples of the same size, the width of the confidence intervals varies substantially, reflecting district-level differences in sample coverage and in the underlying ICC. Many of the intervals are too wide for precise estimation, but the data in the figure can be used to classify coverage into two or more categories.

Any consistent categorization is permissible as long as it is useful and described clearly. The programmatic threshold of 95% coverage is important for measles campaigns. Several logical coverage categorizations are described in Annex N.

Figure 9. Measles SIA coverage and confidence interval and bounds for 24 fictional districts and the province that they comprise; districts are sorted by estimated coverage

****Note: The distributions are plotted with equal areas, corresponding to 95% confidence for each district, so those with narrow confidence intervals appear taller and those with wider intervals have very little height. Tick marks near the left edge of each distribution indicate the 95% one-sided lower confidence bound; those near the right edge indicate the 95% one-sided upper confidence bound. The red vertical line indicates a programmatic threshold of 95% coverage. Districts coloured green are 95% likely to have coverage ≥ 95%. Those coloured red are 95% likely to have coverage < 95%. Those coloured yellow cannot be classified as above or below 95% with this sample of 150 respondents.

# Interpret, format, and share results

This chapter describes how to draft the survey report and present or summarize the survey results and their implications of the results for immunization programmes.

The coordinator and statistician prepare a primary report of the vaccination coverage survey to communicate their findings and make recommendations to the commissioning authority. This report must be submitted to the ministry of health for their review and approval. After receiving approval, the coordinator can revise the report and work with the national EPI manager to prepare simpler and shorter reports, describing survey results and recommendations for health service workers in the areas covered by the survey. It is recommended also to share the findings with other stakeholders such as an immunization interagency coordinating committee.

The primary report should be attractively prepared and presented to encourage readership. The key points to include in the report are shown in Box 3.

----------------------------------------------

Box 3. Essential components of a report

**Title.** Give a title that clearly describes the location, year and purpose of the survey.

**Acknowledgements.** Acknowledge the source of the funding and others who made the survey possible.

**Executive summary.** Summarize the methods, key results, and implications for action. An executive summary is extremely important, and should contain enough information about survey methods and any limitations so that results can be interpreted correctly. Often, the summary is the only part of the report that is read by senior officials, survey funders, and vaccination programme partners.

**Background.** Givebrief information about the country and its demographics, the health services organization, the vaccination programme, and vaccination trends over time. Explain why the survey was done and what its objectives are.

**Survey methods.** Include details of the sampling frame used, as well as any regions excluded from the survey due to security problems or other access problems. Describe how the survey was implemented and the quality control methods used. Also describe the data transmission, processing and analysis methods.

**Results.** This section includes tables and charts accompanied by explanatory text.

**Discussion**. Discuss the main survey findings and their implications for action, as well as the survey limitations and how these may affect interpretation of the results. Be sure to discuss sources of uncertainty in the results of this survey and, if relevant, the uncertainty of other data with which the findings are being compared.

**Recommendations**. Make recommendations that focus on next steps for the ministry of health, and recommendations for programmatic action. If necessary, the report can also recommend further investigations, such as further study of factors that have affected coverage or differences in coverage between subgroups.

**Appendices**. Include copies of data collection forms, descriptions of the sample and weighting frame, a cluster list and a list of personnel involved.

------------------

## Draft the background section

Give a brief overview of the country, its demographics and its health services organization, as well an overview of the target population of the survey. Also give an overview of the vaccination programme, including the vaccination schedule(s) and trends in vaccination over time. Finally, explain why the survey was undertaken and the survey objectives.

## Draft the survey methods and limitations

Explain the survey design and the reasons for choosing the design. Highlight the aspects of the survey design that make it different from previous surveys, if applicable. For example, previous surveys may have used a two-stage rather than a single-stage cluster design, may not have used weighted analysis or may not have included record extraction from health facility registers.

Include details about the sampling frame used and how the sample was selected. Note any areas excluded from the survey due to security problems or other reasons. Explain how data was collected in the field from households and health facilities. Also explain the data-checking protocols used to ensure the quality of the data. Briefly explain how data were transmitted and processed, and the protocol for maintaining data integrity in these steps.

