1 **Rapid Mobile Phone-based (RAMP) survey toolkit**

Designing a RAMP survey: technical considerations

November 2012

with support of

Norwegian **Red Cross**

www.ifrc.org
Saving lives, changing minds.

International Federation
of Red Cross and Red Crescent Societies

This work is licensed under the Creative Commons Attribution-Non-Commercial-ShareAlike 3.0 Unported License.

To view a copy of this license, visit:
http://creativecommons.org/licenses/by-nc-sa/3.0/

or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Acknowledgement

This publication was made possible through the financial and technical support of the Norwegian Red Cross.

© International Federation of Red Cross and Red Crescent Societies, Geneva, 2012

Copies of all or part of this study may be made for non-commercial use, providing the source is acknowledged The IFRC would appreciate receiving details of its use. Requests for commercial reproduction should be directed to the IFRC at secretariat@ifrc.org.

The mention of product names (brands of bed nets, drugs, mobile phones, mobile phone service providers etc.) does not constitute an endorsement for the products by any of the organizations or individuals involved in the development of this manual. They are mentioned to provide the users of this manual with information on the types of products that may be used in malaria control programmes and in RAMP surveys and where to obtain further information. None of the authors or reviewers has a monetary link to the private companies mentioned.

Photographs are copyright IFRC, unless otherwise indicated.

Cover photographs:
Small top left: © Melanie Caruso/IFRC
Small bottom left: © Fatima Frank/evalû, LLC
Large right: © Joel Selanikio/DataDyne

P.O. Box 372
CH-1211 Geneva 19
Switzerland
Telephone: +41 22 730 4222
Telefax: +41 22 733 0395
E-mail: secretariat@ifrc.org
Web site: www.ifrc.org

International Federation of Red Cross and Red Crescent Societies (2012), *Designing a RAMP survey: technical considerations*, Volume 1 of the Rapid Mobile Phone-based (RAMP) survey toolkit.

1229700 E 11/2012

RAMP - volume 1

Designing a RAMP survey: technical considerations

Strategy 2020 voices the collective determination of the IFRC to move forward in tackling the major challenges that confront humanity in the next decade. Informed by the needs and vulnerabilities of the diverse communities with whom we work, as well as the basic rights and freedoms to which all are entitled, this strategy seeks to benefit all who look to Red Cross Red Crescent to help to build a more humane, dignified, and peaceful world.

Over the next ten years, the collective focus of the IFRC will be on achieving the following strategic aims:

1. **Save lives, protect livelihoods, and strengthen recovery from disasters and crises**
2. **Enable healthy and safe living**
3. **Promote social inclusion and a culture of non-violence and peace**

Table of contents

Foreword 5
Acknowledgements 6
Glossary 7
Acronyms and abbreviations 11

1. An overview of the Rapid Mobile Phone-based (RAMP) survey 13
1.1 Introduction 13
1.2 Development of the RAMP survey methodology 14
1.3 RAMP design features 14
1.4 Testing of the RAMP concept and tools 16

2. About the RAMP survey toolkit 19
2.1 Purpose of the toolkit 19
2.2 The three volumes of the RAMP toolkit 20
2.3 Audience for Volume 1, *Designing a RAMP survey: technical considerations* 21

3. Survey objectives and what the survey will measure 23

4. Sample size calculation 27
4.1 RAMP surveys 27
4.2 Primary indicators to be measured 28
4.3 Desired precision 30
4.4 Effect of stratified/disaggregated analysis on sample size 32
4.5 Number of domains 32
4.6 Design effect 32
4.7 Calculating sample size needed for percentages/proportions/prevalence (for a single point estimate) 35
4.8 Increasing the number of statistical units and households needed to account for non-response 39
4.9 Number of clusters and number of statistical units and households to be sampled per cluster 40
4.10 Sample size considerations when comparing two surveys, domains or strata 41
4.11 Questions and answers on sampling for RAMP surveys 42

5. Selecting clusters, segments, and simple random sampling of households 47
5.1 Using population data for selection of PSUs/clusters 47
5.2 Next step: simple random sampling or segmentation 48
5.3 Target size of final segment and sub-segmenting 50
5.4 Simple random sampling of households 50

6. Options for data collection 53

7. Data analysis 57

7.1 Data analysis plan 57
7.2 Data cleaning and management 59
7.3 Analysing the data 60
7.4 Calculation of sample weights and selection probabilities 65
7.5 Implementation of the survey 65

8. The RAMP website 69

Annexes 71

A. Magnitude of potential sampling-related bias in a RAMP survey 71
B. The RAMP methodology compared to the gold standard and the original EPI survey methodology 72
C. Adjusting selection probabilities and weights 75
D. Adjusting selection probabilities at the PSU level by counting the total number of households in the PSU 81
E. Survey specifications: planning worksheet 84
F. Data analysis plan: table shells 85
G. Annotated bibliography: where to go for further written advice 92

© Benoit Carpentier/IFRC

Foreword

From the beginning, insecticide-treated mosquito bed net (ITN) distribution has been closely linked to innovative evaluation. The first integrated ITN campaign, carried out in Ghana almost a decade ago, was also the first use of handheld computers for public health evaluation in Africa. The resulting rapid, high quality evaluation had an immediate impact. Within six months of that evaluation, WHO and UNICEF had endorsed mass, free ITN distribution campaigns, laying the policy foundation for the global effort we see today. As the global ITN scale-up has progressed, intensive evaluation has created a knowledge base that allows informed decisions and best practice on delivery methods, hang-up strategies and methods of improving utilization. It is this thoughtful approach to programme planning that has maintained international, national and donor confidence in the overall effort.

A key to rapid implementation has been the high degree of country ownership of bed net programming. However, the overall evolution towards country ownership has not advanced as fast for evaluation as it has for other programme elements. While it is now commonplace for countries to plan, budget and implement sophisticated ITN delivery strategies, state-of-the-art evaluation usually requires external support for financing and technical expertise. A principal reason for this lack of local ownership of evaluation has been that the evaluation tools have not been appropriately adapted to the local requirements.

The three volumes making up the Rapid Mobile Phone-based (RAMP) survey toolkit are an effort to empower local ownership of evaluation. The toolkit is part of a vision that high quality evaluations should be able to be conducted using the expertise and resources available at the district level. It takes advantage of two technologies: locally-available mobile phones and EpiSurveyor (to be re-named Magpi) software which enables mobile phones to be data collection platforms. Importantly, it is simplified without being simple. It brings within reach of all programme managers state-of-the-art methodology in survey sampling and the use of mobile phones to collect data. With it, local programme managers will not only be able to respond to concerns expressed at national and local level, but also to questions posed by international donors.

During an early integrated measles vaccination/ITN campaign in northern Uganda, a doctor told me that if the measles campaign was successful he could "close the measles ward". He added that if the bed net campaign was equally successful at controlling malaria he could "close the hospital". The developers of the RAMP toolkit hope that enhancing evaluations will be one more step towards our ultimate goal of closing the hospitals.

Mark Grabowsky
Deputy Director, National Vaccine Program Office
Department of Health and Human Services
Washington, D.C.

Acknowledgements

The International Federation of Red Cross and Red Crescent Societies would like to thank all those who have contributed to the development of this manual. Principal authors were Mac Otten and Tara Neville, with support from Jenny Cervinskas and Jason Peat. Mark Grabowsky (Centers for Disease Control and Prevention and American Red Cross) was a member of the group that provided the initial vision for the manual. Thom Eisele, Adam Bennett, and Albert Kilian reviewed several sections in the publication and provided valuable comments. Editing by Vivienne Seabright.

The hard work and dedication of staff and volunteers in Kenya Red Cross Society, Namibia Red Cross and Nigerian Red Cross Society and their branch offices in Malindi district, Kenya, Calabar, Cross River State, Nigeria and Katima Mulilo, Caprivi region, Namibia respectively, are much appreciated. Their contributions to the four pilot RAMP surveys conducted in 2011 and 2012 were key to the comprehensive field testing of the RAMP methodology.

Glossary

Accuracy: The average difference between sample estimates and the true population value. Accuracy is a function of both precision and bias as defined below.

Bias: Bias refers to the difference between the estimated value of a health indicator from a survey and the true value found in a population. A method of estimation is ***unbiased*** if the average value of the estimate taken over all possible samples of a given size is equal to the true value. The effect of bias reduces the accuracy of the results, which is a function of both precision and bias. The most important issue of bias in epidemiological research is the fact that direction and magnitude of bias are typically unknown, thus rendering interpretation of biased survey results extremely difficult. Increasing the sample size cannot reduce bias. Surveys can have both sampling-related bias and non-sampling bias. This manual discusses mostly sampling-related bias.

Census: A census describes one method for collecting data about a population in which every member of the population is questioned to gain information about characteristics of interest. This type of data collection can take a great deal of time and money to complete.

Cluster sampling: Statistical units (e.g., persons) can be grouped, for example, by areas, districts, schools, etc. Cluster sampling is sampling in which groups of statistical units (such as areas, districts, schools or villages) are chosen in a first stage, often using a method called probability proportional to estimated size (PPES – see below). The first-stage units are called primary sampling units; they are also called clusters.

Confidence interval: The confidence interval describes a range of values expected to include the true value of the characteristic or parameter being evaluated if the survey were to be repeated many times. The confidence interval provides information about the precision of the survey estimates. The confidence interval and precision are closely related: e.g., when precision increases, the confidence interval narrows (e.g., from ±10 per cent to ±5 per cent).

Confidence level: A percentage value that describes how likely it is that the true value of the characteristic being estimated will fall into the confidence interval defined, if the survey were to be repeated many times. The usual level of confidence is 95 per cent. For example, if the point estimate is 70 per cent and the 95 per cent confidence interval is ±5 per cent, then 95 per cent of the time the true value would fall between 65 and 75 per cent if the survey were to be repeated many times.

Design effect: The design effect compares the variability (variance or standard error) of the sampling method being used to the variability of using a simple random sampling method for the same sample size. Two types of design effect are in common use: DEFF and DEFT. The variance is used for DEFF and the standard error is used for DEFT.

- DEFF = variance of a cluster sampling design ($se_{design\ x}$) / variance of the same indicator estimated from simple random sampling (se_{SRS}).

- DEFT = standard error of a cluster sampling design (sedesign x) / standard error of the same indicator estimated from simple random sampling (se_{SRS}). It can be thought of as the ratio of standard error of cluster sample (or sampling design x) to simple random sampling. It is also the factor by which the sample size for a cluster sample would have to be increased to have sample precision equal to that for simple random sampling (i.e., the penalty for cluster sampling).

Domain weight: A weight that is used when more than one domain is included in the survey design and the sample size for each domain is not allocated proportionally based on domain size. For example, if the rural population is 60 per cent of the population and urban population is 40 per cent of the population and 300 households each were sampled in rural and urban domains, then design weights are needed to increase the weight of sampled rural households since their fraction of the total population is larger.

Enumeration areas (EA): Small areas used by government statistical bureaus for conducting national censuses. These areas can be used for constructing the sample frame for surveys.

Expanded Programme on Immunization (EPI) cluster sampling method: An older sampling methodology for conducting immunization surveys in which 30 primary sampling units (clusters) are chosen in the first stage, then a random direction chosen from the centre of the selected primary sampling units, followed by randomly choosing a house along the selected direction, then going to the nearest next door until seven children within a target age range are found.

Equal probability survey: Survey in which all statistical units have an equal probability of being selected (therefore, the survey weight is the same for all statistical units). A RAMP survey has an equal probability design, but the weights are not equal because they are adjusted for non-response, and, optionally, weights are adjusted to account for inaccurate estimates of PSU size.

GPRS (General Packet Radio Service): The most basic type of cellular service that can carry data. It enables downloading of programmes to mobile phones and uploading of data (such as e-mail).

Multi-stage cluster sampling: Similar to (single-stage) cluster sampling defined above, groups of statistical units (such as provinces, districts, schools, etc.) are chosen in two or more stages. For example, 30 of 100 districts can be chosen using probability proportional to estimated size (PPES – see below) sampling during the first stage, then five of 25 schools in each district could be chosen by PPES during the second stage. Then all students in the five schools would be interviewed for the survey.

Non-sampling error: Non-sampling errors are one source of bias in a survey. Examples of non-sampling errors are those caused by response errors, non-response, faulty questionnaires, interviewer recording errors, data processing errors and others. These errors lead to bias in survey results. The extent of non-sampling errors in a survey is very difficult to estimate, but can be minimized with careful survey preparation.

Point estimate, estimate: Analysis of an indicator in a survey provides two important items: (1) point estimate, (2) 95 per cent confidence interval. The point estimate is a proportion, percentage, or (point) prevalence. The point estimate can be referred to as just an “estimate”. However, survey specialists

and epidemiologists often refer to indicator results from a survey as "point estimates" since they refer to a single point in time.

Precision: In practical terms, precision is the width of the confidence interval for a health characteristic measured in a survey. The precision of survey results can be improved by increasing the sample size. Practically, precision is often described as "plus or minus x per cent" (for example, plus or minus 10 per cent), "x" being one-half of the width of the confidence interval.

Primary sampling unit (PSU): PSU and cluster have the same meaning in this manual. See *Cluster sampling* above.

Probability proportional to estimated size sampling (PPES): This type of sampling at the first and possibly later stages ensures that those primary or secondary sampling units that have a larger size have a larger probability of being selected for sampling. The probability of the sampling unit being selected is equal to the estimated size. The word "estimated" is used because the size is usually estimated and not known with absolute precision. For example, the population size of a district is usually estimated because data come from a census that was carried out more than one week ago (e.g., six months to ten years ago).

Proportion, Percentage, and Prevalence: In this manual, proportion, percentage, and prevalence are used interchangeably. Proportion is a more formal term, expressed as a decimal (0.65, 0.30), and is used in the definitions of the Roll Back Malaria (RBM) Monitoring and Evaluation Reference Group (MERG) malaria indicators. Percentage is the term in more common usage, e.g., percentage of households with at least one ITN, percentage measles immunization coverage, etc. Prevalence can be expressed as a percentage or proportion. Prevalence is used in this manual in the sample size calculations (expected prevalence). "Expected prevalence" could as easily been written as "Expected percentage" or "Expected proportion", but "Expected prevalence" is the most common phrase used in the language of sample size calculations and calculators. Formally, the prevalence used in the sample size calculations is a point prevalence.

Sample: A survey is a method of collecting information about a population which involves gathering data from only a part of the population and estimating from the results what is occurring in the entire population. The part of the population that is selected is referred to as the sample.

Survey design: The survey design refers to the set of specifications that describe, in detail, the target population, sampling frame, the sampling units, the sample size, the sample selection and the indicators to be used for the survey.

Sampling error: The sampling error refers to the error in survey estimates that occurs because the entire population was not included in the sample. In practical terms, each time the survey is conducted, the results may not be exactly the same.

Sampling frame: The sampling frame is a complete list of all the sampling units that cover the target population. It is a critical element in the quality of the survey results. In cluster sampling, the sampling frame is the list of all clusters or primary sampling units from which a sample is drawn during the first stage.

Sampling units: The target population is divided into parts called sampling units. Sampling units are a set of units used for selection at each stage of sampling.

Statistical units: A statistical unit is an entity about which information is sought and for which statistics are ultimately compiled. It is the unit at the basis of statistical aggregates and to which tabulated data refer[1]. For example, if information is collected and analysed about children under five years of age during the household interviews, then children under five would be considered statistical units, even though parents would be reporting the information. RAMP malaria surveys have three types of statistical units: persons, households and nets. The type of statistical unit matches the denominator of the indicator being measured, therefore, a single survey can have several types of statistical units (e.g., persons, households, bed nets). Survey "elements" comprise the same concept.

Stratified/Disaggregated analysis: Analysis of data by a division of the population into levels, classes, or strata according to a specific characteristic, such as age, socio-economic status or rural-urban status. For example, when dividing by gender, male and female would be different strata. When dividing by educational status, those who attended secondary school would be one stratum.

Survey domain: A stratum for which independent and precise estimates are desired within the survey population (e.g., urban and rural).

Survey population: The target population (defined below) that can be reached, taking into account practical constraints. For example, all of the villages in the target population (desired sampling frame) may not be reached because of flooding, insecurity, etc. Those villages that can be reached are the survey population. The survey population is also known as the survey universe. For example, the survey population might be defined as all people in accessible districts that can be reached in the rainy season and within the budget of the survey.

Target population: The term target population refers to the entire population about which information is required. It can be thought of as the population that is the ideal one for meeting a survey's measurement objectives (e.g., proportion of children under five years old with a fever in the past two weeks).

Wealth quintile: Division of households into five equal groupings by the number and type of household assets that they have (e.g., water supply, electricity, vehicle).

1 *United Nations Department of Economic and Social Affairs Statistics Division. Statistical Units. United Nations, New York, October 2007.*

Acronyms and abbreviations

CDC	Centers for Disease Control and Prevention (USA)
DHS	Demographic and Health Survey
EA	Enumeration area
EPI	Expanded programme on immunization
GPRS	General Packet Radio Service
IFRC	International Federation of Red Cross and Red Crescent Societies
IRS	Indoor residual spraying
ITN	Insecticide-treated net
LLIN	Long-lasting insecticide-treated net
MERG	Monitoring and Reference Group of Roll Back Malaria
MoH	Ministry of Health
NGO	Non-governmental organization
PPES	Probability proportionate to estimated size
PSU	Primary sampling unit
RAMP	Rapid Mobile Phone-based survey
RBM	Roll Back Malaria
SRS	Simple random sampling
WHO	World Health Organization

© Benoit Carpentier/IFRC

01.

