“How to carry out LLIN durability monitoring” and FAQ

# Purpose:

* To address critical aspects of all phases of LLIN monitoring (design, implementation, analysis, reporting) by presenting the issues and arguments for the current guidance, discuss alternative options and answer frequently asked questions.

# Target audience

* Researchers and malaria program staff involved in LLIN durability activities (PMI and beyond)

# Part 1: Design

## Moving from study to monitoring

The initial studies on durability of LLIN in the field were mainly focused on looking at the acquisition of holes and loss of insecticidal effects over time in a given environment and in some cases the objective was to compare the field performance of one or more different brands of LLIN. The study sites were usually chosen purposively to “represent” a particular environment and were limited in size to a few communities sometimes in more than one location within the country.

As the objectives of the field assessment of LLIN durability now shifts to monitoring of LLIN survival to inform program decisions on when to replace LLIN and to determine which brands perform best in a given target population (possible procurement implications), the design and sampling strategy also need to be reconsidered. The major objective of the monitoring activity now is to obtain a realistic picture of what happens to the population of nets given out through a campaign. The following considerations present the options, explore strengths and weaknesses and suggest possible solutions.

## Monitoring Protocol

Because to date there is no methodology that allows reliable assessment of insecticidal effectiveness in large field samples, the insecticide aspect of LLIN durability is only examined in smaller sub-samples and this can be integrated in any study design as discussed later.

The principle epidemiological design options for monitoring physical LLIN durability are shown in Figure 1 and can be described as follows:

1. **Prospective, longitudinal:** A sample of nets (cohort) is identified shortly after the distribution campaign and marked with unique identifiers. This cohort is then followed over time in multiple (e.g. annual) surveys.
   1. **Strength**:
      1. For each net a definite outcome can be determined: net still there, or net is lost and reasons for loss, or household moved away and is lost to follow-up. This is the best way to measure attrition and has an additional advantage that it allows to capture changes of time in household or net associated factors that drive durability.
   2. **Weakness:** 
      1. The approach requires that nets from the distribution can be clearly identified (e.g. by brand label, size, color and reported age of net) and distinguished from other nets within the household obtained from other sources (previous campaigns, continuous distribution, or private sector).
      2. Longitudinal follow-up of the same nets has the potential of inducing a Hawthorne effect, i.e. net users – when aware that the net is being followed up – may change their behavior and keep it much longer than they otherwise would. While this has been shown to be true, it does not really interfere with the monitoring of durability as the estimate of net survival always includes the assessment of both, attrition and the physical integrity of surviving nets. This implies that in this design the relative contribution of attrition to physical durability will be lower than in a “natural” setting, but the overall estimate of net survival will be very similar as the proportion of surviving nets in torn condition will be accordingly higher.
2. **Prospective sub-population, cross-sectional (or longitudinal):** A sub-population of the nets designated for distribution (campaign) are marked or “tagged” before the distribution and channeled into the distribution process. They are then identified in the field and sampled for follow-up over time either in multiple independent cross sectional surveys or longitudinally. This design most closely resembles that of previous studies where clearly marked nets are distributed by the study team.
   1. **Strength:**
      1. Nets from the distribution can be easily identified and distinguished from other nets.
   2. **Weakness:**
      1. After distribution the whereabouts of these nets is unknown and the exact location needs to be established before a sampling frame can be established.
      2. If a longitudinal approach is chosen, the potential Hawthorne effect as discussed above will apply.
3. **Retrospective, cross sectional:** Nets are not identified at the time of distribution but rather identified retrospectively at each time point during independent cross-sectional surveys.
   1. **Strength:**
      1. Durability of nets from the campaign can be monitored even when the time point of distribution as start of monitoring activities has been missed.
   2. **Weakness:**
      1. At each survey the number of nets received per household and those lost since the campaign has to be established from recall by the respondent and as time goes on will increasingly be prone to recall bias (under-estimation of attrition). Reliable estimates of net survival can only be obtained if some recall adjustment is made.