Every survey has limitations. Results are more useful when you understand and communicate these limitations to those who will use the data to make decisions about programmes. Discuss common potential sources of error and to what extent these errors were minimized in the survey:

* **Sampling frame**. Were any populations excluded from the sampling frame? How recent was it and what, if anything, was done to improve it? What implications were there for the calculation of sample weights? What are the implications of any deficiencies in the sampling frame for the observed coverage? For example, were excluded populations likely to have lower coverage, and how big were such populations?
* **Sampling procedures**. Report how the survey plan was carried out in the field and any deviations from the survey protocol. These may include the failure to revisit households, failure to record non-responses or what type of non-response occurred (for example, no one at home or refusal), problems with identifying cluster boundaries, or changes in security that prevented the team from reaching some selected clusters. Discuss any likely effect of such deviations on the survey findings.
* **Selection bias.** What proportion of households had a respondent present, and how did this compare with expected levels? What were the participation rates and how might this they have affected results?
* **Information bias.** For what proportion of children was a home-based record available, and how did this vary between strata? If some areas had very few records, what does this imply about the logistics of card distribution or caretakers’ motivation to keep the records? Is there any suggestion that interviewers did not give enough time to caretakers to retrieve the records? Of the cards seen, how many were illegible or had errors (for example, no vaccination dates, or dates out of range such as DTPCV1 before the birthdate)? Did this vary by area? How many children without records could be traced at a health facility to obtain documentation? What was the overall reliance on each caretaker’s verbal history, and how does this compare to previous surveys? What were the results of quality control (use of pictorial prompts, supervision, repeat interviews) to assess the reliability of a verbal history? The proportion of data contributed by a verbal history alone will affect the confidence in the estimates, and will need to be considered when comparing different survey results.
* **Data transcription and data entry errors.** Describe any errors that may have happened in this process, and the proportion of errors detected that were resolved (for example, by referring to a photograph of the record or by revisiting the household). How many values out of range could not be resolved, and how were these handled?
* **Missing data**. What statistical adjustment was made to account for missing data, if any?

## Draft the results section

Review the survey results in detail to determine which ones best answer the questions the survey was designed to answer. Choose which descriptive statistics are most relevant to the objectives of the survey and of most interest to whomever commissioned the survey. You will likely need to include all of the standard analyses (see section 6.3), but you should also consider which of the additional analyses, if any, are appropriate to include (see section 6.4).

Because survey results are based on a sample instead of a full census, they have some inherent uncertainty: if the survey were repeated using the same protocol and sample size, but a different set of households were visited, the results from those sampled households would vary somewhat from the ones sampled in this survey. This element of uncertainty, called *sampling variability* or *sampling error,* affects all survey results and is taken into account in different ways according to the type of result.

Select how you want to present the results, using the format that will make it easiest for the audience to understand the data. Diagrams and graphs are often most useful for communicating survey results. It is difficult to discern trends and draw conclusions from tables, but tables allow more detail to be presented. Tables should, therefore, be complemented by data visualizations. Decide which visualizations are most effective in drawing attention to the most important or relevant aspect of the data. Also consider visualizations that use color, lines and shapes to accurately portray the data. Choose visualizations that eliminate as much graphical clutter as possible.

In this manual we recommend the inchworm plot representation described in Chapter 6 for graphical display of coverage results. We recognize that bar charts are often used to portray coverage and are simpler to make than inchworm plots. If you portray coverage with a bar chart, be sure to include a representation of the 95% confidence interval on the chart to convey the magnitude of uncertainty due to sampling variability.

### Describing results for estimates of coverage

For descriptive results such as estimates of coverage, the precision reflects sampling variability and is usually represented by the 95% confidence interval. The estimated proportion of eligible persons in the population who received each vaccine is called the *point estimate* of vaccination coverage. These estimates are often the main outcome of interest, and significant attention should be given to them.