An overview of the Rapid Mobile Phone-based (RAMP) survey

This section of the manual gives a brief overview of the development and testing of the RAMP survey methodology, its main design features and most appropriate use. It will be most useful to the health programme manager responsible for contemplating a survey, as well as the survey coordinating group, the survey coordinator and those responsible for designing the survey.

1.1 Introduction

Monitoring and evaluation of health interventions in populations is a critical aspect of the work of Red Cross Red Crescent National Society health programme managers, as well as health programme managers in non-governmental organizations (NGOs) and government departments. Accurate information about the effect of health programmes allows programme managers to help advance the health status of the population by:

- assessing health outcomes at the population level
- evaluating intervention strategies
- improving efficiency in the delivery of health programmes
- making the best decisions in terms of spending
- meeting reporting requirements for funding

A health survey is one of several ways to gather information about health interventions. Data are collected, organized and interpreted so that meaningful conclusions can be made and action taken. A sample of the population is surveyed in order to estimate what is happening in the whole population.

Data collection, management and analysis can be performed in many ways. Traditionally, in health surveys, paper questionnaire forms are filled out in the field and the data are summarized and entered into a computer for analysis. The rapid spread and use of mobile technology throughout the globe, however, offers

a new means of data collection. The widespread availability of mobile technology means that paper questionnaires can be stored on mobile phones and used directly by the field team to record responses. One of the main benefits of using mobile phones is that the data can be downloaded and analysed very quickly.

1.2 Development of the RAMP survey methodology

Between 2003 and 2008, the International Federation of Red Cross and Red Crescent Societies (IFRC), in collaboration with epidemiologists from the Global Immunization Division at the Centers for Disease Control and Prevention (CDC) and DataDyne, an organization specializing in data collection using handheld devices, worked on a project for undertaking health surveys using mobile devices for data entry, using the Expanded Programme on Immunization (EPI) survey methodology concepts. In 2008, personnel from the Global Malaria Programme at the World Health Organization (WHO) joined the effort.

During the development of the RAMP survey methodology, it was recognized that methodological improvements could be made to the EPI survey method. In 1994, Dr Donna Brogan, a professor of Biostatistics and international survey expert at the Emory School of Public Health, United States, and her Emory University and CDC colleagues published a paper suggesting ways to improve the EPI survey method using more standard survey techniques[2]. During the development process of the RAMP methodology, many of the methods suggested by Dr Brogan were incorporated.

1.3 RAMP design features

The RAMP survey methodology was designed to support management decision-making. Features include:

- simplicity of sampling design to facilitate survey implementation and reduce field costs
- use of standard survey sampling methods
- web-based questionnaire design
- questionnaires that can be easily downloaded to standard mobile phones
- data collection using low-cost, familiar and widely-available mobile phones
- real-time web-based dataset that can be easily exported for rapid analysis and reporting purposes
- availability of tools to guide survey planning and implementation

The RAMP survey is a two-stage cluster survey with clusters/PSUs as the first stage and households as the second stage with the following characteristics:

- selection of clusters using probability proportionate to estimated size (PPES) (similar to most other cluster surveys)
- segmentation of clusters into smaller segments by PPES if there are too many households for simple random sampling (SRS)
- simple random sampling of a fixed number of households in a selected smaller segment
- adjustment of sample weights for non-response
- optional adjustment of sample weights for imprecise cluster/PSU size

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

Possible very small residual sampling bias related to imprecise estimate of size of segments is discussed in more detail in Annex A.

A RAMP survey is flexible. It can be carried out at different levels (e.g., national, provincial, district or sub-district) to generate results. A RAMP survey may be especially pertinent to guide informed, decentralized decision-making, using local information. District health teams and managers of development projects should consider using a RAMP survey.

The survey can be used in a variety of health and development sectors to address a number of information needs. A RAMP survey would be appropriate, for example, for the following applications:

- surveys where rapid results are key
- surveys where cost is a significant issue
- sub-district surveys involving multiple villages that are typically carried out by non-governmental organizations (NGOs)
- national, provincial or district-level surveys carried out by Ministry of Health (MoH) or government departments (immunization, maternal and child health, malaria, nutrition, HIV/AIDS, water and sanitation, etc.) to collect intervention coverage data quickly and at low cost
- district-level surveys conducted by the district health management team
- baseline and endline surveys for donor-funded projects/programmes
- repeated surveys to track time trends for key indicators

Examples might include:

- surveys to estimate the percentage of households that were visited by community-based volunteers to discuss the care and repair of mosquito nets
- surveys to estimate the percentage of households that are receiving clean water
- surveys to estimate the percentage of six year old female children that are attending school

A primary objective of the RAMP survey method is to enable NGOs and health ministries in countries in the process of development to both collect and analyse survey data with updated modern tools, without external technical assistance.

There are, however, two main types of survey for which a RAMP survey may not be the first choice. First, in very long and complex surveys with a multitude of skip patterns, paper-based data recording may be superior to mobile phone-based data recording. For example, large (more than 5,000 households) and complex (questionnaires of more than 20 pages in length), multi-sector, expensive surveys that are carried out approximately every five years, such as the Demographic and Health Surveys (DHS), may not benefit from using the RAMP methodology. Second, the RAMP approach may not be the first choice for surveys where survey planners feel that a simple random sample (SRS) of households should be done at the second stage involving all households in the cluster, regardless of the size of the cluster. For example, planners of research studies may insist on SRS even in clusters with 200 or more households.

1.4 Testing of the RAMP concept and tools

Once developed, and a draft user manual and training programme produced, pilot RAMP surveys were carried out in three African countries, Kenya, Namibia and Nigeria in 2011 and 2012. The Red Cross National Societies at headquarters and branch levels played a leading role in the surveys, and Red Cross volunteers were recruited and trained to collect the data in the field survey. There are many public health problems in Africa that could have been chosen to pilot the surveys. However, malaria was selected to test the RAMP tools.

The IFRC created a set of standard survey questionnaires based on RBM's Malaria Indicator Survey[3] (see the RAMP website: www.ifrc.org/ramp, for the most up-to-date malaria questionnaires).

The pilot surveys in Africa established conclusively that National Societies can be a core partner in leading a RAMP survey, with community-based volunteers able to collect data using mobile phones, and the results being available within days of the last interviews in the survey. Lessons learnt from the pilot surveys have been used to refine the RAMP survey methodology and tools, and to provide sample materials in the RAMP toolkit.

© Fatima Frank/evalû LLC

3 *Roll Back Malaria, MEASURE Evaluation, MEASURE DHS, USAID, UNICEF, World Health Organization, PATH-MACEPA, US Centers for Disease Control, World Bank. 2012.* Household Survey Indicators for Malaria Control. *MEASURE Evaluation: Calverton, MD. See: www.rbm.who.int/mechanisms/merg.html.*

© Karen Bramhill/IFRC

02. About the RAMP survey toolkit

This section describes the three volumes that make up the RAMP toolkit, and their different purposes. It will be most useful for the health programme manager, the survey coordinating group and the survey coordinator, to understand the most appropriate use of each of the volumes.

2.1 Purpose of the toolkit

The RAMP survey toolkit is specifically designed to be a management tool for Red Cross Red Crescent National Society health programme managers to achieve their management priorities. It is meant to be a comprehensive resource to empower National Societies to oversee the planning, implementation and analysis of health and development surveys. The toolkit will help National Societies conduct simple, low-cost surveys using standard sampling methods with no or limited external technical assistance. The RAMP method may also be used by government, ministry of health programmes, NGOs and others interested in monitoring and evaluating health and development programmes.

There are three primary purposes of the RAMP survey toolkit. The first is to improve on the sampling methods of the original EPI cluster methodology. The second is to add mobile phone data collection, helping to reduce the time to availability of results to one to three days, rather than the more common three months or more. The third is to reduce the need for external technical support for the design, conduct and analysis of surveys.

The practical example shown in the toolkit is a survey to measure ownership, usage and hanging rates of long-lasting insecticide-treated nets (LLINs) following a national or sub-national mass distribution campaign. Included are a number of resources to aid the carrying out and completion of the tasks required.

EpiSurveyor (now named Magpi) was chosen to illustrate the RAMP survey processes. It is a freely-available suite of tools that can be used to design survey questionnaires, collect data in the field and transmit data to an internet database using low-cost mobile phones.

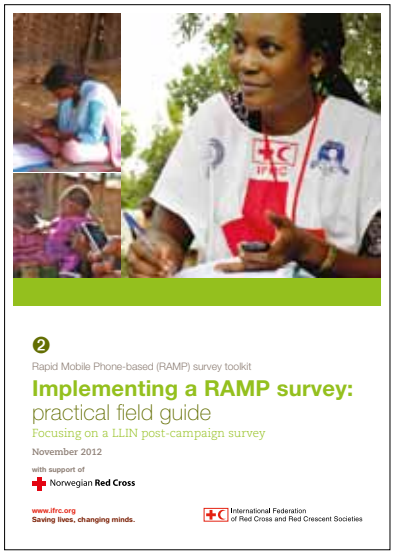
The RAMP survey toolkit, in addition to the three volumes described in Section 2.2, has a website (www.ifrc.org/ramp) where the latest version of many tools, forms and information can be found. See Section 8 of this manual for more details of the contents of the website, which will be updated on a regular basis to serve as a dynamic resource.

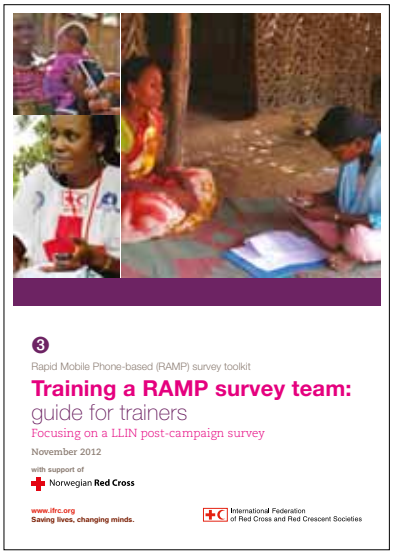
2.2 The three volumes of the RAMP toolkit

Designing a RAMP survey: technical considerations is the first volume in a set of three linked publications making up the RAMP toolkit. It discusses the technical aspects of survey design, sampling decisions, indicators to be measured, data collection and analysis, and the approach taken by a RAMP survey. The second volume, *Implementing a RAMP survey: practical field guide,* is a comprehensive resource for the organizers of the survey, giving information on timings, field logistics, staffing, field procedures, data management in the field, and reporting.

Neither *Designing a RAMP survey* nor the practical field guide is intended to be used directly by the implementing field survey team. The third publication in the toolkit is a training manual, *Training a RAMP survey team: guide for trainers,* designed as a main resource for preparing trainers to facilitate the training of the interviewers and supervisors who will carry out the RAMP malaria survey in the field. The training manual is fully compatible with the RAMP survey methodology described in the technical manual and field guide. In all three volumes, the same example of a RAMP malaria survey is used: a survey to measure ownership, usage and hanging of insecticide-treated bed nets, following a mass distribution campaign. Although a malaria example is used, the tools in this toolkit can be used to measure most other types of health or development indicators.

The three publications together provide a methodology, operations protocol and numerous tools to carry out a RAMP survey. This is meant to ensure that health surveys will be conducted at reduced costs, in a timely fashion, and with limited external technical assistance. Users of the manual should make necessary adaptations from the manual to suit the specific characteristics of their survey.





2.3 Audience for Volume 1, *Designing a RAMP survey: technical considerations*

The primary users of this technical manual will be programme managers and decision-makers considering whether to conduct a survey to gather health or development data. Once it has been decided to conduct a survey, this manual will help programme managers, the survey coordinator and the data manager/data analyst with the design, planning and the technical aspects of the survey. Members of the team may concentrate on specific sections of the manual that are relevant to their role, responsibilities and interests. For example, details regarding the survey sampling approach that is specific to a RAMP survey are provided and will be of most relevance to the survey leaders and the data manager/data analyst. It is expected that the survey coordinator would benefit by becoming familiar with all aspects of the RAMP methodology.

Besides those actively involved in a RAMP survey, the manual can also be relevant to others, for example as a reference guide for planning and carrying out population-based surveys. Organizations and individuals might find specific content and tools helpful and flexible for adaptation to their objectives and local situation.

Designing a RAMP survey is organized into eight sections, each covering an important aspect of survey design, plus annexes that provide highly detailed information that might be of particular relevance to technical specialists, such as sampling specialists or data analysts who are involved in the design of the survey, the selection of the sample and data management.

03. Survey objectives and what the survey will measure

Section 3 describes the setting of the survey objectives and the indicators used to measure them. The example in the manual is a survey to measure bed net use in a population following a mass LLIN distribution. This section will be most useful for the programme manager, survey planners, the survey coordinating group and the survey coordinator.

The first step in planning a health survey is to prepare a statement of its objectives. Clear objectives help to explain the survey to national authorities, interested partners, the overall survey team, including the field team, and the target population. Without a clear statement of objectives, it can be easy to get distracted by the details of the planning and make decisions that will not provide reliable results[4].

Survey objectives should include statements about the following:

- why the survey is needed
- what the survey will measure
- target population
- survey population
- stratification/survey domains
- desired precision for all point estimates
- who will be responsible for overseeing the survey
- the timeframe to complete the survey and analysis
- how survey findings will be used

One of the most important things to remember when considering survey objectives is to keep the survey as simple as possible. Only data to assess those indicators needed and intended to be used to inform programme development should be collected. This will increase the effectiveness and accuracy of the survey results[5].

4 *Cochrane W G (2007)* Sampling techniques, *3rd edition, A Wiley Publication in Applied Statistics. John Wiley &Sons.*

5 *Measuring Mortality, Nutritional Status and Food Security in Crisis Situations: SMART Methodology. Version 1, April 2006. See: www.smartindicators.org/SMART_Methodology_08-07-2006.pdf.*

Indicators describe what the survey will be measuring. In this manual, a survey to measure LLIN coverage in a population after a mass distribution will be used as an example. The main indicators for such a survey would be:

- ownership: proportion of households that own at least one insecticide-treated net (ITN)[6]
- access: proportion of persons (all ages) in a household with access to an ITN, assuming that one ITN covers two persons
- access: proportion of households with at least one ITN for every two household members
- use by persons: percentage of persons (all ages) in the household who slept under an ITN, measured as the percentage of persons who slept under an ITN the night before the survey

A supplemental indicator would be the percentage of ITNs that were used. This is measured as the percentage of the ITNs that were slept under by at least one person during the night before the survey.

It should be noted that the denominators are different for each indicator. This is important in the calculation of the sample size and analysis phase:

- for household ownership of at least one ITN, the denominator is *households*
- for use/hanging rates of ITNs, the denominator is *ITNs*
- for usage by persons, the denominator is *persons*

In addition the survey is designed to collect information about:

- the study population (total number of individuals in the household)
- household wealth
- the effectiveness of mobilization efforts

It is important to remember that the results will describe the sampling frame surveyed (e.g., nation, region, state, district) as a whole. You CANNOT compare health indicators between clusters or primary sampling units (e.g., villages, sub-districts, districts, enumeration areas).

6 *Most mass distribution campaigns distribute LLINs. Other insecticide-treated nets, however, whether treated during production or by the end user, also provide protection, although usually for a more limited length of time. When assessing coverage and use, the standard indicators include insecticide-treated nets (ITNs) of all types, not just LLINs.*

RUMAN
MUBI
BUKALO
PRIMARY
NSUNDANO
IMELDER
MUTENWA
POLICE
KATUPISA
HIGH LIFE
S.D.A
SHOPPING CENTRE
BUKALO H.C.
TO SCHOOL
SANJO
NJAI
TO KAB
TO HEALTH C.
MARKET
BUILD TOGATH
MATAPI
NGULWA
MULENENI
N.A.C
MUHONGO
REGINER
SHAMBWE
MASISHATI
MATENDU
MAMUNDIYA
KUTA
MAMULAKO
C.F.
MANALUCA
TO MULENENI
FOREST
SAKUTU
MASANG

04. Sample size calculation

Section 4 describes how to calculate the sample size in order to meet the survey objectives and adequately measure the indicators. This section will be of most use to the survey planners, survey design specialists and the data manager/data analyst. Some specialized technical content is contained in the annexes.

4.1 RAMP surveys

In general, the RAMP survey has a two-stage cluster design. At the first stage, clusters or primary sampling units (e.g., villages, sub-districts, districts, enumeration areas) are chosen from a list constructed by survey planners. Within clusters, a certain number of households are selected for interview. All the statistical units (e.g., households, persons, children, pregnant women) within those selected households are included in the survey. The methods used to select the clusters and households are described in Volume 2 of the RAMP toolkit, *Implementing a RAMP survey: practical field guide*.

In general, a RAMP survey involves 30 clusters and between 10 and 40 statistical units per cluster in a single domain. This section, however, involves calculating the exact number of statistical units and households needed, number of clusters, and number of households per cluster for sampling in order to meet the objectives of the survey. Survey planners also need to determine if they want more than one survey domain, for example, separate rural and urban domains.

Sample size of a RAMP survey will vary depending on the kind of survey that is to be carried out. In general, the larger the sample size, the more precise the survey results will be. The sample size chosen for health surveys is usually a balance between the precision required and the increased survey costs associated with obtaining a larger sample size.

This section discusses calculation of the sample size and some aspects of the survey design. A number of different factors and decisions have to be taken into consideration in the calculation of the sample size:

- primary indicators to be measured, expected prevalence, and the number of statistical units (persons) per household
- precision desired
- stratification/disaggregation analysis considerations
- number of domains (one or several, rural and urban, provinces, etc.)
- the design effect, which reflects the level of homogeneity of indicator measures within clusters

Once these have been factored in, the resulting sample size is adjusted upward to account for potential non-response. The adjusted sample size is determined in terms of statistical units (persons or households)[7] and number of households that contain the required number of statistical units, after which survey planners can decide on the number of clusters and households per cluster required.

Below, each factor needed in the calculation of the sample size is explained in more detail.

4.2 Primary indicators to be measured

The final sample size will come from a single indicator that needs the largest number of households to measure the indicator with specified precision. For example, if 300 households are needed to measure the proportion of households with sufficient ITNs, 400 households to measure ITN use in persons of all ages, and 1,000 households to measure the proportion of children under the age of five years with fever that had blood taken for malaria testing, then 1,000 households would be chosen as the final sample size for a survey that would measure those three indicators.

Surveys usually have more than 20 questions, which means that many indicators can be measured. Sample size planning is best conducted using only the main indicators, which should be listed. Three examples of main indicators are shown in the table, one for a malaria bed net survey, one for a malaria survey that includes questions about malaria testing and treatment, and the third for a micronutrient survey[8]. These main indicators and their target populations will be the examples used in this manual to illustrate the process of calculating the sample size for a survey.

Table 1: Main indicators by type of survey

Malaria survey: bed net indicators
1. ITN use during the previous night, persons of all ages
2. Households with at least one ITN
3. Households with sufficient ITNs to cover all members of the household
Malaria survey: malaria testing and treatment in children under 5 years old
1. Among children under 5 years old with fever in the last two weeks, proportion that had blood taken from finger or heel for testing
Micronutrient survey (abbreviated indicators)
1. Vitamin A capsule coverage in children under 5 years old
2. Anaemia in children under 5 years old
3. Households using iron fortified product
4. Households using iodized salt

7 *Note: if the statistical unit is households, then the number of statistical units and the number of households needed for sampling will be the same. However, if the statistical unit is persons of all ages, then the number of households needed to be sampled will be much fewer than the number of statistical units.*

8 *Adapted from Gorstein J, Sullivan K M, Parvanta I, Begin F (May 2007),* Indicators and Methods for Cross-Sectional Surveys of Vitamin and Mineral Status of Populations. *The Micronutrient Initiative (Ottawa) and the Centers for Disease Control and Prevention (Atlanta). See: www.micronutrient.org/CMFiles/PubLib/Indicators-for-Cross-Sectional-Surveys1IYA-3242008-2823.pdf.*

The main indicators should also be listed according to whether they are individual or household type indicators, so that the required sample size of households can be correctly calculated. The main indicators from a malaria bed net and micronutrient survey divided by individual and household type indicators are shown below as examples.

Table 2: Indicators by individual and household types

Malaria survey: bed net indicators
Based on individuals: 1. ITN use during the previous night, persons of all ages **Based on households:** 1. Households with at least one ITN 2. Households with sufficient ITNs to cover all members of the household
Malaria survey: malaria testing and treatment in children under 5 years old
Based on individuals: 1. Among children under 5 years old with fever in the last two weeks, proportion that had blood taken from finger or heel for testing
Micronutrient survey
Based on individuals: 1. Vitamin A capsule coverage in children under 5 years old 2. Anaemia in children under 5 years old **Based on households:** 1. Households using iron fortified product 2. Households using iodized salt

When the denominator of the indicator is individuals, a second step is needed. The number of households needed to be sampled to yield the number of individuals (statistical units) required must be calculated. For this step, one needs to know how many individuals among common target population groups are expected to be found in each household. Common target population groups might include the following:

- all persons
- children under five years old
- children under five years old with fever in the last two weeks (for malaria surveys in malaria endemic areas)
- women of child-bearing age
- pregnant women

The table below shows the average number of target persons (statistical units) per household. This example comes from malaria endemic areas in sub-Saharan Africa.

Table 3: Percentage of total population and average number of target persons per household for common target groups.

Target group	Percentage of total population	Average number of target persons (statistical units) per household (assume 5.0 persons/ household)
1	2	3
Children under 5 years old	17	0.85
Children under 5 years old with fever in the last two weeks	25% of 17 = 4.25 (assuming 25 per cent will have had fever in the last two weeks)	0.2125
Women of child-bearing age	25	1.25
Pregnant women	4	0.20

Column 3 = Column 2 (as a proportion or decimal) times 5.0 persons per household (e.g., children under five years old, 0.17 x 5 = 0.85).

The third column is important to calculate the overall number of households that need to be sampled to yield the estimated number of target population (statistical units) needed.

4.3 Desired precision

Survey planners must decide what precision is desired for each primary indicator to be measured. For many indicators, plus or minus 10 per cent is good enough for management decisions (plus or minus 10 per cent was the target of the original EPI survey). However, some indicators might require greater precision, such as plus or minus 5 per cent (e.g., percentage of children with anaemia or households using iodized salt).

The tables below show examples of desired precision for each indicator for surveys with a single domain. For this example, precision of plus or minus 10 per cent was chosen for the malaria indicators and precisions of plus or minus 3 per cent and plus or minus 5 per cent were copied from the micronutrient survey guidelines[9].

9 *Adapted from Gorstein J, Sullivan K M, Parvanta I, Begin F (May 2007),* Indicators and Methods for Cross-Sectional Surveys of Vitamin and Mineral Status of Populations. *The Micronutrient Initiative (Ottawa) and the Centers for Disease Control and Prevention (Atlanta). See: www.micronutrient.org/CMFiles/PubLib/Indicators-for-Cross-Sectional-Surveys1IYA-3242008-2823.pdf.*

Table 4: Target population and desired precision for selected main indicators

a) Malaria survey: bed net indicators			
Main indicators	**Denominator**	**Target population**	**Desired precision (±%)**
Based on individuals: 1) ITN use during the previous night, persons of all ages	Individuals within selected households	Individuals within households in target/ programme area	10
Based on households: 1) Households with at least one ITN 2) Households with sufficient ITNs	Sampled households	Households within target/ programme area	10 10

b) Malaria survey: malaria testing and treatment in children under 5 years old			
Main indicators	**Denominator**	**Target population**	**Desired precision (±%)**
Based on individuals: 1) Among children under 5 years old with fever in the last two weeks, percentage that received a finger prick or had blood taken	Children under 5 with fever in past two weeks	Children under 5 with fever in the last two weeks in target/programme area	10

c) Micronutrient survey (abbreviated list of indicators)			
Main indicators	**Denominator**	**Target population**	**Desired precision (±%)**
Based on individuals: 1) Vitamin A capsule coverage in children under 5 years old 2) Anaemia in children under 5 years old	Children under 5 years old within selected households	Children under 5 years old in target/programme area	3 5
Based on households: 1) Households using iron fortified product 2) Households using iodized salt	Sampled households	Households in target/ programme area	5 5

4.4 Effect of stratified/ disaggregated analysis on sample size

When estimating the sample size, it is sometimes important to consider stratified/disaggregated analyses that survey planners may want to perform after the data are collected (e.g., analysis by age group, gender, rural/urban status, by province, etc., often called *post hoc* stratification by survey specialists). If the survey planners are not concerned with achieving a specified level of precision for disaggregated point estimates, then the sample size would not be affected. For example, given an overall precision of plus or minus 5 per cent, planners may analyse data by three age groups with resultant precision ranging from 7 to 15 per cent, but would accept that level of uncertainty compared to the added cost of increasing the sample size to reduce precision for each age group (stratum). However, if a specified level of precision is desired within strata, say 10 per cent for each of the three age groups, then sample size would be calculated for each stratum and added together to get the total sample size (the sample size is increased by the factor of the number of strata).

4.5 Number of domains

Most surveys will have one domain (with 30 clusters). However, if results with reasonable-sized confidence intervals are needed from more than one domain (e.g., both rural and urban, by province, by district), then more domains (usually each with 30 clusters) can be added. As another example, if separate results with certain precision (plus or minus 5 to 10 per cent) are needed for each of 10 provinces, then 10 domains would be needed with 30 clusters each[10]. If there are more than five domains, then fewer than 30 clusters may be sufficient (20 to 29)[11], but the confidence interval may exceed the target precision (confidence interval width) for some of the domains. A summary result across domains can then be calculated using domain weights. The financial cost of the survey depends heavily on the number of domains since the total number of clusters in the survey increases dramatically with increased number of domains (30 clusters per domain).

4.6 Design effect

The design effect is an important concept in calculating the sample size. Because most health and intervention population characteristics have a tendency to be homogeneous within clusters used for a two-stage cluster sampling design, observations are correlated within clusters relative to the other clusters. In other words, characteristics (such as ITN use) are more likely to be similar in nearby households (within 100 metres) compared to households in other clusters (for example, 50 to 100 kilometres away). To account for this within-cluster homogeneity, standard errors (used to compute confidence intervals about all point estimates) are estimated using special statistical methods such as the Taylor Series Linearization or the Robust Sandwich Estimator. Such standard errors and confidence intervals are typically larger than those estimated if simple random sampling had been used for sampling. The design effect is a

10 *There are ways to design surveys other than choosing 30 clusters in each domain, but those designs are complicated and require a survey design specialist. For example, a DHS may use over- and under-sampling, explicit and implicit stratification, and proportional and power allocation. Macro International Inc.,* Sampling manual. DHS-II basic documentation, number 6. *Calverton, Maryland. 1996.*

11 *It is generally recommended not to have fewer than 20 clusters, since the confidence interval could increase substantially.*

measure of this loss (or gain) in precision when a two-stage cluster sampling design is used instead of simple random sampling.

Technically, the design effect for a two-stage cluster survey design is the comparison of the variance or standard error of a cluster survey with that of the same survey assuming simple random sampling was used. There are two types of design effects in common use: DEFF and DEFT. The former compares the variance (square of the standard error) and DEFT the standard error of the two types of surveys (cluster versus simple random sample). DEFF and DEFT are mathematically related since DEFT is the square root of DEFF. Both DEFF and DEFT are used in this manual.

DEFT has two additional features: (1) it is the factor by which the sample size for a cluster sample would have to be increased to have sample precision equal to that for simple random sampling (SRS) (i.e., the penalty for cluster sampling); (2) it is the inflation factor of the confidence interval in a cluster survey compared to a SRS survey. Note: DEFT = √DEFF; DEFF of 2 = DEFT of 1.4; DEFF of 3 = DEFT of 1.7; DEFF of 5 = DEFT of 2.2; DEFF of 10 = DEFT of 3.2.

The design effect is related to heterogeneity/homogeneity between clusters and within clusters. Higher design effects occur when heterogeneity between clusters is higher or the homogeneity within clusters is higher.

It is important to note that the design effect is different for each indicator. In addition, design effects are different for each survey.

Therefore, survey planners need to estimate what the design effect will be for all main indicators in the survey to estimate the sample size.

It is optimal to examine the design effects of similar indicators from previous surveys with similar designs. The DEFTs from previous Demographic and Health surveys (DHS) (each DHS report has an Appendix B that shows DEFTs for most indicators) should be examined for similar indicators. However, the DEFTs in the DHS surveys should be taken as the minimal values since DHS are often done with 250 clusters or more (that yield smaller confidence intervals) whereas RAMP surveys usually have 30 clusters.

Below is an example of the use of data from previous surveys to guess at the design effect. The table below shows the results from two indicators from the four pilot RAMP surveys. When interpreting the design effects of previous surveys, it is useful to know the total number of clusters, number of households per cluster, and the usual size (number of households) of the area in which SRS was used to select households. In the example below, all four surveys had 30 clusters, 10 households per cluster, and SRS was used to select households from segments with 15 to 20 households.

Table 5: Precision and design effect (DEFF) of two main malaria bed net indicators from the four pilot surveys

	Kenya	Namibia I	Nigeria	Namibia II
% households with at least one ITN				
Width of confidence interval (±x%)	6	8	7	7
Design effect (DEFF)	1.4	2.1	3.3	1.7
% ITN use, all persons				
Width of confidence interval (±x%)	6	9	8	8
Design effect (DEFF)	5.1	7.4	9.7	6.7

Notice that the median DEFF for the first indicator is approximately 2.0 and for the second indicator is approximately 7.0. The precision (plus or minus 6 to 9 per cent) was similar for both household and individual indicators despite the design effect being much higher for the individual indicator. Since the individual indicator target population was persons of all ages, the sample size of statistical units was much higher (5.0 statistical units per household) compared to the household indicator.

Usually, DEFF is near or below 2.0 for most health indicators. However, a DEFF of 1.5 or 3.0 is also common. DEFFs greater than 3.0 are possible: water and sanitation-related indicators often have design effects greater than 3.0.

The table below shows examples of design effects for the example indicators.

Table 6: Design effects (DEFFs) of selected main indicators

Main indicators	Design effect (DEFF)
RAMP Malaria survey: bed net indicators	
- ITN use during the previous night, persons of all ages	7.0
- Households with at least one ITN	2.0
- Households with sufficient ITNs	2.0
RAMP Malaria survey: malaria testing and treatment in children under 5 years old	
- Among children under 5 years old with fever in the last two weeks, percentage that received a finger prick or had blood taken	2.0
Micronutrient survey	
- Vitamin A capsule coverage in children under 5 years old	2.0
- Anaemia in children under 5 years old	2.0
- Households using iron fortified product	3.0
- Households using iodized salt	3.0

Notice that all the design effects are 2.0 or 3.0 except for the indicator about ITN use in all ages. From previous surveys, the micronutrient community knows that the design effects for households using an iron fortified product and households using iodized salt have been closer to 3.0 than 2.0, whereas for vitamin A capsule coverage and anaemia, the design effects have been near 2.0. The design effect for ITN use in all ages was listed as 7.0 because ITN use was highly homogeneous within households and PSUs (as expected) in the four pilot

surveys. As mentioned previously, water and sanitation indicators are often more similar within clusters, and therefore may have a higher design effect (e.g., 5.0 to 10.0).

4.7 Calculating sample size needed for percentages/ proportions/prevalence[12] (for a single point estimate)[13]

There are two steps for sample size estimation in RAMP surveys. The first step is the calculation of number of statistical units needed. The second step is the calculation of the number of households that need to be sampled that contain the number of statistical units needed. If statistical units are households (e.g., percentage of households with sufficient ITNs), then steps one and two yield the same number.

For the first step, the sample size (number of statistical units) for a proportion (or percentage or prevalence) is calculated using five factors:

1. Expected prevalence (column 1 of Table 7). This is the prevalence (percentage, proportion) that one would expect to see after the survey is completed. Survey planners would guess at the expected prevalence based on previous experience and knowledge. The maximal variance (and the maximum number of statistical units needed) occurs at an expected prevalence of 50 per cent. The sample size needed to achieve a certain precision declines as the prevalence changes from 50 per cent towards 100 per cent and from 50 per cent towards 0 per cent. Given the other factors are the same, the sample size needed at prevalence of 10 per cent and 90 per cent is identical (and is identical for 20 per cent and 80 per cent, and 30 per cent and 70 per cent). The sample size needed to achieve a certain precision at 10 per cent or 90 per cent prevalence is often half of that needed when the prevalence is 50 per cent. Therefore, planners should carefully consider the expected prevalence.

2. Absolute precision (column 2). This is the half width of the confidence interval that one would like to observe, e.g., plus or minus 3, 5 or 10 per cent, etc.

3. Design effect (column 3). DEFF is used in the web-based OpenEpi tool (www.openepi.com/OE2.3/Menu/OpenEpiMenu.htm) and DEFT is used in the Excel tool provided on the RAMP website (www.ifrc.org/ramp).

4. Sample size of statistical units needed (column 4) to measure specific indicators, e.g., number of persons of all ages, number of children under five years old, number of children under five years old with fever in the previous two weeks, number of households. The two available tools (web-based OpenEpi and Excel-based) use the first three factors to calculate this number.

5. Average number of statistical units per household (column 5). This shows the number of households needed to provide the necessary number of statistical units. For example, if 200 children under five years old are needed, how many households must be sampled to yield this number? In sub-Saharan African

12 *Most survey indicators are percentages, proportions, or prevalence. Most sample size tools use percentages/ proportions/prevalence to calculate sample size. Survey indicators can be means, but most sample size calculation tools will not accept means. In this manual, we exclusively use percentages, proportions, or prevalence to demonstrate sample size estimation.*

13 *Most RAMP surveys are designed to provide an estimate of a single point estimate (for example, 0.60 of households or 60 per cent of households with at least one ITN). Other surveys are designed to compare two or more point estimates across domains or strata, for instance. Sample size calculations for surveys comparing more than one point estimate are discussed in section 4.10.*

countries, children under five years old are usually 17 per cent of the total population and households have an average of 5.0 persons per household. The average number of children under five per household will be 0.85 (17 per cent or 0.17 times 5.0). Therefore, 235 households (200/0.85) need to be sampled to find 200 children under five years old. In another example, assuming that 25 per cent of children under five years old will have had fever in the previous two weeks, approximately 4.25 per cent of the population will be children under five years old that have had fever in the previous two weeks. The number of children with fever per household will be 0.2125 (25 per cent x 17 per cent x 5), therefore, 471 (100/0.2125) households would be needed to find 100 children under 5 years old with fever in the last two weeks.

The web-based OpenEpi tool (www.openepi.com/OE2.3/Menu/OpenEpiMenu.htm) and the Excel tool (contributed by Dr. Thom Eisele, Tulane School of Public Health and Tropical Medicine) provided on the RAMP website (www.ifrc.org/ramp) can be used to calculate survey sample sizes for proportions.

Table 7 provides an overall view of the number of households (sample size of households) that need to be sampled under various conditions. Table 8 shows specific examples for malaria and micronutrient surveys.