### Describing results for classifying coverage

To classify coverage, calculate the upper and lower 95% confidence bounds and compare those bounds to a pre-specified coverage threshold. It is always best to state classification rules clearly and report the upper and lower 95% confidence bounds to help readers gauge the strength of classification conclusions. Classify coverage as follows:

* When the lower 95% lower confidence bound falls **above** the threshold, confidently classify coverage as high; true coverage is very likely to be above the threshold.
* When the upper 95% confidence bound falls **below** the threshold, confidently classify coverage as low; true coverage is very likely to be below the threshold.
* When the threshold falls between the two bounds, conclude that the sample was not large enough at this level to classify with 95% confidence whether true coverage is above or below the threshold.

See Annex N for classification examples.

Some reports summarize classification results by simply listing which strata were classified as being high or low, but this practice is discouraged. Reporting only the qualitative result may be helpful for simplicity, but it comes at the cost of omitting important information that may be useful to some readers of the report. Consider listing the confidence bounds for every classification result, so that readers of the report can compare coverage to other thresholds that may be of secondary interest. It is helpful to report and plot the 95% confidence interval along with the upper and lower confidence bounds for each coverage result, as shown in section 6.5.

### Reporting aggregated results

If the sample was stratified and data were collected in all districts, the results may be combined or aggregated up to the next highest level (province), and the process of either estimation or classification may be repeated. If each province contains at least several strata, then the 95% confidence interval may be quite narrow and the results might be summarized using the interval. Whether they are narrow or not, it is also possible to use the upper and lower confidence bounds to classify coverage in the district as likely to be above or below important programmatic thresholds.

If data were collected in all districts, the results may be aggregated again to estimate coverage nationally. It will usually be the case that the national confidence interval is quite narrow and results are reported that way. It is also valid to calculate upper and lower confidence bounds to classify national coverage with respect to important thresholds.

## Draft the discussion section

The discussion section of the report is a guide to interpreting the results. Discuss the main survey findings and their implications for action, as well as the survey limitations and how these they may affect interpretation of results. Be sure to remind readers of how uncertainty (sampling error) may have affected the results.

The report should describe clearly what rules and methods are used for classification, along with the qualitative descriptive labels that may be applied. In the methods section, it might specify that if the lower 95% confidence bound fell above 80%, the district was classified as having high coverage and was otherwise classified as having low coverage. Translate this into a sentence your audience can understand. In this case, “high” might be interpreted to mean that we are 95% confident that the population coverage is above 80% and “low” means that we cannot be that confident. This report recommends listing the 95% confidence bounds along with the classification labels, to avoid ambiguity associated with qualitative labels for coverage categories.

## Develop implications and recommendations

Though your readers and stakeholders may be able to draw some of their own conclusions from the data, they rely on you to explain the data and what it means for the programme. The objectives of the EPI are to provide protection against vaccine-preventable diseases as completely and as early as possible. The data collected during the survey provide detailed operational information on the EPI performance, and therefore on the possible causes of obstacles to reaching their objectives. Below are some of the most common programmatic implications of the data.

### Coverage of each vaccine and of fully vaccinated children

This is the indicator most commonly used at national and international levels to measure overall performance. People will want to know the coverage of each vaccine and the percentage of fully vaccinated children (and 95% CI), and how they compare to results from administrative data and from other surveys. How has it coverage progressed over time and what are some likely reasons why?

An important factor in interpreting these results is the proportion of children who had documented evidence of vaccination and how this proportion compares to other surveys. The proportion of data contributed by a verbal history alone will affect the confidence in the estimates, and will need to be considered when comparing different survey results.

The study of the pattern of dropout between doses of vaccine, in which many children are given a vaccine early in the series but not given the later doses, will provide clues about where programme problems may lie and should be addressed.

### Reaching a birth cohort

Sometimes the results indicate that there was difficulty in reaching a certain birth cohort.

* DPT1/BCG crude coverage by record, history, and register is an indicator of access to vaccination services (the percentage reached at least once as well as the percentage of children who have never received these vaccines). Access should be quite high in most settings nowadays. It is worth looking carefully at the clusters where none of the children in the survey received DPT1/BCG. Which health facilities serve those clusters, and how is it possible that so many children in the sample were not vaccinated?
* Dropouts between the first vaccine and the last vaccine to be provided (for DPTCV1 and MCV1) may be an indicator of the EPI’s capacity of the EPI to follow-up with each birth cohort (and of utilization). What was the dropout between first and final doses of multi-dose vaccines? What were the likely reasons? Have dropout rates fallen since the last survey? If dropout rates are still high (for example, above 10 %), what strategies are needed to ensure that all children who start the vaccination series complete it? The data can provide clues about where programme problems may lie. Infant death rates and migrations influence the dropout rate, but so does the dissatisfaction of the families with the programme (irregular vaccination sessions, lack of information, fever or abscess following vaccination, etc.).