Table 7: Example sample size estimations of the number of statistical units needed and households needed to be sampled for indicators with different types of denominators

Indicator denominator	Expected prevalence	Absolute precision (±x%)	Design effect, DEFF	Sample size of statistical units needed*	Average number of statistical units per household (for malaria-endemic sub-Sahara Africa populations)	Number of households that need to be sampled (increased by 5% to account for expected non-response) (Col. 4/Col. 5)
	-1	-2	-3	-4	-5	-6
Household indicators	50 50 20 / 80	10 5 5	2.0	193 768 492	1.0	203 806 517
All persons	50 20 / 80 50 20 / 80	10 10 5 5	7.0	673 431 2688 1721	5.0	141 90 565 361
Children under 5 years old	50 20 / 80 50 20 / 80	10 10 5 5	2.0	193 123 768 492	0.85	238 152 945 608
Child-bearing age women	50 50 20 / 80	10 5 5	2.0	193 768 492	1.25	162 645 414

Indicator denominator	Expected prevalence	Absolute precision (±x%)	Design effect, DEFF	Sample size of statistical units needed*	Average number of statistical units per household (for malaria -endemic sub-Sahara Africa populations)	Number of households that need to be sampled (increased by 5% to account for expected non-response) (Col. 4/Col. 5)
	-1	-2	-3	-4	-5	-6
Pregnant women	50 50 20 / 80	10 5 5	2.0	193 768 492	0.20	1013 4032 2583
Children under 5 years old with fever in last 2 weeks	50 20 / 80 50 20 / 80	10 10 5 5	2.0	193 123 768 492	0.2125	953 608 3795 2431

* Sample sizes of statistical units for column 4 were from the sample size tool at www.openepi.com.

Notice the wide variation in the number of households needed, depending on the expected prevalence, desired precision, and average number of statistical units per household. To measure indicators within plus or minus 10 per cent requires just 100 to 250 households when the denominator is all persons, children under five years old, women of child-bearing age or households. However, measuring those indicators to within plus or minus five per cent increases the number of households needed three-fold to approximately 400 to 750. Changing the expected prevalence from 50 per cent (maximum variance) to 20 per cent or 80 per cent decreases sample size by approximately one-third. The sample sizes of households for pregnant women and children under five years old with fever in the last two weeks are much higher than for indicators with the other denominators (because of the low proportion of statistical units per household). Their sample sizes are near 1,000 for expected prevalence of plus or minus 10 per cent and 2,000 to 4,000 for expected prevalence of plus or minus five per cent.

Table 8 shows the number of statistical units and households to be sampled for malaria and micronutrient surveys.

Table 8: Estimating the number of statistical units and households to be sampled in the entire sampling frame for malaria and micronutrient surveys

a) Malaria surveys						
Main indicators	**Expected prevalence**	**Absolute precision (±x%)**	**Design effect (DEFF)**	**Sample size of statistical units needed**	**Average number of statistical units per household**	**Number of households that need to be sampled***
	-1	-2	-3	-4	-5	-6
Bed net main indicators:						
Individual: ITN use last night, all persons	60	10	7.0	646	5.0	129

a) Malaria surveys						
Main indicators	**Expected prevalence**	**Absolute precision (±x%)**	**Design effect (DEFF)**	**Sample size of statistical units needed**	**Average number of statistical units per household**	**Number of households that need to be sampled***
	-1	-2	-3	-4	-5	-6
Bed net main indicators:						
Household: Households with at least one ITN	60	10	2.0	123	1.0	123
Household with sufficient ITNs	50 (max. variance)	10	2.0	193	1.0	193
Household with sufficient ITNs	50	7	2.0	392	1.0	392
Testing/treatment main indicators:						
Individual: Children under 5 years old with fever that were tested for malaria	20	10	2.0	123	0.2125**	579
Individual: Children under 5 years old with fever that were tested for malaria	50 (max. variance)	10	2.0	193	0.2125**	908

* Unadjusted for household non-response rate.
** Calculated by multiplying the percentage of the total population that are children under five years old (17 per cent) times the percentage of children under five years old that would have had fever in the last two weeks (25 per cent) times 5.0 persons per household.

b) Micronutrient survey						
Main indicators	**Expected prevalence**	**Absolute precision (±x%)**	**Design effect (DEFF)**	**Sample size of statistical units needed**	**Average number of statistical units per household**	**Number of households that need to be sampled***
	-1	-2	-3	-4	-5	-6
Individuals						
Vitamin A capsules, children 0—59 months	80	3	2.0	1365	0.85**	1606
Anaemia, children 0—59 months	50	5	2.0	768	0.85**	903

b) Micronutrient survey						
Main indicators	**Expected prevalence**	**Absolute precision (±x%)**	**Design effect (DEFF)**	**Sample size of statistical units needed**	**Average number of statistical units per household**	**Number of households that need to be sampled***
	-1	-2	-3	-4	-5	-6
Households						
Households using iron fortified product	30	5	3.0	968	1.0	968
Households using iodized salt	60	5	3.0	1106	1.0	1106

* Unadjusted for household non-response rate.
** Calculated by multiplying the percentage of the total population that is under five years old (17 per cent) by the average number of persons in a household (5.0 persons/household).

Notice that survey planners can examine different scenarios on different rows in the table by varying one or more factors. For example, the precision was varied (plus or minus 10 per cent, plus or minus 7 per cent) for the malaria indicator on percentage of households with sufficient ITNs. The expected prevalence was varied (20 per cent, 50 per cent) for the malaria-testing indicator.

The number of households needed for a survey will be the highest number of households (column 6) among the main indicators in the survey. For example, in the micronutrient table above, 1,606 households is the highest number of households of the four indicators. For a malaria bed net survey, 193 households are the highest if a precision of plus or minus 10 per cent is desired or 392 households if a precision of plus or minus 7 per cent is desired. If the survey requires malaria testing and treatment indicators to be measured in children with fever, then 579 households would be needed if a prevalence of 20 per cent is expected and 908 households if a prevalence of 50 per cent is expected.

4.8 Increasing the number of statistical units and households needed to account for non-response

The household non-response rate in most African household surveys of non-sensitive indicators is usually five per cent and rarely more than ten per cent. Therefore, the sample size might be increased by at least five per cent to account for non-response. Survey planners may need to increase the adjustment factor if they anticipate that the target population will be particularly hard to reach (high percentage of households with no respondents present) or the questions are sensitive in nature.

4.9 Number of clusters and number of statistical units and households to be sampled per cluster

Thirty clusters have worked well as a common standard for field surveys for more than 30 years. As a first step, the total number of statistical units needed should be divided by 30 to yield the number of statistical units per cluster. In the malaria example, 203 statistical units (193 increased by 5 per cent to account for non-response) are needed for bed net indicators (e.g., proportion of households with sufficient ITNs); dividing by 30 clusters yields seven statistical units per cluster (since the denominator of the bed net indicator used for calculating sample size was households, the number of statistical units and households needed is the same). For the pilot malaria surveys, the number of statistical units/households per cluster was rounded up to 10. To measure malaria testing adequately in children under five years old with fever at a prevalence of 20 per cent, 129 statistical units and 608 households (123 and 579 increased by 5 per cent) are needed, or approximately 20 households per cluster. In the micronutrient example, 1,433 statistical units and 1,686 households are needed accounting for non-response. Dividing by 30 clusters yields 48 statistical units and 56 households per cluster.

Several experts counsel that the number of statistical units per cluster should not exceed approximately 40 (1,200 statistical units in 30 clusters). If more than 1,200 statistical units are needed, then clusters should be added.

The number of households needed per cluster can be larger (much larger) than the number of statistical units needed per cluster. Therefore, the number of households per cluster can easily exceed the maximum number of statistical units per cluster (40). The number of households will be higher than the number of statistical units when the average number of statistical units per household is less than one (children under five years old, children under five years old with fever in the last two weeks, pregnant women). Indicators with pregnant women as the statistical unit are a good example. If the precision desired is plus or minus 5 per cent, expected prevalence is 20 per cent, and design effect is 2.0, then 492 statistical units and 2,583 households are needed (see Table 7). Four hundred and ninety-two statistical units would result in 17 to 20 statistical units per cluster with 30 clusters, fewer than the maximum of 40 statistical units per cluster. The number of households per cluster would be approximately 90 (2583/30 = 86) after accounting for non-response.

If more than two domains are used, then 25 clusters per domain could be used instead of 30 clusters. It is generally never recommended to have fewer than 20 clusters per domain.

Assuming a constant sample size, more clusters will usually result in a narrower confidence interval. However, the financial cost will also be higher. It should be noted that given a particular sample size, as the number of clusters decreases, precision will go down (i.e., 95 per cent confidence intervals will get larger). It is therefore not recommended to reduce the number of clusters below 30 and increase the number of households selected per cluster, even though the largest driver of cost is often the number of clusters in a given survey.

The following worksheet provides a summary of the survey specifications for a survey to measure the coverage of a mass LLIN distribution in a population. A blank planning worksheet is provided in Annex E if it is required to change the sample size, confidence level or number of clusters in the survey.

Figure 1: Survey specifications: planning worksheet

Summary of LLIN distribution survey	
Health indicator:	Percentage of households with sufficient ITNs
Denominator of indicator:	Households
Confidence level:	95%
Precision/Confidence interval:	+/- 10% or 0.10
Estimate design effect:	2.0
Expected value:	50% or 0.50
Sample size of statistical units:	193 households
Sample size of households:	193 households
Final sample size of households after increasing to account for expected non-response of 5%	**203 households**
Number of clusters:	30
Statistical units/households to interview per cluster:	10 (7 rounded up to 10 to be safe)

4.10 Sample size considerations when comparing two surveys, domains or strata

The sample size calculations above derive a sample size for a single point estimate in time with a specified degree of precision (or sample sizes for each survey domain with desired precision); it does not provide an estimation of the sample size needed for making statistical comparisons between survey domains or multiple survey rounds.

The sample size calculations for determining if point estimates for two surveys or two survey domains are statistically different require other considerations. For example, survey planners may want to make a statistical comparison of two survey domains consisting of different geographical areas (for example, comparing a survey with 30 clusters in Province A with another 30-cluster survey carried out in Province B in the same month). In another example, survey planners may desire to compare two surveys from the same place at different times (for example, comparing a survey in Province C in 2011 (baseline) with a follow-up survey carried out in 2013).

For those wanting to conduct sample size calculations for statistical comparisons, the following information sources are useful:

1. **Excel sample size tool (see www.ifrc.org/ramp). This tool estimates sample sizes for comparisons using factors listed below (design effect, expected proportions, type 1 and 2 errors).**
2. **Section 3.5.2 Sample size calculation for comparing two surveys, pages 31-33 in Gorstein J, Sullivan KM, Parvanta I, Begin F (May 2007) *Indicators and Methods for Cross-Sectional Surveys of Vitamin and Mineral Status of Populations*. The Micronutrient Initiative (Ottawa) and the Centers for Disease Control and Prevention (Atlanta). See: www.micronutrient.org/CMFiles/PubLib/Indicators-for-Cross-Sectional-Surveys1IYA-3242008-2823.pdf, www.micronutrient.org**

In general, the sample size calculations for making statistical comparisons between two proportions will require the following information:

- Design effect (DEFT or DEFF)
- Expected population prevalence/proportion in domain 1 (or baseline)
- Expected population prevalence/proportion in domain 2 (or follow-up)
- Desired probability of committing a type-1 error, typically set at 5 per cent (Z = 1.96 with 2-tailed test or 1.645 with a 1-sided test). This stipulates the probability of *rejecting* the null hypothesis (no difference between point estimates) when in reality there truly is a difference.
- The desired probability of committing a type-2 error, typically set at 80 per cent (Z = 0.84). This stipulates the probability of *accepting* the null hypothesis (no difference between point estimates) when in reality there truly is a difference.

4.11 Questions and answers on sampling for RAMP surveys

Questions	Answers
Q. What is the width of the confidence interval needed to make management decisions?	**A.** Traditionally, plus or minus 10 per cent has been used for management decisions. Generally, a confidence interval of 15—20 per cent is felt to be too wide. The information in this manual aims at plus or minus 5—10 per cent.
Q. To get plus or minus 5—10 per cent, how many statistical units do you need to include in the survey?	**A.** The width of the confidence interval mostly depends on three things: (1) sample size, (2) the number of clusters and (3) the variability between clusters (PSUs). A large variation between clusters will result in wider confidence intervals. An example of variation would be if zero per cent ITN use is found in several clusters and more than 80 per cent ITN use in other clusters, compared to another survey in which ITN use was 40—50 per cent in all clusters. Thirty clusters are usually chosen so that the intra- and inter-cluster variability is reasonable. Assuming that the variation between clusters is minimal, then a design effect of 2.0 can be used to calculate the sample size. If a large amount of variability between the clusters is anticipated, a design effect of 3.0 or higher can be used.

Questions	Answers
	Calculating the sample size for malaria surveys is more complicated than other surveys because there are two types of statistical units: *persons* and *households*. For example, the statistical unit in the percentage of households that have at least one ITN is the *household* compared to the percentage of persons sleeping under an ITN last night where the statistical unit is *persons*. In most African countries, the number of households is five times fewer than the number of persons (average 5.0 persons per household). To calculate the sample size, you need to input three numbers into a sampling calculator: *Width of the confidence interval (precision) that you want* (for example, plus or minus 8 per cent or 0.08) *Estimate of the design effect*. Use 2.0 for the design effect. If you think that there will be much difference in the point estimate from cluster to cluster then use 3.0. For example, if some villages may have been totally missed during LLIN distribution, use 3.0. *Estimate of the prevalence*. For percentage of households with at least one ITN, 50 per cent should be used as the prevalence if the quality of mass LLIN distribution may not be optimal. If it is considered that most households were reached with LLINs, 70, 75 or 80 per cent can be used. *Example:* Using the web-based survey calculator found at Emory University School of Public Health (www.openepi.com/OE2.3/Menu/OpenEpiMenu.htm[14]), at plus or minus 10 per cent, design effect of 2.0 and prevalence of 70 per cent, then 162 households are needed. With a design effect of 3.0, then 243 households are needed (95 per cent confidence is standard).
Q. Is a different sample size needed for large populations such as one million, 10 million or 100 million?	**A.** Surprisingly, 300 households will provide an estimate of plus or minus 10 per cent regardless of the size of the population. One does not need to survey more than 300 households, even for a population with more than 10 million households.
Q. Three hundred households will provide a single estimate (for example, 77 per cent of the households had at least one LLIN) for the whole sampling frame (such as province or state). If a point estimate at plus or minus 10 per cent for each of ten districts is wanted, how many households and how many clusters are needed?	**A.** If ten point estimates (one per district) are desired, then the sample size and number of clusters would be 10 times 300 households and 10 times 30 clusters (300 clusters). It might be possible to cut corners by using 20 clusters per district, but the confidence interval might be higher than 10 per cent in some districts. Having fewer than 20 clusters per survey domain can be risky, since the confidence interval could reach plus or minus 15 per cent or higher.
Q. If data by district are desired, should 200—300 clusters be randomly chosen from the whole province/state or 20—30 clusters per district?	**A.** It does not make much difference. Modern software will calculate the estimates and confidence intervals appropriately by district with either method. There are advantages to each method. If there are large differences in the population between districts, then there will be an unequal probability of selection between districts if you choose 20—30 clusters per district, but this does not have any impact on the validity of the point estimates and confidence intervals. Choosing 200—300 clusters for the whole state or province maintains an equal probability of selection at the state or province level.

14 *Sullivan KM, Dean A, Soe MM ,* OpenEpi: A Web-based Epidemiologic and Statistical Calculator for Public Health. *Public Health Rep. 2009 May-June; 124(3): 471–474. See: www.ncbi.nlm.nih.gov/pubmed/19445426.*

Questions	Answers
Q. Can the data be analysed by urban/rural status, geographical area (e.g., north versus south, by province), and wealth quintile?	**A.** Yes. Modern analysis software (STATA, SAS, SPSS, EpiInfo) can provide appropriate point estimates and confidence intervals regardless of how the sampling frame was constructed[15]. However, the confidence intervals for each stratification or disaggregation level are likely to be wider than the target width of the confidence interval for the single point estimate. For example, if the target confidence interval is plus or minus 10 per cent for the whole survey sampling frame or domain, the confidence interval for each stratification level (for example, if you have two, three, or four stratification levels, etc.) will likely be greater than 10 per cent. Analyses should be avoided that have too many stratification levels yielding confidence intervals that are so wide that the results are difficult to interpret.
Q. Do the strata for analysis have to be decided before choosing the method of sampling?	**A.** No. Modern software will appropriately produce the correct estimates and confidence intervals assuming that the sampling weights are properly calculated.
Q. Can the design effect be calculated before the survey is done? Additionally, is there one design effect for each survey?	**A.** No, the design effect cannot be calculated before the survey data are collected—one can only make an educated guess based on previous experience. See Appendix B of the most recent national DHS report where the DEFTs listed for similar indicators can provide some guidance. Remember, there is a different design effect for each question/indicator/analysis element. For example, the design effect for the percentage of persons sleeping under an ITN last night might be different from the percentage of households with at least one ITN. In practical terms, it is often assumed that there is one design effect per survey, but this is not true. There is a design effect for each indicator.

15 *Brogan D.* Household Sample Surveys in Developing and Transition Countries. *Chapter XXI Sampling error estimation for survey data.* See: unstats.un.org/unsd/HHsurveys/pdf/Chapter_21.pdf.

© Benoit Carpentier/IFRC

05. Selecting clusters, segments, and simple random sampling of households

This section gives information on the selection of clusters in advance of survey implementation in the field and the eventual selection of households. After clusters are selected, planners need to provide rules on how to proceed: segmentation or simple random sampling? This section will be most useful to the survey coordinating group, the survey designers and planners, and the survey coordinator.

Having calculated the necessary sample size and the number of clusters needed, the next stage is to select the clusters for the survey, a task which starts with identifying the population data.

5.1 Using population data for selection of PSUs/clusters

In the first place, a list is made with the population of the smallest area unit with reasonable quality population data. For national, regional or provincial surveys, this is likely to be a list of census enumeration areas. For district or sub-district surveys, it may be a list of villages or sub-district administrative units. It is best to list both the population and the estimated number of households.

For national or provincial surveys, data from the most recent census, available from the government department responsible for national statistics, may be used. Immunization data or other sources of data may be more recent, however, and those data sources may be the best to estimate the relative size of small population or area units for district or sub-district surveys.

The smallest unit for which there are good data should be used to construct the sampling frame for selecting clusters at the first stage.

It may be necessary to make an official request to the national statistics bureau to obtain the population data from the most recent national census for selecting the clusters. This can take up to a few weeks and there may be a cost connected to the release of the data. Sometimes, the statistics bureau can provide the service of selecting the clusters, according to the information provided by the survey planner. The statistics bureau often has a headquarters in the capital of a country, and regional or district offices. In addition, maps (e.g., of enumeration areas, of a district, and so on) can often be obtained from a government agency such as the bureau of statistics or cartography units. The maps can help to plan the survey schedule and aid survey teams to locate the selected clusters.

If a district-wide or other sub-national survey is to be conducted, the same procedures described in this manual for a nationwide survey should be followed, but a complete list of enumeration areas or primary sampling units from the area of interest would need to be developed. The same number of clusters and households as described for a nationwide survey would be selected. A list of villages can be used to select clusters at the first stage if the list is complete and population size estimates exist (census data, MoH data, EPI data, or other data) for all the villages.

The number of clusters to be selected can be 30, 50, 100, 250, or any higher number. In the fictional example used in this manual, as would be usual in a RAMP survey, 30 clusters are selected. The method described to select clusters is called probability proportional to estimated size (PPES). That is, the probability of a geographical unit (district, village, enumeration area) being selected as a cluster to be included in the survey depends on its estimated size. The term "estimated" is used before "size" because the exact population size is rarely known. The clusters can be selected using paper or a computer spreadsheet program or can be selected using a computer statistical program like STATA or SAS. The procedures for using STATA and SAS can be found on the RAMP website at www.ifrc.org/ramp. The paper and spreadsheet method is described in Volume 2 of the RAMP survey toolkit, *Implementing a RAMP survey: practical field guide,* Sections 4.1 and 4.2, and is also described in the EPI survey manual[16].

5.2 Next step: simple random sampling or segmentation

Survey planners need to decide the sampling step that will next be used after the clusters have been chosen. Do they use simple random sampling (SRS) from an enumeration of all households in a selected cluster, or do they segment the cluster into smaller parts before conducting simple random sampling?

Simple random sampling at second stage generally yields higher precision than using the segmentation method (given the same sample size), but doing so can be difficult and the cost may be high if the number of households in the cluster is high, say more than 30 to 100. Simple random sampling requires enumerating all households in the cluster. The gold-standard listing method is to visit all households, inquire about the name of head of household, record the name on a numbered list, and then give a unique identification number to each house so that selected households can be found at a later time.

16 *World Health Organization.* Immunization coverage cluster survey-Reference manual. *Department of Immunization, Vaccines and Biologicals, World Health Organization, Geneva, 2005, WHO/IVB/04.23, Annex D. See: www.who.int/vaccines-documents/DocsPDf05/WWW767.PDF*

Because of the main RAMP objectives of simplicity and low cost, segmentation (and sub-segmentation if required) is recommended in RAMP surveys in most situations. However, survey planners can use any cut-off for SRS, for example, 100, 50 or 30 households. If the cluster sizes are nearly all fewer than 30 households, then SRS can be used exclusively. Since a typical survey has clusters with more than 30 households, segmentation would be used most often.

Planners need to establish rules for sampling within the clusters. Before establishing rules, planners need to specify:
- number of households to be sampled in each cluster
- estimated final segment size (usually 1.5—2 times the number of households to be sampled)
- maximum number of segments if segmentation is used (for example, 10)

Next, planners need to specify one of three types of rules:
- **Rule type 1:** Always start with segmentation of cluster
- **Rule type 2:** Never segment clusters. Always start with simple random sampling
- **Rule type 3:** Start with segmentation if the number of households is more than "x". Start with simple random sampling if the number of households is fewer than or equal to "x".

For a RAMP survey, it is most likely that segmenting clusters will be the method chosen. The following is an example format of the written instructions that would be provided to supervisors by the survey planners:

Selecting the households to be sampled: instructions to supervisors
10 households are to be sampled in each cluster
Final segment size should be between 15 and 20 households

1. Always start with segmentation if the number of households in the cluster is more than 30 (nearly all clusters).

2. If there are fewer than 30 households in the cluster, then start with simple random sampling immediately.

3. In the first phase of segmentation, the maximum number of segments should be 10. Try to create segments that are approximately 15—20 households (they can be larger, but not smaller than 10 households).

4. After the first phase of segmentation, if the segment selected has fewer than 20 households, then proceed to simple random sampling.

5. After the first phase of segmentation, if the segment selected has more than 20 households, then sub-segment into sub-segments that have between 10 and 20 households, randomly select one sub-segment, then proceed to simple random sampling.

5.3 Target size of final segment and sub-segmenting

The target size of the final segment or sub-segment depends on the total fixed number of households to be selected in each cluster. The target should be at least 1.5 to 2.0 times the number of households required per cluster. For example, if 10 households are to be selected for each cluster, then the final segment or sub-segment size should be 1.5 to 2.0 times 10, or 15 to 20 households. If 20 households are required per cluster, then the final segment or sub-segment should be 1.5 to 2.0 times 20, or 30 to 40 households. Once the target size of the final segment has been decided, then instructions to leaders of the interview teams can be formulated about segmentation.

Clusters can be segmented to the final target size in one step, or if the clusters are too large, in two steps (segmenting, then sub-segmenting). For example, if the cluster has 200 households and the final target segment size is 15 to 20 households, then a two-step process would be undertaken. If the cluster has only 100 households, however, then 5 to 7 segments of approximately 15 to 20 households would be made in a one-step process without the need for sub-segmentation. It is not advised to exceed 10 segments in any one cluster. Unequal size segmenting and sub-segmenting are carried out using PPES in a process similar to selecting clusters.

Detailed instructions and steps for carrying out segmentation and sub-segmentation can be found in Volume 2 of the RAMP toolkit, *Implementing a RAMP survey: practical field guide,* Section 5.

5.4 Simple random sampling of households

Once the final segment is selected, then households are selected for interview using simple random sampling.

Detailed instructions and steps for carrying out the selection of households for interview by SRS can be found in Volume 2 of the RAMP toolkit, *Implementing a RAMP survey: practical field guide,* Section 5.5.

© J. Cervinskas/IFRC

06. Options for data collection

In this section, questionnaire development, data collection, data management and data quality are discussed in terms of a comparison between expensive-to-implement and low-cost-to-implement mobile phone software and the paper and pencil method. The advantages of using Magpi (previously EpiSurveyor) for the RAMP survey are described. This section will be most useful for the survey coordinating group, survey designers and planners and the survey coordinator.

Since a RAMP survey has low cost and simplicity as two of its main objectives, the first pilots used mobile phone software that was low-cost-to-implement. The Magpi (previously EpiSurveyor) software system was the software chosen to implement the RAMP pilot surveys because it was a fully functioning system. RAMP surveys can be, however, implemented with other types of software or paper/pencil systems. More expensive-to-implement software and paper/pencil systems may work better for some types of surveys, for example for very long and complex surveys (like DHS).

A comparison of three methods of data collection demonstrates a number of advantages and disadvantages for each. These methods are as follows:
- by mobile phone using low cost software (using Magpi as the example)
- by mobile phone using more expensive-to-implement software
- using paper and pencil

The less expensive, or often free, software may not have robust relational database management and relational error-checking capabilities. More expensive commercial software usually has both, but is costly and technical support may be needed to establish relational questionnaires and relational error-checking. Surveys with a single database (individuals only like EPI surveys rather than both individuals and households like malaria surveys) do not need relational database capabilities. Paper-based systems do not have the advantages of immediate and simple data acquisition and immediate data cleaning of mobile phone-based methods.

The Magpi system has three main components:
- web-based questionnaire design
- data entry on common mobile phones
- data transmission over mobile phone networks to a web-based database

The table below gives comparisons between Magpi, more expensive-to-implement, commercial software, and paper and pencil systems.

Table 9: Characteristics of Magpi, more expensive commercial software, and paper/pencil systems

In the table below, "-" indicates no, "+" indicates yes, and "++" or "+++" indicates yes to greater degrees.

Characteristics	Magpi	Other commercial	Paper/Pencil	Comment
Simplicity (external technical support not needed)	+++	-	+	More expensive-to-implement commercial software is complex and may require external technical support.
Cost	Free or low cost	High cost	+	Magpi's free account is limited to questionnaires of 100 questions or fewer, and the uploading of 6,000 completed questionnaires per year. Unlimited paid accounts are also available.
Simple web-based questionnaire design	+++	+	-	Simple web-based design allows better collaborative questionnaire development.
Data entry	+++	+++	-	In Magpi, data are entered on Java, Android, and iPhone phones. Inexpensive Java phones were used for pilot RAMPs.
Data transmission over mobile phone network to web database	+++	+++	-	Instant data transmission when in range of 2G+ mobile network.
Real-time error analysis of web data	+++	+++	-	Data quality improved by real-time error analysis.
Relational database structure	-	+++	+++	For the pilot surveys, EpiSurveyor (Magpi) did not have relational structure, but this is currently being added. More expensive-to-implement commercial software is likely to have relational structure.
Data error checks	+	+++	+	For the pilot surveys, EpiSurveyor (Magpi) had basic error checks, currently being improved. More expensive-to-implement commercial software may have better relational error checks.

More expensive-to-implement commercial software's biggest disadvantages are cost and complexity of operation, usually requiring external technical assistance. Magpi's main advantages for a RAMP survey are simplicity and low cost. Its main disadvantage for the pilot surveys was the lack of a relational structure, but this is currently being addressed. This is not an issue in surveys that have a single questionnaire, but malaria surveys have three questionnaires: household, individual, and net. The weakness was overcome by using two paper job aids to ensure, as far as possible, that cluster and household ID numbers were correct. The relational error checking weakness was partially overcome by real-time error analysis of incoming data.

Mobile phone data entry has three advantages over paper/pencil:
- simplicity of data management. Paper/pencil data entry method often uses numerous sheets of paper, which then have to be saved and protected after completion of interviews
- speed of data availability for data analysis and report production
- real-time error analysis and field correction to improve data quality

It is possible to use double data entry in paper/pencil systems for RAMP surveys. Double data entry greatly reduces data entry errors compared to single data entry. If well organized, data can be double-entered and checked in one week. However, double data entry and comparison may not be easily organized by many NGOs.

The experience of the four pilot surveys showed the many advantages of Magpi (then called EpiSurveyor):
- The three malaria questionnaires (household, individual, net) were developed quickly and collaboratively on the web.
- Inexpensive (approximately US$ 80) Java-capable phones were purchased within countries.
- The phones were quickly activated.
- Data were easily entered by Red Cross Red Crescent volunteers, even those with no previous survey experience.
- Data were instantaneously sent to the web database once in range of the 2G mobile network at the end of each survey day.
- Data were immediately available for error checking and analysis throughout the survey.
- Error rates that were detected were very low, on the order of one error per 1,000 records.
- Results were available for feedback to survey workers within 12 hours of the last interview.

When the RAMP survey method was first designed, EpiSurveyor had the best software fit because of the questionnaire web design feature, robustness and reliability. However, the mobile environment is changing rapidly and it is anticipated that many different types of mobile tools will be available in the near future to implement RAMP surveys. For example, less expensive-to-implement commercial software may add relational database and relational error-checking (as Magpi is doing) and other such features that would be useful in RAMP surveys.

Detailed steps for the collection of quality data in the field and its management can be found in Volume 2 of the RAMP toolkit, *Implementing a RAMP survey: practical field guide,* Sections 6.4 to 6.6.

© Melanie Caruso/IFRC

07. Data analysis

In this section, the data analysis plan and how the plan is used in the design of the survey are outlined, and the table shells that will form the results bulletin are described. Finally, the procedures for managing and cleaning the data and for analysing data are explained. This section is most useful for the survey coordinator, survey designer and data manager/data analyst.

7.1 Data analysis plan

The data analysis plan should be prepared as part of the survey planning process. In general, the plan provides critical information for sample size calculations, questionnaire development and questionnaire finalization. The main part of the analysis plan includes empty data analysis table shells for presenting all key indicators outlined in the survey objectives. The plan allows the survey planners to lay out the key socio-demographic characteristics by which the indicators will be stratified/disaggregated. This then helps planners to collect the relevant information in the questionnaire for stratification/disaggregation (e.g., wealth status, mother's educational status, etc.).

The data analysis plan allows those responsible for survey design to:

- show indicators that can be analysed from the questions in the questionnaires and help ensure that all main indicators can be constructed from the draft questionnaire
- modify the questionnaires if desired indicators cannot be produced from the first draft questionnaires
- create separate lists of primary (those with the highest priority) and secondary indicators
- provide feedback into the sample size calculation and survey design process since the survey must be able to produce the appropriate precision of indicators in the data analysis plan table shells
- decide which supplementary analyses/questions are "nice to have" but may not be necessary and could be deleted to shorten the questionnaire
- think about the importance of stratified/disaggregated analysis (by rural/urban status, by age, by wealth quintile), for example, whether certain stratified/disaggregated analyses can or cannot be safely done within the planned sample size or whether the sample size needs to be increased to provide stratified/disaggregated estimates with reasonable precision
- consider the precision for each indicator and stratified/disaggregated analysis in relation to the planned sample size

The analysis plan and table shells provide the core tables and basis for key graphs that provide the structure for the survey report and data dissemination products.

Table 10 contains several components of a data analysis plan for RAMP malaria surveys. The main indicators that must be incorporated into the survey are shown, followed by a list of table shells. See Annex F for the actual table shells.

Table 10[17]: Main indicators for RAMP malaria surveys[18]

Prevention	
Vector control via ITN and Indoor Residual Spraying (IRS)	Proportion of households with at least one ITN *(net database questions 10, 11, and 12 indicating whether a net was an ITN, related to household database)*
	Proportion of households with at least one ITN for every two people *(net database questions 10, 11, and 12 indicating whether a net was an ITN, persons database shows number of persons in the household, both related to household database)*
	Proportion of population with access to an ITN within their household *(net database questions 10, 11, and 12 indicating whether a net was an ITN, persons database shows number of persons in the household, both related to household database)*
	Proportion of population who slept under an ITN the previous night *(net database questions 10, 11, and 12 indicating whether a net was an ITN and questions 14—18 indicating who slept under the net last night, related to the persons database)*
	Proportion of children under five years old who slept under an ITN the previous night *(net database questions 10, 11, and 12 indicating whether a net was an ITN and questions 14—18 indicating who slept under the net last night, related to the persons database, question 8 about age)*
	Proportion of households with at least one ITN and/or sprayed by IRS in the last 12 months *(household database question 20 about IRS and net database questions 10, 11, and 12 indicating whether a net was an ITN, related to household database)*
Case management	
Diagnosis	Proportion of children under five years old with fever in the last two weeks who had a finger or heel stick *(persons database question 10 about blood taken, question 9 about fever in the last two weeks, and question 8 about age)*
Treatment	Proportion of children under five years old with fever in the last two weeks for whom advice or treatment was sought *(persons database question 11 about advice or treatment, question 9 about fever in the last two weeks, and question 8 about age)*
	Proportion receiving first line treatment, among children under five years old with fever in the last two weeks who received any anti-malarial drugs *(persons database questions 17, 19, and 21 about first-line or any anti-malarial drugs, question 9 about fever in the last two weeks, and question 8 about age)*

17 *The question numbers noted in this table refer to the example questionnaires published in Volume 2 of the RAMP toolkit,* Implementing a RAMP survey: practical field guide.

18 *Indicators for pregnant women and those requiring blood taken from children (e.g., anaemia or malaria parasitaemia) have not been included.*

The list of table shells for the RAMP malaria survey is as follows:

1. key descriptive information about survey (population in the sampling domain, number of clusters, numbers of households and persons sampled, etc.)
2. number of ITNs needed (ITN gap) to reach universal coverage (estimated number of ITNs in the sampling frame and the gap or number of ITNs needed)
3. main and secondary indicators (main MERG indicators[19], plus supplementary indicators of value to IFRC)
4. selected ITN indicators by wealth quintile
5. age of ITNs in years
6. percentage of households that received a home visit from a community volunteer in the last six months (of interest to IFRC)
7. source of ITNs (mass campaign, market, health facility, etc.)
8. greatest source of information about ITN use (community volunteer, health facility, radio, etc.)
9. malaria testing and treatment indicators
10. precision and design effect of key indicators (useful to guide sample size calculations for future surveys)
11. ITN use by age group
12. type of net (brand of LLINs, ITNs, non-ITN nets)
13. percentage of persons sleeping under ITNs last night (for example, shows percentage of nets that had three or four persons sleeping under them the previous night)

See Annex F for a fuller description of the table shells.

7.2 Data cleaning and management

Once data are entered into the mobile phone, they may be sent directly from the phone to a web database. To do this, however, a 2G mobile network connection must be available. In addition, the gateway from the national to the international network must be working. If interviewers are working in remote or country areas without a 2G network connection, they can wait until the end of the day to send the data. A 2G network service is available on most main roads and in district capital cities in most countries. If the 2G network or the international gateway connections are not working, then data can be downloaded to a computer using several methods (Bluetooth, cable, micro-USB card).