### Quality of recording and reporting vaccinations

Sometimes the data suggest that the issue is not necessarily with administration of the vaccine, but with reporting when vaccines were administered and to whom. Questions to consider:

* Is there an important difference between survey coverage results and administrative reports, and does this vary by stratum? The data may help indicate potential problems with the numerators, denominators or both used in administrative estimates.
* What proportion of individuals in the survey had a home-based record available, and has this improved since previous surveys? What proportion said that they had received one but did not have it available at the time of the survey? What are the likely reasons for lack of cards (such as stockouts, failure to emphasize their importance, caretakers not keeping them after the vaccination series has been completed)?
* How well were vaccination records completed? What proportion of records had dates that were outside the valid range (for example, date of birth after date of DTPCV1)? What proportion had illegible or missing dates (for example, tick marks instead of dates)?
* What proportion of children’s vaccination records was located in health facility records? For those not located, what were the likely reasons (for example, migrants, poor storage of health facility records, stockouts)?
* What was the quality of health facility records? How many illegible entries or out-of-range entries were found?
* Depending on survey design, health facility records may have been sought for all children, or only for children who did not have a home-based record. If sought for all children, how did data from health facility records compare with those from home-based records, and what are some likely reasons for discrepancies?
* Are dropout rates in the survey similar to those reported from administrative data? If routine reports show much lower dropout rates than survey results, investigate how well health workers are recording each dose of vaccine. Sometimes they may intentionally record the first or second dose of DTPCV1 incorrectly as the third dose, because they know that the third dose is monitored more closely.

### Invalid doses and timely encounters

Many problems with low vaccination rates can be corrected by better adherence to the vaccination schedules and standards.

* The gap between the crude and the valid data figures from records is often due to doses given too early, making them invalid. National programmes must implement the WHO recommendations for minimum ages at each dose and intervals between doses. The reasons for early doses may include inadequate screening for age by the EPI staff (for example, no card or no date of birth on the card) or ignorance of the strict vaccination schedule. The gap between the crude and the valid data shows what the performance might have been if the staff had followed the recommendations more closely.
* The analysis of *missed opportunities for vaccination during vaccination sessions* documents the scenario of each encounter between a child and the EPI team. It looks at the date of each dose received and whether the child was eligible to receive any other doses on that date (for example, whether the child was eligible for BCG at the time he received DPTCV). If the child does not receive all the vaccines he was eligible for, it is a missed opportunity. If the missed dose was given later it is a corrected missed opportunity; otherwise it is an uncorrected missed opportunity. The analysis provides information on the screening capacity of the EPI team and on vaccine stock management, but should also provide an opportunity to probe about possible misunderstanding from the staff on the so-called dangers of multiple injections on the same day, or misperceptions about vaccinating sick children (that is, implementing false contraindications for vaccination).
* Vaccinations should be provided as early as possible to protect the child before exposure to the infection. The percentage of children fully immunized by 12 months is one indicator of the timeliness of vaccination. Comparing ages at vaccination with the recommended schedule (using, for example, histograms) provides more detailed information on timeliness. Delayed vaccination can be caused by ignorance of the vaccination calendar, missed opportunities, ignorance of the need for the child to receive all recommended doses of a vaccine, EPI not informing the mother when to return, irregular vaccination sessions, breakdown in the vaccine supply, or temporary migration for cultivation or cattle rearing.
  + Note that although it is best to give vaccines as early as the schedule allows, it is better to complete the schedule, even at a later age, than to leave the child unvaccinated or under-vaccinated. For example, MCV1 should be given at age 9 months in countries where measles is still endemic, but if a child is seen at a health facility after 12 months of age and has not yet received MCV, it is important to administer the vaccine at that opportunity. Thus, when presenting these data, take care not to suggest that it is wrong to administer vaccines at older ages, but rather emphasize that it is even better to administer them close to the scheduled age.
* Interruptions in delivery for a given vaccine can be documented by the distribution of doses by calendar month. It should be more or less evenly distributed. Variations might be explained by: seasonal distribution of births, supply breakdown, inaccessibility due to weather (for example, monsoon), or absences of health workers due to illness, training workshops, or other interventions such as SIAs. It is also useful to compare the patterns between vaccines that are scheduled to be administered at the same time (for example, DPTCV and polio; DPTCV and PCV). If differences are observed, they might very well be due to stockouts and supply problems.