The data analyst tasks require good access to the internet to see and retrieve the web data. Therefore, if good internet access is not available at the local site of the survey, the data analyst tasks may need to be performed by other than the data manager, who will be present on the survey site. Since data are available each evening, the person performing the data analyst tasks can download and closely examine the data for potential errors using error-detection programs, and start the data cleaning process. An example of an error-detection program in STATA can be found on the RAMP website (www.ifrc.org/ramp). The error-detection program should be examined, modified, and finalized prior to the start of the survey by the data manager/data analyst. An advantage of mobile phone data collection and transmission is error-detection from the first day and throughout the survey. Immediate error-detection allows immediate error correction by calling or sending text messages to survey and interview team leaders.

19 *See www.rbm.who.int/mechanisms/merg.html, reference documents tab.*

Data cleaning, or validation, is used to ensure the validity and accuracy of the survey data once they have been uploaded and combined into a dataset. Even though the data are in an electronic dataset format, cleaning often includes manual inspection of the data to ensure no large irregularities exist. Data quality checks programmed in a statistical package, such as STATA, are needed to identify inaccurate or invalid data ranges, missing values, consistency checks and aggregate descriptive statistics to detect strange patterns in the data[20].

Cody (1999)[21] defines data cleaning to include the following:

- ensuring that the raw data were accurately entered into a computer-readable file
- checking for uniqueness of certain values, such as unique identification numbers for each observation
- checking for and eliminating duplicate data entries
- checking that numeric values are within predetermined ranges
- checking that character variables contain only valid values
- checking if there are missing values for variables where complete data are necessary
- checking for invalid date values and invalid date sequences

Sample STATA code showing the various data cleaning elements above can be found on the RAMP website (www.ifrc.org/ramp).

7.3 Analysing the data

After two to three days of data collection (in the middle of the survey), the data analysis programs can be run as a "test" and, if issues arise, can be modified. The steps in data analysis are outlined below. The data analysis process will be demonstrated using STATA, but EpiInfo, R, SPSS, and SAS can be used to analyse survey data. In general, spreadsheet programs like Microsoft Excel are not used to analyse survey data as they do not have the capacity to properly estimate standard errors for complex survey designs. The example below will be assumed to be from a survey with a single domain. The analysis of a single malaria indicator (proportion of households with at least one ITN) will be used to illustrate the analysis process.

Step 1: import the clean data from Excel and prepare the data for analysis using STATA commands (re-code date, character to numeric variable conversion, etc.).

Step 2: combine the household and net databases.

Step 3: calculate the unadjusted sample household selection probability/weights.

Step 4: calculate the selection probability/weights adjusted for non-response.

Step 5: issue commands that specify the survey design (name of the cluster and weight variables).

Step 6: issue the command specifying the indicator type (proportion, mean, etc.) that provides the point estimate and confidence interval.

Step 7: issue command that results in the design effects.

Step 8: issue command that yields results for a stratified analysis by rural/urban status.

20 *Society for Clinical Data Management. 2003.* Good Clinical Data Management Practices: Version 3.

21 *Cody, R. 1999.* Cody's Data Cleaning Techniques using SAS Software. *Cary, NC: SAS Institute Inc.*

The STATA code for the eight steps is shown below. Text is not processed as STATA commands, so an explanation of the code is given. The indicator is the proportion of households with at least one ITN. Data are needed from two databases (household and net) to analyse this indicator.

Step 1. Data imported from Excel, cleaned, and prepared for analysis

```
cd "/STATA_example_code" // This command sets the working directory or folder.

This command imports household data from Excel file worksheet called "Household".
"/RAMP survey.xlsx", sheet ("Household") firstrow clear
save household, replace // Save household data into database called "household".

This command imports net data from Excel file worksheet called "Net"
"/RAMP survey.xlsx, sheet ("Net") first row clear
A variable called "itn" is created (steps not shown)
save net_itn, replace // Save net data into database called "net_itn"

Many additional commands are issued to prepare data for analysis. For example,
- some variables need to be changed from character to numeric or vice versa
- some variables need re-coding to account for missing information
- some variables need re-coding from character Yes/No to number 0/1 variables

The complete set of data preparation commands can be found at www.ifrc.org/ramp
```

Step 2. Combine household and net database

The data for the number of ITNs is in the net database (one record for each net). The net data by net needs to be converted into net data by household with the number of ITNs by household. The household database contains a record of all households, including those that do not have any ITNs. Therefore, the household and net data need to be merged.

```
use net_itn, clear // Load net database.
gen num_itns = 0 // Create variable for showing the number of ITNs in each household.
replace num_itns = 1 if itn == "Yes"
collapse (sum) num_itns , by(hhid) // Sums ITNS by household ID number.
save nethhnumitns, replace // Save to new database with the new variable – number of ITNs per household.

use household, clear // Load household database.
merge 1:1 hhid using nethhnumitns, keepusing(num_itns) // Merge the household database with the new net database
gen hhitn = 0 // Create new variable indicating whether the household has at least one ITN
replace hhitn = 1 if num_itns > 0 & num_itns < 100 // hhitn = 1 if there are any ITNs
save household, replace // Save combined datasets with new variables
```

Step 3. Calculation of unadjusted sample household selection probability/weight

Unadjusted selection probability of statistical units/households and weights. Unadjusted selection probability of statistical units and households is the product of the number of clusters (30) multiplied by the number of households targeted per cluster (10) and divided by the total number of households in the sampling frame. Remember that RAMP has an equal probability design so the unadjusted weight is the same for all households. Also remember that the household, person, and net weights are all the same since all persons and nets were sampled in each selected household (person and net probability of selection was 1 if the household was selected for sampling).

```
Use sam_frame_data, clear // Sampling frame database (one record per cluster) contains number of households in each cluster.
```

First, we need to create three new variables:

```
gen targethhsampled = 10 // Creates new variable with fixed number of households sampled in each cluster, (10 in our example).
gen numclusters= _N // Creates new variable with the number of clusters in the sampling frame (30 in our example).
egen totalhhs=total(households) // Creates new variable with the total number of households in the sampling frame.
```

Then we create the unadjusted statistical unit/household selection probability by multiplying the number of clusters (30) by the number of households targeted per cluster (10) divided by the total number of households in the sampling frame.

```
gen unadj_selprob=(numclusters * targethhsampled)/totalhhs
```

Note: The unadjusted weight would be the reciprocal of the unadjusted selection probability.

```
save sampframedata, replace // File name is similar to the imported data, but not exactly the same.
```

Step 4. Adjusting selection probability/weights for non-response

Discussed in more detail in Section 7.4, “Calculation of sample weights and selection probabilities”.
In Step 4, there are several sub-steps. First, create a household response variable in the household database, then create a new small database that shows the number of households that were interviewed by cluster (10, 9, 8, etc). Second, relate the small newly-created database with the sampling frame database. Once the household response variable is merged into the sampling frame database, the adjusted selection probability and weight can be calculated for each cluster.

Note: the variable for household non-response is hhinterviewed_yn

```
use household, clear
gen hhinterviewed = 1
replace hhinterviewed = 0 if hhinterviewed_yn == “No”
collapse (sum) hhinterviewed, by(hh_clusternumber)
gen targethhsampled = 10 // The number of households sampled per cluster (10 in our example).
```

In the new “collapsed” database, create new variable called “response” = number of households actually interviewed divided by households targeted for interviewing (10).

```
gen response = hhinterviewed /targethhsampled
destring hh_clusternumber, replace
save nonresponse_psu, replace
```

Merge new “response” variable with sampling frame database that contains the unadjusted selection probability variable.

```
use sampframedata, clear
merge 1:1 hh_clusternumber using nonresponse_psu, keepusing (response)
```

drop _merge
Adjusted selection probability = unadjusted selection probability times new response variable.

```
gen adj_selprob = unadj_selprob * response
```

Weight is the reciprocal of the total adjusted selection probability.

```
gen adj_weight = 1 / adj_selprob
tostring hh_clusternumber, replace
save sampframedata, replace
```

Merge adjusted weight into the household database

```
use household, clear
merge m:1 hh_clusternumber using sampframedata, keepusing(adj_weight)
drop _merge
drop if hhinterviewed_yn == “No”
save household, replace
```

Step 5. Issue survey command specifying the survey design and key variables

```
svyset hh_clusternumber [pweight=adj_weight]
```

hh_clusternumber is the cluster (PSU) variable and adj_weight is the adjusted sample weight.

```
pweight: adj_weight // pweight is the probability weight.
VCE: linearized // VCE is the method used to calculate confidence interval. “Linearized”= Taylor linearized variance estimation method.
Single unit: missing // No strata with only one sampling unit.
Strata 1: <one> // Strata = <one>, only one stratum.
SU 1: hh_clusternumber // SU, sampling unit, PSU variable = hh_clusternumber.
FPC 1: <zero> // FPC, finite population correction, not used in this example.
```

Step 6. Issue the survey command specifying the type of indicator (proportion, mean, etc.)

```
svy: proportion hhitn // hhitn is the variable showing “1” if at least one ITN in the household and 0 if no ITNs. In addition to the point estimate, results show 95% confidence interval and the standard error.
```

	Linearized			
	Proportion	Std.Err.	[95% Conf.	Interval]
hhitn				
0	.3541667	.0401521	.2720465	.4362869
1	.6458333	.0401521	.5637131	.7279535

Note: Proportion for hhitn = 1 (.6458333) is the desired indicator; 65% of households have at least one ITN. Confidence interval is 56% to 73%.

Step 7. Examine the design effects (DEFF, DEFT)

```
estat effects, deff deft
```

	Linearized			
	Proportion	Std.Err.	DEFF	DEFT
hhitn				
0	.3541667	.0401521	2.09336	1.44685
1	.6458333	.0401521	2.09336	1.44685

NOTE: The design effect DEFF = 2.1 and DEFT = 1.4 for this indicator.

Step 8. Perform stratified (disaggregated) analysis by rural/urban status

The subpopulation command "subpop()" is used for stratified analysis.
The subpopulation variable must = 1 for the subpopulation chosen and = 0 for others.
The estimate will be provided for those with subpop variable = 1
In our example, rural/urban status is the subpop variable, with rural = 1 and urban = 0.

```
svy, subpop(rur_urb): proportion hhitn
```

	Linearized			
	Proportion	Std.Err.	[95% Conf.Interval]	
hhitn				
0	.3798165	.0376776	.3027572	.4568758
1	.6201835	.0376776	.5431242	.6972428

Note: Percentage of households with at least one ITN is 62% in the rural population.

Information on the data management and analysis tasks for data manager/data analyst can be found in Volume 2 of the RAMP toolkit, *Implementing a RAMP survey: practical field guide,* Sections 7.1 to 7.3, and discussion of the rapid dissemination of results, including detailed guidance on the production of reports in Section 7.4.

7.4 Calculation of sample weights and selection probabilities

This section introduces the methods for calculating sampling weights for non-response and imprecise estimates of PSU size. The RAMP method includes adjustment for non-response as a standard feature and adjustment for imprecise estimates of PSU size as an optional feature. Full details of calculating sampling weights are shown in Annex C. Details about the optional operational procedures for counting all households in each cluster for adjustment for imprecise estimates of PSU size are shown in Annex D.

Adjusting sampling weights for non-response has been made a main feature of a RAMP survey for several reasons. First, it is part of standard survey methodology. Second, non-response becomes more important when dealing with sensitive questions (sexual practices, violence, finances, HIV, etc.) where non-response may climb to 10 to 50 per cent. Currently, although household and individual non-response are low (less than 10 per cent) and of little consequence for most indicators in many African household surveys, non-response may increase in the future.

Adjustment for household non-response requires just one additional question (household interviewed, yes or no) and modern statistical software. All modern statistical software can adjust both the point estimates and confidence intervals for non-response. The statistical software code for adjusting for non-response is found in Step 4 of section 7.3, *Analysing the data,* and in the complete STATA data analysis code (found at www.ifrc.org/ramp).

Related to the non-response issue, RAMP surveys have a fixed number of households that are selected for interviewing per cluster. Additional or replacement households or individuals (statistical units) are not sought if households or individuals are not available for interviewing (not found or refused). Non-response often results in a different number of interviewed statistical units per cluster. Households are re-visited to re-attempt interview, but if they cannot be found after one attempt, those households are marked as "not interviewed". Quota sampling (a fixed number of statistical units per cluster, for example, 20 children per cluster) is not used. Adjusting for non-response is one of the features that brings RAMP more in line with standard survey methodology compared to the old EPI method.

7.5 Implementation of the survey

Using the example of a survey to measure bed net ownership and use following a mass LLIN distribution campaign, a RAMP survey has been planned. The steps leading to creating the survey design include:

- setting the survey objectives and what it should measure
- setting up a data analysis plan
- calculating the sample size in order to meet those objectives
- deciding on the number of households to be interviewed per cluster
- selecting clusters
- setting up the segmentation procedures and deciding on the target size of the final segment

All of these steps should take place well in advance of implementation in the field. There are many further steps to take before the RAMP survey can be carried out efficiently and effectively. Many of those steps are described in Volume 2 of the RAMP toolkit, *Implementing a RAMP survey: practical field guide*. Those steps include:

- setting up appropriate partnerships with stakeholder organizations, and appointing a survey coordinating group to oversee the design, planning and implementation of the survey
- organizing the survey: ethical considerations, time schedule, budgeting, developing the survey questionnaires, identifying and recruiting personnel, and training
- organizing the logistics and preparation of materials
- conducting fieldwork and ensuring data quality
- organizing local data management and analysis
- producing and disseminating results of the survey through reports

Training of the field personnel is a further consideration. Good training is an essential element in the success of a RAMP survey. Volume 3 of the RAMP toolkit, *Training a RAMP survey team: guide for trainers,* is a comprehensive modular five-day training programme, using the forms and tools contained in Volumes 1 and 2, intended to guide those responsible for training supervisors and interviewers who will carry out the fieldwork.

© Benoit Carpentier/IFRC

08. The RAMP website

The RAMP website (www.ifrc.org/ramp) contains electronic versions of the three-volume RAMP toolkit and other RAMP-related materials. It is intended as a dynamic resource and will be kept updated as new materials and information become available.

Each volume of the RAMP toolkit is available as a .pdf file. In addition, many tools and forms contained within the RAMP toolkit are accessible separately in a format that allows the user to modify the file, (e.g., as a word processor or spreadsheet file) through a link. The questionnaires for a post LLIN campaign malaria survey are also available. Three types of data analysis files are available:

- sample dataset for a malaria survey
- data cleaning programs, in STATA, for a malaria survey
- data analysis programs, in STATA, for a malaria survey

Links are provided for users to download the data analysis programs and the sample dataset.

Two sets of questionnaires and data programs are available on the website, one set matching the information in the published RAMP toolkit. A second set of questionnaires and data programs will be available that match any updated malaria questionnaires. As recommendations change, the malaria questionnaires will be adjusted, for example, as new questions about physical status and care and repair of nets are added. While WHO and MERG have not yet (November 2012) included indicators for physical care and repair of nets in their standard set, these are important new areas of investigation.

Other resources, such as non-malaria questionnaires, for example those covering emergencies, immunizations, nutrition or community health, will be placed on the website as they become available.

Links are also provided to other useful sites, for example Magpi, where the complete Magpi Mobile Phone User Client Guide, available in English, Spanish, French and Swahili, can be downloaded.

© Karen Bramhill/IFRC

Annex A

Magnitude of potential sampling-related bias in a RAMP survey

The RAMP method has two primary potential sampling-related biases:
- inaccurate estimate of segment sizes during selection of segments
- inaccurate estimate of size of PSUs contained in the sampling frame

The effect of the first bias, inaccurate estimate of segment sizes, on indicator estimates is likely to be very small (e.g., ≤1 per cent). A potential bias of the point estimate with magnitude of ±1 per cent or less is unlikely to be practically important if the target precision for the main indicators is as high as ±5—10 per cent.

Second, inaccurate estimates of size of PSUs contained in the sampling frame may lead to bias. This potential bias and an option to address it are discussed more fully in Annex D. In the pilot RAMP surveys, this potential bias seemed to be small. Sensitivity analyses using the four pilot surveys to estimate the size of the potential bias showed that point estimates changed by 1 per cent or less when the true size was varied by up to 100 per cent for 25 per cent of the PSUs.

Annex B

The RAMP methodology compared to the gold standard and the original EPI survey methodology

In this annex, the technical aspects of the RAMP methodologies are described in more detail, together with a side-by-side comparison of the original EPI survey[22], a gold standard survey, and the RAMP survey methodologies.

The table below shows the two key features of the gold standard method, seven features of the original EPI method that are different from the gold standard, and three key features of the RAMP method that are different from the gold standard.

Table Annex B1: Comparison of the original EPI, gold standard and RAMP survey methodologies

	Original EPI	Gold standard	RAMP
Gold standard procedures			
Enumerate all households in each PSU, comparing the count to the estimate in the sampling frame, and adjusting the selection probability with weights	No	Yes	No. An optional procedure
Select target number of households to be interviewed from a list of all households in the PSU	No	Yes	No

22 *Henderson RH, Davis H, Eddins DL, Foege WH.* Assessment of vaccination coverage, vaccination scar rates, and smallpox scarring in five areas of West Africa. *Bulletin of the World Health Organization 1973;48:183-194.*

	Original EPI	Gold standard	RAMP
EPI survey method			
Select first household in PSU by selecting a random direction from the centre of the PSU	Yes	No. Randomly selects from a long list (e.g., 100—200 households).	No. Randomly selects from a short list (e.g., 15—30 households).
Select second and subsequent households using next-nearest-door technique	Yes	No. Randomly selects from a list.	No. Randomly selects from a list.
Go back to selected households where no one is home at least twice to limit bias associated with non-response	No. Alternate households are selected to reach a quota of x statistical units per PSU	Yes. Interviewers go back twice at different times.	Yes. Interviewers return one time.
Interview a non-fixed number of households until data obtained on a fixed number of statistical units	Yes	No. A fixed number of households are interviewed.	Same as gold standard
Obtain data on a fixed number of statistical units	Yes	No. Often results in a variable number of statistical units in PSUs.	Same as gold standard
Adjust selection probability for non-response in the analysis	No	Yes. Records and adjusts for non-response.	Same as gold standard
Adjusts selection probability for differences between the estimated and counted number of households in each PSU in the analysis	No	Yes. Adjusts for selection probability according to the count.	Optional. Yes if “count” option is conducted.
RAMP survey method			
Segment and sub-segment using available data and sketch maps until a segment with 15—50 households is identified.	No	No, in rural PSUs. Sometimes segment urban PSUs, but segments are large (100—200 households)	Yes
Simple random sample of X households (where “X” is the number of households per cluster determined by the sample size calculation) from the list of 15—50 households.	No	Yes, simple random sampling but the list of households is usually much larger based on complete enumeration of PSU (usually 100—200 households)	Yes
Number of PSUs is generally 30, but any number can be used.	Yes	Generally has 200—250 PSUs across all survey domains.	Yes

Note: the immunization community revised recommendations for EPI surveys in 2005[23,24]. The revised recommendations re-emphasized the need to select the first household through a complete listing of all PSU households if possible, with the original method (random direction) as the back-up method if household listing was not possible. The method of selection of subsequent households was the same as the original EPI method, the nearest-next-door method. In addition, sampling a fixed number of statistical units (quota sampling) per PSU was maintained from the original EPI method instead of probability sampling (fixed number of households, followed by taking all statistical units in those households). The 2005 version also recommends sampling one of several target age persons if there is more than one in the household.

Quota sampling does not permit adjustment of selection probabilities for non-response. Compared to the 2005 EPI guidelines, the RAMP method has the following advantages if a PSU household list is not available: (1) provides a less costly and unbiased method of selecting the first household in all cases (when SRS cannot be done), (2) provides a better method for selecting subsequent households (through simple random sampling), (3) maintains probability sampling and eliminates quota sampling, and (4) adjusts for non-response.

23 *World Health Organization,* Immunization coverage cluster survey - reference manual. *See: www.who.int/vaccines-documents/DocsPDF05/www767.pdf.*

24 *World Health Organization.* Training for mid-level managers (MLM). 7. The EPI coverage survey. *See: www.who.int/immunization_delivery/systems_policy/MLM_module7.pdf*

Annex C
Adjusting selection probabilities and weights

The purpose of this annex is to discuss the details of calculating unadjusted selection probabilities and weights, then to present three potential adjustments of sampling weights: (1) for non-response, (2) for imprecise estimates of PSU size, and (3) with domain weights. Adjustment of the sampling weights for non-response is a standard feature of all RAMP surveys. Adjustment for imprecise estimates of PSU size is an optional feature (see Annex D for a more detailed discussion). Adjustments using domain weights are performed if the survey design includes more than one domain[25]. If two or more domains were used (for example, one rural and one urban, five provinces, or ten districts) and each domain had 30 clusters and the same number of households sampled per domain, then domain weights are needed. If there were 30 clusters in each domain and the same number of households sampled per domain, then this would be called "disproportionate stratification" because the number of households sampled in each domain is not proportional to its domain size (population). The following example illustrates the need for domain weights. Assume a survey scenario where 30 clusters are surveyed in each province or district in the country. If a national manager wants a summary national estimate, then the selection probabilities need to be adjusted because the population sizes of the provinces or districts (survey domains) will be unequal.

If there is more than one domain, then the domain weights are combined with the adjustments for non-response and the optional adjustment for imprecise estimates of PSU size to produce the final adjusted sample weights.

Unadjusted selection probabilities/weights

This section describes the calculation of the "unadjusted" selection probabilities and weights due to the two-stage cluster survey design before adjustment for non-response and other factors. There are four or five phases of the RAMP sampling process (if there are only segments, then the third phase of sub-segments is skipped):

- First phase: selection of PSUs by PPES
- Second phase: selection of one segment by PPES
- Third phase (if done): selection of one sub-segment by PPES
- Fourth phase: selection of a fixed number of households by SRS from the segment/sub-segment.
- Fifth phase: selection of all household members in the target population

Based on the RAMP design, statistical units have an "unadjusted" selection probability and weight prior to being adjusted for non-response and other

25 *In this manual, we use the term "domain weights" and not "strata weights". A domain is a specific type of stratum. Survey specialists use the term "strata weights". However, the RAMP method does not involve non-domain strata considerations, so the term "domain weight" is more applicable.*

factors. Prior to adjustment, the RAMP method has an equal probability sample design since it uses PPES and SRS to select households. An equal probability design means that each statistical unit (household or person) is assumed to have an equal probability of being selected. Simplistically, the "unadjusted" selection probability of all statistical units is the same: the total number of households sampled divided by the total number of households in the sampling frame.

The formula for illustration of selection probabilities has six parts:

1. First part: number of PSUs selected.
2. Second part: size of PSU/size of target population from the sampling frame. Size can be population[26] or households. Households will be used as the measure of size in the example.
3. Third part: size (number of households) of selected segment/size of PSU (in number of households).
4. Fourth part: size (number of households) of selected sub-segment/size of the segment (in number of households).
5. Fifth part: number of households selected by SRS/size (number of households) of selected sub-segment.
6. Sixth part: probability = 1 because all members of the target population in each household are selected (all members in both numerator and denominator).

Using the six parts, a formula for calculating the unadjusted selection probabilities for individuals is the following:

$$\underset{\text{Part 1}}{\textit{Num PSUs}} \times \underset{\text{Part 2}}{\frac{\textit{PSUsize}}{\textit{TotPopsize}}} \times \underset{\text{Part 3}}{\frac{\textit{Segsize}}{\textit{PSUsize}}} \times \underset{\text{Part 4}}{\frac{\textit{SubSegsize}}{\textit{Segsize}}} \times \underset{\text{Part 5}}{\frac{\textit{numHHbySRS}}{\textit{SubSegsize}}} \times \underset{\text{Part 6}}{1}$$

where:

- "Num PSUs" is the number of PSUs selected at the first stage/phase (30 in our example)
- "PSUsize" is the size (preferably using households; data from the sampling frame) of the selected PSU
- "TotPopsize" is the size of the total target population (preferably using households; data from the sampling frame)
- "Segsize" is the size (in households) of the one segment selected by PPES (from the sketch map or household list)
- "SubSegsize" is the size (in households) of the one sub-segment selected by PPES (if done) (from the sketch map or household list)
- "numHHbySRS" is the number of households selected by SRS (10 in our example) in each selected segment/sub-segment in each PSU
- "1" is the probability of selection in the household, which equals "1" because all target population members in the households are selected

From the formula above, notice that many of the terms are similar and can be cancelled:

$$30\ \textit{PSUs} \times \frac{\cancel{\textit{PSUsize}}}{\textit{TotPopsize}} \times \frac{\cancel{\textit{Segsize}}}{\cancel{\textit{PSUsize}}} \times \frac{\cancel{\textit{SubSegsize}}}{\cancel{\textit{Segsize}}} \times \frac{\textit{numHHbySRS}}{\cancel{\textit{SubSegsize}}} \times 1$$

26 *It is optimal to use households as the measure of size if available, since households are used as the measure of size in the second and subsequent stages. If one has to use population and convert population into households by using an average number of persons per household for all PSUs, then a small amount of bias could be introduced since average household size could be different by PSU. However, the amount of bias is likely to be very small and unlikely to change the interpretation of results.*

After cancelling terms, the following un-cancelled terms are left:

$$\frac{(30 \text{ X } numHHbySRS)}{TotPopsize} = n / N;$$

Where TotPopsize is measured in terms of number of households; n = total number of households selected over all clusters in the sampling frame; N = total households in the sampling frame.

This makes sense since the selection probability is the number of PSUs (usually 30 PSUs) times the number of households selected per PSU (for example, 10) divided by size (in households) of the target population in the sampling frame (for example, 300,000 households or 1.5 million persons). The selection probability would be calculated as 300 households ÷ 300,000 households or 1 ÷ 1,000. Therefore, each person sampled represents 1,000 persons in the sampling frame and the unadjusted sampling weight is 1,000 (the weight is the reciprocal of the selection probability). The selection probability for households is the same as for individuals (1 ÷ 1,000), and therefore each household represents 1,000 households (household sampling weight is 1,000).

Adjusting selection probabilities

However, the final probability of selection of statistical units is not equal for all households/persons in most surveys because of non-response and imprecise measures of size used for selection of PSUs. In addition, selection probabilities need to be adjusted if the survey design included more than one domain and had disproportionate stratification. Below are four separate possible adjustments to selection probabilities in a RAMP.

1. Non-response, household
2. Non-response, individual, given that household responded
3. Imprecise size, PSU level
4. Domain weights if the survey design included more than one domain (e.g., 30 clusters in each domain)

Since non-response and imprecise size can be different by PSU, adjustments are made by PSU. The following paragraphs look at adjustments to the sampling weights instead of adjustments to selection probabilities. Mathematically, they are equivalent since the sampling weight is the reciprocal of the selection probability.

Non-response: There can be two stages of non-response: household and individual. In household surveys in much of Africa, non-response rates are low (5 per cent or less). In those instances, non-response is unlikely to change the survey result by more than 1—2 per cent. However, non-response rate can be higher if there are questions about sensitive subjects such as sexual practices or financial issues.

Non-response at the household level: A question about household non-response must be added to the household questionnaire to measure non-response. Once this question is added to the survey, it is easy to adjust the results for household non-response.

Non-response at the individual level: If some questions are sensitive, individuals may refuse to respond even though they are at home and available to respond. If this situation is likely, then a question about individual non-response should be added to the individual questionnaire. The categories might be[27]:

- Completed interview
- Not at home
- Postponed
- Refused interview
- Partly completed interview
- Incapacitated

In the RAMP malaria individual questionnaire, there is no individual non-response question because malaria information is not sensitive.

The unadjusted sampling weights are adjusted for non-response by multiplying by the reciprocal of the response rate for each PSU; the response rate = number of households on which data were collected through interviewing ÷ number of households that were supposed to be interviewed. For example, if 10 households were selected for interviewing and only 9 are available for interviewing in PSU Number 24, then the response rate is 90 per cent or 9 out of 10 households. The unadjusted weight (reciprocal of the selection probability) for those nine households would be increased by multiplying by the reciprocal of the response rate, 10 ÷ 9 or 1.11. This makes sense: if one or two households could not be interviewed, then the "weight" of the information of those 8—9 remaining households needs to be increased since they will be "representing" all 10 households in the analysis.

Individual non-response rates: If you include a question about individual non-response in the individual questionnaire, then the sampling weights should be adjusted by multiplying by the reciprocal of the individual response rate. The individual response rate (in each PSU) = number of individuals on which data were collected ÷ number of individuals that were eligible for data collection. For example, if young women 15—24 years old were the target population and 4 of 17 young women did not agree to participate or were not available to be interviewed (despite availability of other members of the household), then the individual response rate would be equal to (17 − 4) ÷ 17 or 13 ÷ 17 and the sampling weight would be adjusted for individual non-response by multiplying the sampling weight by the reciprocal (17 ÷ 13).

Adjustments due to imprecise size estimation of PSU: Bias can be introduced if the ratio of the estimate of the PSU size to the true size is different among PSUs. For example, if data from a 10 year old census were used for the sampling frame and the growth rate was 10 per cent in one PSU and 40 per cent in another PSU, then some bias may be introduced (if the indicator measures are different) if adjustments are not made. (If the ratio is similar for all PSUs, for example 20 per cent growth rate for all PSUs, then bias will not be introduced). Imprecise estimates of PSU size can be adjusted if all the households in each PSU are counted. The topic of counting all the households in a PSU is addressed in Annex D.

Theoretically, the formula for the weight for statistical unit x (e.g., household or individual) that includes adjustment for imprecise PSU size is:

- Weight statistical unit $\chi = 1 \div [a(M_{\alpha} \div \Sigma M_{\alpha}) \times (b \div B_{\alpha})]$

27 *Demographic and Health Survey.* Interviewer's Manual. *ICF International. Calverton, Maryland. May 2012.*

where:

- a is the number of clusters in sample
- b is the number of households selected per cluster
- M_α is the estimated size for cluster α (from sampling frame data)
- B_α is the actual size of cluster α (from the count)
- ΣM_α is the estimated number of households in the whole sampling frame.

Practically, the weights would be adjusted by multiplying by the reciprocal of the ratio of estimated number of households from the sampling frame (M_α) divided by the number of the "counted" households (true number) (B_α) for each selected PSU. For example, if the estimated number of households in a PSU was 400 and a house-to-house count revealed 500 households, then the weights would be increased by multiplying by 500 ÷ 400 or 1.25. This makes sense since if the PSU is truly larger than estimated, then the "weight" of data from those households should be increased.

Domain weights if the survey design included more than one domain: If the survey design included more than one domain and uses the RAMP method of 30 clusters in each domain, then domain weights need to be calculated. The most common scenario for this is when independent samples are done within survey domains; either because independent point estimates within domains are sought with a specified level of precision, or because comparisons between domain point estimates are desired.

Estimating the domain weights is quite simple. Consider a different and simpler scenario where we have three domains (districts) making up a province. In the example, 30 clusters were sampled in each of the three districts in the province and the provincial manager wants an overall provincial estimate. The sampling weight for each stratum (district) is calculated using the following formulae:

Weight domain 1 = target population size domain 1 (district 1) divided by the target population across domains 1—3 (province)

Weight domain 2 = target population size domain 2 (district 2) divided by the target population across domains 1—3 (province)

Weight domain 3 = target population size domain 3 (district 3) divided by the target population across domains 1—3 (province)

Σ domain weights 1—3 = 1.0 (The sum of the three domain weights equals one)

These domain weights can be calculated using any computer analysis program, including Excel.

The domain weights are used in adjusting the final weights as shown in the next section.

Summary formula for calculating the final, adjusted weights in each PSU: This formula has four potential "adjustment" terms: (1) non-response at the household level (2) non-response at the individual level (3) imprecise estimate of size at the PSU level (4) domain weights:

$$= Unadj.wt \times \frac{Sel.HHs}{Interviewed\ HHS} \times \frac{Sel.Ind.}{Interviewed\ Ind.} \times \frac{Ct.HHs.PSU}{Est.HHs.PSU} \times \text{domain weights}$$

where:

- Unadj. wt = unadjusted sampling weight
- Household non-response: Sel. HHs = selected households (number of households that were supposed to be interviewed). Interviewed HHs = households that were interviewed.
- Individual non-response: Sel. Ind. = selected individuals (number of target persons that were supposed to be interviewed). Interviewed Ind. = interviewed targeted persons
- Inaccurate PSU size: Ct. HHs, PSU = counted (true) households at the PSU level. Est. HHs, PSU = estimated households at the PSU level.

"Domain weights" are included if multiple domains were present and disproportionate stratification was used.

All RAMP surveys will have at least one adjustment term (household non-response). Some surveys may also be adjusted for individual non-response, inaccurate PSU size, and with domain weights.

Adjusting selection probabilities and weights can take time and extra effort. Survey planners need to decide which adjustments may provide sufficient information value based on additional cost. Adjustment for household non-response is easy and should be done in all surveys. Adjustments for non-response at the individual level depend on the likely frequency of individual non-response and the sensitivity of the questions. Adjustments due to imprecise estimates at the PSU level will require extra effort and cost that must be weighed against the added benefit (see a fuller discussion of this issue in Annex D).

These adjustments to the weights can be calculated using various types of computer analysis programs. The STATA code for calculating the final adjusted weights can be found in the complete STATA code for the example malaria survey: see the RAMP website (www.ifrc.org/ramp).

Annex D

Adjusting selection probabilities at the PSU level by counting the total number of households in the PSU

Potential bias can result if the estimated sizes of the PSUs used in the sampling frame are inaccurate. This potential bias can be reduced or eliminated if all the households in the PSU are counted and that count used to adjust the selection probabilities. Counting all the households in the PSU to adjust selection probabilities is not a standard part of the RAMP survey method, but can be added as an option. This annex describes how to add the "counting" option, including the theoretical basis of "counting", and corresponding operational issues.

As part of a gold-standard cluster survey, the number of households in the PSU is counted in all PSUs. During analysis, the selection probability is adjusted by comparing the number of counted households with the estimated number of households on the sampling frame (see Annex C) that was used to select PSUs.

There will be no bias due to varying probability selections across observations (and counting would not be necessary) if the difference between the sampling frame estimate and counted households is zero or if the percentage difference between PSUs is the same. For example, if the data from a 10 year old census are used in the sampling frame and all PSUs truly have exactly 20 per cent more households compared to the census (see PSUs 1—5 in the Table below), there will be no bias due to varying probability selections across statistical units.

However, if the percentage difference between the sampling frame estimate and the true number (counted) is different between PSUs (see PSUs 6—8 in the Table), for example, PSU 6 has twice the number of counted households, PSU 7 has 50 per cent fewer households, and PSU 8 has 200 per cent more, then bias may result due to varying probability selections across observations. Bias will result if two conditions are met: (1) there are differences in the percentage difference of counted (true) versus estimated households as discussed above, and (2) the prevalence of the survey indicators in those PSUs with a deviation in the

percentage difference in the counted versus estimated (e.g., clusters 6—8 in the Table) is different compared to the other PSUs. For example, bias will result if the percentage difference is 100 per cent in five PSUs, 20 per cent different in the rest of the PSUs, and the intervention prevalence is 40 per cent in the five PSUs and 80 per cent in the rest. The magnitude of the bias depends on both factors.