### Evaluating supplementary immunization activities

Supplementary immunization activities (SIAs), also called vaccination campaigns, are used regularly to improve immunity to vaccine-preventable diseases. This is currently the case with polio, or measles, or measles-rubella campaigns. Managers are encouraged to learn the campaign results and use survey results to inform decisions on when to go through and vaccinate those who were not vaccinated during the SIA, and over what geographical area. A post-campaign survey should include questions on whether children had received a dose of the relevant vaccine(s) in the routine programme, so it can highlight areas where the routine programme is weak. Those areas can be targeted for extra action after the survey.

If clusters are identified with alarmingly few children vaccinated in the survey sample (for example, zero or one), officials should be notified that there may have been an important campaign failure in that area, and follow-up investigation is warranted.

## Revise the report and obtain clearance on final draft

A draft report should be prepared as soon as possible after the survey ends, and presented to national authorities (and, if possible, to health authorities in each stratum of the survey). Often, when presenting results, additional issues will be raised that will lead to fuller discussion of the results and their implications. The report should be updated accordingly, and the final report submitted to all relevant institutions. It will be necessary to obtain clearance on the final report from the ministry of health before distributing it widely.

## Share the results

Although the coverage survey results might initially seem to be a technical subject only, in practice they can become political and sensitive, and should be approached as such. The survey organizers should be aware that survey results could sometimesbe perceived as an assessment of the performance of the specific programmes implementing the SIA, and indirectly of the institutions (EPI, ministry of health), and potentially of the government and the political parties in power.

The survey is not simply an academic exercise or a formal requirement for international donor and technical agencies. Rather, it produces data that could be used to improve the EPI at each level. Therefore, it is essential that each level of stakeholders understands the implications of the results, and how they can facilitate corrective actions. Because of this, it is important to think through how you will share the results. Below are some steps to consider taking as you plan to share the survey results.

### Choose the key messages

There are usually a few main themes that emerge from the data. Create messages based on these themes. Consider the survey goals and the political context as you create messages.

Phrase the conclusions and the recommendations of the report in objective, moderate terms, stating the facts and their meaning. The general tone should be not to blame but to emphasize progress, while documenting the possible reasons for slow progress and opportunities for improvement. The recommendations should stem practically from an interpretation of the results that would have been expressed in operational terms. For example, a DTPCV first dose coverage of 87% will be interpreted as 13% of children not being reached, and the recommendation will be to document the profile of these children in order to reach them in the future.

Ideally, the following will be true when it is time to share the results:

* The organization of the survey (including the selection of a contractor) has been requested and approved by the Ministry of Health at the beginning of the process.
* The reliability of the data collection and of the data processing have been checked and documented thoroughly.
* The interpretation of the data has been peer reviewed at least by the following: the steering group, the survey coordinator, the statistician, the EPI director, and senior members of the ministry of health.

If any of these has not been done, you will need additional talking points to address what happened and why.

### Select audiences

Consider who needs to learn about the survey results and how best to communicate them. Priority should be given to the policymakers, but also to the local EPI managers who will take corrective action. The goal of the survey is to provide actionable information to the EPI managers at different levels to take corrective action. It is essential that the presentation of the results and their implications go beyond a national presentation to multiple administrative levels (provinces, districts, etc.). The level that can make the most impactful changes based on survey results is probably the lowest statistically valid stand-alone level, the strata at the lowest level of administrative hierarchy.