Table Annex D1: Example data: adjusting selection probabilities due to imprecise size estimate of PSUs/ clusters

Number of the PSU	Estimated number of households in the PSU from the sampling frame	Number of households in the PSU during the count*	% difference, using the sampling frame estimate as the denominator
1	100	120	+ 20%
2	150	180	+ 20%
3	200	240	+ 20%
4	100	120	+ 20%
5	150	180	+ 20%
6	100	200	+ 100%
7	100	50	- 50%
8	100	300	+ 200%
…			

* The "count" of all households in the PSU is assumed to equal the true number of households in the PSU.

Operational factors

The gold-standard method of counting would be for personnel employed by the survey organizers to visit each household and number with chalk or list all the households in each PSU. Another less expensive (and possibly less accurate method) is to ask community members or volunteers to do the counting by visiting each household or by using local data.

The sampling weights adjusted for inaccurate PSU size for each statistical unit are then estimated:

Weight for statistical unit x = $1 \div [a(M_{\alpha} \div \Sigma M_{\alpha}) \times (b \div B_{\alpha})]$;

where a is the number of clusters in sample, b is the number of households selected per cluster, M_{α} is the estimated size for cluster α (from sampling frame), B_{α} is the actual size of cluster α (after counting). This weight can be calculated in STATA or Excel prior to importing the data into STATA for analysis.

Practical guidance

If the data from the sampling frame are recent (recent immunization data, census data within the last 3—5 years, etc.), thought to be reasonably accurate, and differences between the PSUs thought to be similar, then counting all households in the PSU may not be cost-beneficial. However, if the sampling frame data are old (census is more than 10 years old, urban areas have grown much faster than rural areas, inaccurate (census data known to be inaccurate), and major differences are possible between PSUs, then counting may be cost-beneficial.

Cost

In a RAMP survey, segmentation and random sampling of households generally takes up to one-half day and interviewing usually takes a further one-half day per PSU. Counting is likely to take a full day, resulting in a doubling of the field costs if payment of personnel is necessary to complete the counting. Field costs represent approximately one-third to one-half of the costs of a RAMP survey.

Counting of all households in a PSU is likely to be impractical and very costly if several PSUs have more than 300 households.

Annex E

Survey specifications: planning worksheet

Health indicator: _____

Denominator of indicator: _____

Confidence level: _____

Precision/Confidence interval: _____

Estimate design effect: _____

Expected value: _____

Sample size of statistical units: _____

Sample size of households: _____

Final sample size of households after increasing to account for expected non-response: _____

Number of clusters: _____

Statistical units/ households to interview per cluster: _____

Annex F
Data analysis plan: table shells

The table shells below, which coincide with table shells listed in Section 7.1 are examples of what survey planners might prepare for the data analysis plan for a malaria survey using the RAMP method. The figures, tables and table numbers below match the figures and tables in the "results bulletin" shown in Volume 2 of the RAMP survey toolkit, *Implementing a RAMP survey: practical field guide,* Annex I.

Analysis plan table 1: Key descriptive information about survey

Indicators	n	%
1. Sample domain (population) *Population variable in "SAM FRAME data" (sampling frame) database*	xxx	
2. No. cluster *Household database, question 4 cluster number*	xxx	
3. No. households per cluster *Household database, survey design document or question 4 (cluster number) and 5 (household number)*	xxx	
4. No. of households targeted for interview *Household database, question 5 household number*	xxx	
5. No. of households interviewed *Household database, question 6 (interviewed household yes or no)*	xxx	X% of number targeted
6. No. persons (all ages) *Person database, count*	xxx	
7. Average household size *Relate person database to household database, question 5 (household number) in household database*	xxx	
8. No. children under five years old *Person database, question 8 about age in years*	xxx	X% of all persons
9. Rural households (%) *Household database, question 9 rural/urban status*		X% of total households
10. Average unadjusted household sample weight *Household database, unadjusted sample weight calculated in analysis code, averaged using question 5 household number*	xxx	

11. Average adjusted household sample weight *Household database, adjusted sample weight calculated in analysis code, using question 5 household number and question 6 about household interview status*	xxx	
12. No. sleeping spaces *Household database, question 14 number of sleeping spaces*	xxx	
13. No. nets *Net database, count of records*	xxx	X% of sleeping spaces
14. No. nets observed *Net database, question 6 about observation*	xxx	X% of nets
15. No. nets hanging last night *Net database, question 7 about hanging*	xxx	X% of nets
		X% of sleeping spaces
16. No. ITNs *Net database, questions 10, 11, and 12 indicating whether a net was an ITN, related to household database*	xxx	X% of all nets
		X% of sleeping places
17. No. persons per sleeping space *Person database, number of persons in each household related to household database with question 14 on number of sleeping places*	xxx	
18. Average no. persons that slept under a net (from net roster data) *Net database, questions 14—18 on persons sleeping under each net*	xxx	

Analysis plan table 2: Number of ITNs needed (ITN gap) to reach universal coverage

	n	95% confidence interval (CI)
Total ITNs needed for universal coverage "SAM FRAME data" (sampling frame) database, population variable		—
Estimated number of ITNs in the whole survey domain Net database, questions 10, 11, and 12 indicating whether a net was an ITN in the net database and calculated adjusted sample weight		
ITN gap Subtraction of previous two indicators		

Analysis plan table 3: Main and secondary malaria indicators for main graph

Indicator	Point estimate	95% CI
Main indicators		
1. Proportion of households with at least one ITN *Net database, questions 10, 11, and 12 indicating whether a net was an ITN, related to household database*		
2. Proportion of households with at least one ITN for every two people *Net database, questions 10, 11, and 12 indicating whether a net was an ITN, person database shows number of persons in the household, both related to household database*		
3. Proportion of population with access to an ITN within their household *Net database, questions 10, 11, and 12 indicating whether a net was an ITN, person database shows number of persons in the household, net database related to person database*		
4. Proportion of population who slept under an ITN the previous night *Net database, questions 10, 11, and 12 indicating whether a net was an ITN and questions 14—18 indicating who slept under the net last night, related to the person database*		
5. Proportion of children under five years old who slept under an ITN the previous night *Net database, questions 10, 11, and 12 indicating whether a net was an ITN and questions 14—18 indicating who slept under the net last night, related to the person database, question 8 about age*		
6. Proportion of households with at least one ITN and/or sprayed by IRS in the last twelve months *Household database, question 20 about IRS; net database, questions 10, 11, and 12 indicating whether a net was an ITN, related to household database*		
7. Proportion of children under five years old with fever in the last two weeks who had a finger or heel stick *Person database, question 10 about blood taken, question 9 about fever in the last two weeks, and question 8 about age*		
8. Proportion of children under five years old with fever in the last two weeks for whom advice or treatment was sought *Person database, question 11 about advice or treatment, question 9 about fever in the last two weeks, and question 8 about age*		
9. Proportion receiving first line treatment, among children under five years old with fever in the last two weeks who received any anti-malarial drugs *Person database, question 17, 19, and 21 about first-line or any anti-malarials, question 9 about fever in the last two weeks, and question 8 about age*		
Secondary indicators		
10. Proportion of ITNs that were used last night *Net database, questions 10, 11, and 12 indicating whether a net was an ITN in the net database and questions 14—18 indicating who slept under the net last night*		
11. Proportion of households sprayed by IRS in the last 12 months *Household database, question 20 about IRS*		
12. Proportion of households in which a community volunteer visited this household in the last six months to talk about malaria or mosquito nets *Household database, question 17 about visit from community volunteer*		

Analysis plan table 4: ITN indicators by wealth quintile

Wealth quintile data come from household database questions 23—54.

Proportion of households with at least one ITN
Net database, questions 10, 11, and 12 indicating whether a net was an ITN, related to household database

Use of ITNs, all persons
Net database, questions 10, 11, and 12 indicating whether a net was an ITN and questions 14—18 indicating who slept under the net last night, related to the person database

Wealth quintiles	Proportion of households with at least one ITN	Use of ITNs, all persons
Q1		
Q2		
Q3		
Q4		
Q5		
Ratio Q5/Q1		
Total (all 5 wealth quintiles)		

Analysis plan table 5: Age of ITN in years

Net database, question 8 about age of net

Age of ITN in years	Percentage	Cumulative percentage
<1		
1		
2		
3		
4		
5+		
Do not know		

Analysis plan table 6: Percentage of households that received a home visit from a community volunteer in the last 6 months

Household database, question 17 about home visits

	Point estimate	95% CI
Percentage of households that received a home visit from a community volunteer in the last 6 months		

Analysis plan table 7: Source of ITNs

Net database, question 9 about where obtained net

	Point estimate	95% CI
Mass campaign		
Market		
Clinic		
Other		

Analysis plan table 8: Greatest source of information about ITN use

Household database, question 18 about where greatest source of information on the use of nets

	Point estimate	95% CI
Health centre staff		
Community-based volunteer		
Community leader		
Radio		
Neighbour		
Other		

Analysis plan table 9: Malaria testing and treatment

	Point estimate	95% CI
Children with fever in the last two weeks *Person database, question 9 about fever and question 8 about age*		
Proportion of children under five years old with fever in the last two weeks who had a finger or heel stick *Person database, question 10 about blood taken, question 9 about fever, and question 8 about age*		
Proportion of children under five years old with fever in the last two weeks for whom advice or treatment was sought *Person database, question 11 about advice or treatment, question 9 about fever, and question 8 about age*		
Proportion receiving first line treatment, among children under five years old with fever in the last two weeks who received any anti-malarial drugs *Person database, questions 17, 19, 21 treatments taken, question 9 about fever, and question 8 about age*		

Analysis plan table 10: Precision and design effect of key indicators

	CI, ±%	DEFF	DEFT
Household ITN ownership *Net database, questions 10, 11, and 12 indicating whether a net was an ITN, related to household database*			
ITN use *Net database, questions 10, 11, and 12 indicating whether a net was an ITN and question 14—18 indicating who slept under the net last night, related to the person database*			
Proportion of children under five years old with fever in the last two weeks who had a finger or heel stick *Person database, question 10 about blood taken, question 9 about fever, and question 8 about age*			

Analysis plan table 11: ITN use by age group

Net database, questions 10, 11, and 12 indicating whether a net was an ITN and questions 14—18 indicating who slept under the net last night, related to the person database, question 8 about age in person database

Age groups	Percentage
0-4y	
5-9y	
10-14y	
15-24y	
25-44y	
45-59y	
60+y	
Total	

Analysis plan table 12: Type and brand of net

Net database, question 10 about net brand and questions 10, 11, and 12 indicating whether a net was an ITN

	n	% of total
LLIN Brand XXX1		
LLIN Brand XXX2		
LLIN Brand XXX3		
Non-LLIN ITN		
Non-ITN		
Missing or did not know		

Analysis plan table 13: Percentage of persons sleeping under ITNs last night

Net database, questions 14—18 about persons sleeping under net and questions 10, 11, and 12 indicating whether a net was an ITN. Net database related to person database.

	n, nets	% of total
1 person		
2 persons		
3 persons		
4+ persons		

Mean number of persons per ITN = x.xx

Annex G

Annotated bibliography: where to go for further written advice

The three volumes making up the RAMP toolkit contain the most detailed and comprehensive operational guidance on conducting rapid, inexpensive, mobile phone-based field surveys. Below, other types of guidance about surveys in developing countries are listed. Survey planners may find these documents useful to give more details about the technical aspects of survey sampling, or about surveys in more specific areas (emergencies, nutrition, mortality, etc.).

1. General survey reference on the technical aspects of sampling

Levy PS and Lemenshow S (1999), *Sampling of Populations: Methods and Applications.* Third ed. New York: John Wiley and Sons, Inc.

This is a comprehensive book on sampling principles and methods, covering topics from the design of surveys, implementation of various sampling methods (simple random sampling to complex multi-stage designs), standard error estimation to issues concerning the interpretation of survey results.

2. General household survey guidelines in developing countries

United Nations Statistical Division household survey guidelines: *Designing Household Survey Samples: Practical Guidelines,* 2008. See: unstats.un.org/unsd/publication/seriesf/Seriesf_98e.pdf

This contains extensive guidelines on many aspects of household surveys in developing countries. It is one of the most complete guides.

3. Gold-standard general health surveys

Demographic and Health Surveys. *Measure DHS*. Four sets of manuals: *Fieldwork Manuals, Sampling Manuals, Report Tabulation, Statistical and Methodological Documentation*.
See: www.measuredhs.com/What-We-Do/Survey-Types/DHS-Manuals.cfm

The manuals cover very large (more than 5000 households, more than 200 clusters), complex, resource-intensive surveys. Data are first recorded on paper in most surveys. These may be the most up-to-date guidelines for large gold-standard health-related household surveys in developing countries.

United Nations Children's Fund (UNICEF). Division of Policy and Planning. *Multiple indicator cluster survey manual*, 2005.
See: www.childinfo.org/files/Multiple_Indicator_Cluster_Survey_Manual_2005.pdf

Guidance on UNICEF's Multiple Indicator Cluster survey.

4. Gold-standard malaria surveys

Roll Back Malaria (2012). *Malaria Indicator Survey: Basic Documentation for Survey Design and Implementation*.
See: www.who.int/malaria/publications/atoz/9241593571/en/index.html

Very large (more than 5000 households, more than 200 clusters), resource-intensive surveys. Data are first recorded on paper in most surveys. These guidelines are similar to the DHS survey guidelines and may be the most up-to-date guidelines for large gold-standard malaria-related household surveys in developing countries.

5. Vitamin and mineral status surveys

Gorstein J, Sullivan KM, Parvanta I, Begin F (2007). *Indicators and Methods for Cross-Sectional Surveys of Vitamin and Mineral Status of Populations*. The Micronutrient Initiative (Ottawa) and the Centers for Disease Control and Prevention (Atlanta).
See: www.micronutrient.org/CMFiles/PubLib/Indicators-for-Cross-Sectional-Surveys1IYA-3242008-2823.pdf

A recent publication with many important and useful topics on field surveys.

6. Emergencies

Standardized monitoring and assessment of relief and transitions (SMART). *Measuring mortality, nutritional status and food security in crisis situations*. Version 1
See: www.smartindicators.org

This provides information on how to measure nutrition, mortality and other indicators in emergency settings. It does not contain many operational details.

Bilukha OO, *Old and new cluster designs in emergency field surveys: in search of a one-fits-all solution*. Emerg Themes Epidemiol. 2008 Jul 8;5:7.
See: www.ncbi.nlm.nih.gov/pmc/articles/PMC2474831

This is a recent article discussing 30 by 30, 33 by 6, 30 by 7 and other survey designs in emergency situations.

7. EPI/immunization surveys

World Health Organization. *Immunization coverage cluster survey-Reference manual*. Department of Immunization, Vaccines and Biologicals, World Health Organization, Geneva, 2005, WHO/IVB/04.23. See: www.who.int/vaccines-documents/DocsPDF05/www767.pdf

The current guidance is to conduct simple random sampling at the second stage if possible, with the old EPI method as a backup if simple random sampling cannot be done.

World Health Organization. *The EPI coverage survey*.
See: www.who.int/immunization_monitoring/routine/EPI_coverage_survey.pdf

This is the original EPI survey method, 1991 version. This method started the genre of rapid field surveys in the 1970s.

8. Rapid surveys

UCLA Department of Epidemiology. School of Public Health.
See: www.ph.ucla.edu/epi/rapidsurvey.html

This is a history of the WHO immunization coverage surveys, methods, software, and online course. From a historical perspective, RAMP surveys are related to the concepts described at this website.

9. International Federation of Red Cross and Red Crescent Societies publications

Planning and evaluation department, 2010. *Project/programme planning (PPP): Guidance manual*. See: www.ifrc.org/Global/Publications/monitoring/PPP-Guidance-Manual-English.pdf

Planning and evaluation department, 2011. *Project/programme monitoring and evaluation (M&E) guide*. See: www.ifrc.org./Global/Publications/monitoring/IFRC-ME-Guide-8-2011.pdf

Planning and evaluation department, 2011. *IFRC Framework for Evaluation*. See: www.ifrc.org/Global/Publications/monitoring/IFRC-Framework-for-Evaluation.pdf

The Fundamental Principles of the International Red Cross and Red Crescent Movement

Humanity The International Red Cross and Red Crescent Movement, born of a desire to bring assistance without discrimination to the wounded on the battlefield, endeavours, in its international and national capacity, to prevent and alleviate human suffering wherever it may be found. Its purpose is to protect life and health and to ensure respect for the human being. It promotes mutual understanding, friendship, cooperation and lasting peace amongst all peoples.

Impartiality It makes no discrimination as to nationality, race, religious beliefs, class or political opinions. It endeavours to relieve the suffering of individuals, being guided solely by their needs, and to give priority to the most urgent cases of distress.

Neutrality In order to enjoy the confidence of all, the Movement may not take sides in hostilities or engage at any time in controversies of a political, racial, religious or ideological nature.

Independence The Movement is independent. The National Societies, while auxiliaries in the humanitarian services of their governments and subject to the laws of their respective countries, must always maintain their autonomy so that they may be able at all times to act in accordance with the principles of the Movement.

Voluntary service It is a voluntary relief movement not prompted in any manner by desire for gain.

Unity There can be only one Red Cross or Red Crescent Society in any one country. It must be open to all. It must carry on its humanitarian work throughout its territory.

Universality The International Red Cross and Red Crescent Movement, in which all societies have equal status and share equal responsibilities and duties in helping each other, is worldwide.

For more information on this IFRC publication, please contact:

In Geneva
Jason Peat
Senior Health Officer, Malaria
E-mail: jason.peat@ifrc.org

www.ifrc.org
Saving lives, changing minds.