It is also important to give feedback to all relevant partners (which could be done in a meeting of the interagency coordinating committee), including health facility workers, other providers in the area and senior health officials. Feedback should be ideally provided soon, ideally within one month, and is most effective if provided through meetings or newsletters. Feedback helps make health facility staff feel that they are an important part of the vaccination services, thereby increasing their motivation. Communities covered by the survey should also receive feedback, presented in ways appropriate to a lay audience.

The survey budget should include the costs of workshops designed for the EPI manager or deputies and the local EPI staff. During these sessions, attendees will discuss the probable causes of any weaknesses or incomplete performance, and identify corrective actions along with their costs and timetables.

If there are topics for which the team does not have enough information, the EPI should do focused operational research. For example, they could look into health facility studies of missed opportunities and their causes, or conduct focus groups to assess community attitudes and knowledge. In any case, once completed, a survey is likely to translate in to strategies for improvements, and such strategies cost money. At the time of evaluating the survey and its financial feasibility, also investigate the availability of funds to implement its recommendations.

### Create communication pieces and presentations

Prepare other forms of communication in addition to the survey report. The purpose of the survey, ultimately, is to improve vaccination practices and vaccination rates, so it is essential to communicate with district health offices so they can learn from the survey results and improve practices.

Use slide presentations for feedback workshops, and create a brief summary of the national survey results, as well as province- or district-specific results and recommendations, for all districts. You can also print four or five of the key charts and tables (for example, coverage by the time of the survey, by stratum), histogram of age at DTPCV1 and MCV1, and tables summarizing crude and valid coverage and missed opportunities), with bullet points showing their implications for action. These can be used for widespread distribution to health workers.

The high officials attending the presentation meeting should be provided the report or an executive summary in advance, to give them an opportunity to voice any questions and receive satisfactory answers before the beginning of the meeting.

### Contribute micro-data, documentation, and code to the Coverage Survey Repository

Coverage data and documentation of survey methods will be of greatest use if they are made freely available via the Internet, as is done for DHS and MICS. Details of survey methods (including the codes used to analyse the data) should accompany the micro-data of the survey. See the American Association for Public Opinion Research Transparency Initiative for more information on this issue ([www.aapor.org/AAPORKentico/Transparency-Initiative.aspx](http://www.aapor.org/AAPORKentico/Transparency-Initiative.aspx)).

References

Brogan D, Flagg EW, Deming M, Waldman R (1994). Increasing the accuracy of the Expanded Programme on Immunization's cluster survey design. Annals of epidemiology. 4(4):302–311.

Brown LD, Cai TT, DasGupta A (2001). Interval estimation for a binomial proportion. Statistical Science. 16:101–133.

Cochran, WG (1977). Sampling techniques, third edition. New York: John Wiley and Sons, Inc.

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).

Dayan, GH, Shaw, KM, Baughman, AL, Orellana, LC, Forlenza, R, Ellis, A, Chaui, J, Kaplan S, & Strebel, P. (2006). Assessment of delay in age-appropriate vaccination using survival analysis. Am J Epidemiol. 163.6:561–570.

Driscoll J, Lidow, N (2014). Representative surveys in insecure environments: a case study of Mogadishu, Somalia. J Surv Stat and Methodol. 2:78–95.

Fleiss JL, Levin B, Paik MC (2013). Statistical methods for rates and proportions. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Grais, R. F., Rose, A. M., & Guthmann, J. P. (2007). Don’t spin the pen: two alternative methods for second-stage sampling in urban cluster surveys. Emerg Themes Epidemiol, 4(8), 1-7.

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).

Heeringa SG, West BT, Berglund PA (2010). Applied survey data analysis. Boca Raton, FL, USA: CRC Press.

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(2):253.

Hutchins SS, Jansen HA, Robertson SE, Evans P, Kim-Farley, RJ (1993). Studies of missed opportunities for immunization in developing and industrialized countries. Bull of the World Health Organ. 71(5):549.

ICF International (2012a). Demographic and Health Survey sampling and household listing manual. MEASURE DHS, Calverton, MD, USA: ICF International.

ICF International (2012b). Survey organization manual for demographic and health surveys. MEASURE DHS, Calverton, MD, USA: ICF International.

Kish, Leslie (1965). Survey sampling. New York: John Wiley and Sons, Inc.

Korn EL, Graubard BI (1998). Confidence intervals for proportions with small expected number of positive counts estimated from survey data. Survey Methodology. 24:193–201.

Levy PS, Lemeshow S (1980). Sampling for health professionals, vol. 14. Belmont, CA, USA: Lifetime Learning Publications.

Levy PS, Lemeshow, S. (2013). Sampling of populations: methods and applications. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Lohr, Sharon (2009). Sampling: design and analysis, second edition. Boston, MA, USA: Cengage Learning.

MacNeil, A., Lee, C. W., & Dietz, V. (2014). Issues and considerations in the use of serologic biomarkers for classifying vaccination history in household surveys. Vaccine, 32(39), 4893-4900.

Meyer SA, Kambou JL, Cohn A, Goodson JL, Flannery B, Medah I, Messonnier N, Novak R, Diomande F, Djingarey MH, Clark TA, Yameogo I, Fall A, & Wannemuehler K. (2015). Serogroup A meningococcal conjugate (PsA-TT) vaccine coverage and measles vaccine coverage in Burkina Faso—Implications for introduction of PsA-TT into the Expanded Programme on Immunization. Vaccine, 33(12), 1492-1498.

Payton ME, Greenstone MH, Schenker N (2003). Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? J of Insect Science 3.1:34.

Rao JNK, Scott, AJ (1979). Chi-squared tests for analysis of categorical data from complex surveys. Proceedings of the Survey Research Methods Section, ASA. 58–66.

Rao JNK, Scott AJ (1981). The analysis of categorical data from complex surveys: chi-squared tests for goodness of fit and independence in two-way tables. J of the Am Stat Assn. 76:221–230.

Rao JNK, Scott, AJ (1984). On chi-squared tests for multiway contingency tables with cell properties estimated from survey data. Ann of Statist. 12:46–60.

Rao JNK, and Scott AJ (1987). On simple adjustments to chi-square tests with survey data. Ann of Stat. 15:385–397.

Rhoda DA, Fernandez SA, Fitch DJ, Lemeshow S (2010). LQAS: user beware. Int J of Epidemiol. 39.1:60–68.

Schenker N, Gentleman JF (2001). On judging the significance of differences by examining the overlap between confidence intervals. Am Stat. 55.3:182–186.

Sridhar S, Maleq N, Guillermet E, Colombini A, Gessner BD (2014). A systematic literature review of missed opportunities for immunization in low- and middle-income countries. Vaccine. 32(51):6870–6879.

Valliant R, Dever JA, Kreuter F (2013). Practical tools for designing and weighting survey samples. New York: Springer.

World Health Organization (2005). Immunization coverage cluster survey: reference manual. WHO/IVB/04.23

World Health Organization (2011). Documenting the impact of Hepatitis B immunization: best practices for conducting a serosurvey. (<http://whqlibdoc.who.int/hq/2011/WHO_IVB_11.08_eng.pdf?ua=1>; accessed March 6, 2015)

1. In this manual, *vaccination* refers to the administration of antigenic material (a vaccine) to stimulate an individual's immune system to develop adaptive immunity to a pathogen. *Immunization* refers to the process by which an individual's immune system produces an immune response. Immunity can occur due to natural exposure to infectious agents or artificially through the administration of vaccine. Vaccination may not result in immunity, due to impotent vaccine (through exposure to heat or freezing), host factors, the child not receiving all doses of a multi-dose vaccine, the child receiving the vaccine before the recommended minimum age, the child receiving a subsequent dose of a multi-dose vaccine before the recommended minimum interval between doses, or the efficacy of the vaccine itself. This manual describes how to conduct surveys that measure the number of children vaccinated without making claims as to their immunological status or how that status was acquired. [↑](#footnote-ref-2)
2. It will be helpful for the survey steering group to review the latest vaccination schedule and discuss which vaccines to assess and whether recent changes or vaccine introductions will make the survey especially complicated. For example, if new home-based records or cards are issued that list new vaccines, then survey staff will need to be trained to read both the old and the new cards. [↑](#footnote-ref-3)
3. The definition of ‘fully vaccinated’ may vary from country to country, may vary over time, and it may include only a subset of all vaccines; make the definition clear from the very start of the project. [↑](#footnote-ref-4)
4. There are appropriate quantitative tests to evaluate whether an observed difference is *statistically significant* but further judgment will be needed to decide whether the differences are meaningful or *programmatically significant*. [↑](#footnote-ref-5)
5. Respondents who gave birth in the past 12 months are used for evaluating Td or TT coverage because this yields information about the most recent vaccination activities (that is, those that occurred within the past year) and the protection of the most recently born children and their mothers. Surveys that evaluate tetanus toxoid coverage usually involve interviewing women who gave birth in the last year, but might also include a selection of women of childbearing age regardless of when they last gave birth, if this group was targeted for Td or TT vaccination. [↑](#footnote-ref-6)
6. Coverage figures are proportions, and the confidence interval (CI) for a proportion is essentially symmetric when the proportion is near 50%, but it is skewed if the proportion is near 0% or 100%. In this document, the sample sizes are designed so both sides of the CI are smaller than the precision target. That is, if you select a sample size to yield ± 5% precision, both the shorter and the longer sides of the CI should be ≤ 5%. [↑](#footnote-ref-7)
7. Several readers have commented that some surveys will require even more time than suggested here, so use insight from other recent quality surveys in your country. [↑](#footnote-ref-8)
8. The 2005 reference manual always used a design effect of 2. In practice, the design effects observed in vaccination coverage surveys have often exceeded 2, so this manual recommends a more conservative value of 3 if there are 7 respondents per cluster, or 4 if there are 10 respondents per cluster. [↑](#footnote-ref-9)
9. <http://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/VarianceEstimation/intro.htm> [↑](#footnote-ref-10)
10. No judgment is implied about the relative accuracy of home-based versus health facility records. [↑](#footnote-ref-11)
11. This is needed only if the survey instructs interviewers to ask neighbours how many survey-eligible children live in a household, when no one in the household is at home. [↑](#footnote-ref-12)
12. Sometimes there is a so-called *grace period* where the dose can be administered up to 4 days early and still considered valid. This grace period may vary by country, and if it is to be used for survey analysis, it needs to be defined in the survey protocol. [↑](#footnote-ref-13)
13. When effective sample sizes are large and when coverage estimates fall between 20% and 80%, it does not make much difference which method is used. But for small samples, or coverage very near 0% or 100%, different methods yield different intervals. Consult a survey statistician if you want to explore options for less conservative intervals. [↑](#footnote-ref-14)
14. The list of vaccines and doses may need to be adjusted to fit the context of the survey. [↑](#footnote-ref-15)
15. The definition of ‘fully vaccinated’ varies from country to country. Specify this clearly in the analysis plan so it will be clear and in the survey report, document the definition clearly. [↑](#footnote-ref-16)
16. See earlier footnote on documenting the definition of ‘fully vaccinated’. [↑](#footnote-ref-17)
17. Adjust the list as appropriate for the schedule in the country being surveyed. [↑](#footnote-ref-18)
18. A high-quality analysis of missed opportunities depends very much on having a high-quality dataset of vaccination dates. Yet experience has shown that data entry clerks are more likely to make typographical errors when entering dates than when entering other types of data. It will be prudent to compare the dates on the photographs of home-based records and EPI registries with the dates in the dataset to evaluate the quality of the dataset. In order to ensure excellent data quality, it may be necessary to use photos of vaccination cards to confirm every date in the dataset. [↑](#footnote-ref-19)
19. If you want to conduct a simple comparison of unweighted properties of the sample (% of male vs. female children sampled) then it is permissible to use the traditional Pearson chi-square test. For most comparisons of survey outcomes, however, you will draw conclusions about differences in the populations, not the samples, so it will be important to use procedures like Rao-Scott chi-square that take the survey design and weights into account. [↑](#footnote-ref-20)
20. http://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/VarianceEstimation/intro.htm [↑](#footnote-ref-21)
21. Recall that we say informally that we are 95% confident that the true coverage falls within the 95% CI. We also say that we are 95% confident that true coverage falls somewhere above the 95% lower confidence bound, and we are 95% confident that the true coverage falls somewhere below the 95% upper confidence bound. [↑](#footnote-ref-22)