Figure : Design options for post-campaign monitoring of LLIN durability



*What should be considered for the inclusion or not of measurement of insecticidal effectiveness?*

## Design considerations for monitoring insecticide

The above considerations refer only to the monitoring of the physical durability of LLIN. Aspects of monitoring the insecticidal effectiveness need special considerations:

* As stated previously, to date there is no methodology that would allow routine monitoring of insecticidal effectiveness through bio-assays or chemical residue determination without taking samples from a net to a laboratory or insectary, i.e. without destroying the net. Therefore, nets for these assessments need to be collected separately from the cohort of nets to be monitored for physical durability if the (recommended) prospective design is used.
* Because the methodology to test insecticidal effectiveness is either labor intensive (bio-assays/chemical analysis) or expensive (chemical analysis), the number of nets to be monitored at each time point needs to be much smaller than for physical durability and the general recommendation is to use a sample of 30 per site and time point.
* The decision about which time points to include insecticidal effectiveness monitoring depends on the availability in-country of an entomological laboratory and insectary, its capacity, as well as on availability of funds for the chemical residue analysis. The minimum would be to include insecticidal effectiveness monitoring at the end of the activity, i.e. after three years, but an annual assessment of at least bio-assays would be desirable.

## Sampling

### Sampling Strategies

In this section sampling strategies are considered, first establishing the sampling frame and selecting households to be sampled and then discussing the question whether only one or all eligible nets per household should be included in the monitoring.

### Tagged net distribution (sub-population) versus post-campaign sampling

**Prospective, longitudinal**

In this design campaign nets are distributed through the campaign and some might be “lost” during the process and some may remain in stores (Figure 2). The majority will, however, reach the target population and these nets form the “population of interest” of which the monitoring activity aims to get as realistic a picture as possible. Ideally, all these nets would form the sampling frame, but as such lists do not exist, all households within the target area are considered the sampling frame. Sampling then proceeds as in any representative household survey with a two stage cluster sampling where first clusters (villages) are selected with probability proportionate to size (PPS) and the households are selected from an updated list of all households in the cluster (with sectioning in case of communities larger than 200 households) either by using random number lists or systematic sampling (“random walk” methodology should be avoided if possible). Sample size considerations are not presented here in detail, but the number of clusters per sampling domain should not be lower than 15, but could be selected from any target area (district, province, or region). All households that have received nets from the campaign are included in the sample and sampled households that never got a campaign net (usually very few) will be replaced. Campaign nets in sampled households are marked with inedible ink and a unique identifier (ID number or bar-code) and form part of the study cohort. This sampling exercise should ideally take place within three months following the campaign and not later than six months after distribution. This sampling approach provides an optimal “representativeness” of campaign nets within the target area equivalent to any MIS or DHS survey.

Figure : Sampling in prospective, longitudinal design: A) nets are distributed during campaign of which the majority is reaching the targeted households (dark grey circles); B) all households in the population of interest form the sampling frame and first clusters (shapes) and then households (squares) are sampled (red squares) so that they represent the target population of campaign nets.



**Prospective sub-population, cross-sectional (or longitudinal)**

When nets are tagged before distribution (Figure 3) a sampling frame cannot immediately be established as it remains unclear where the tagged nets channeled into the distribution end up. The first step, therefore is to ensure that **all tagged nets are distributed** and do not end up somewhere else as otherwise the exact number from which a sample is drawn cannot be established. This can be achieved by introducing them directly at the various distribution points and not at the warehouse level. It usually means that the area of distribution is limited to a few points and cannot easily cover larger areas such as entire districts, provinces or regions. In a second step the tagged nets need to be identified in the field as this will establish the sampling frame of all households with tagged and distributed nets. Once this sampling frame is established a representative sample can be drawn from this list using either random numbers or systematic sampling. In addition to the limitation of covering larger areas with the “tagged” net approach, there are two critical aspects with respect to representativeness of the resulting sample:

1. Is the sample of distributed “tagged” nets identical with the overall net distribution? Households attending distribution points at different times (e.g. early risers or late comers, first day or later etc.) may represent very different sections of the target population with different behaviors and unless the distribution of tagged nets mirrors that of all campaign nets, there is potential for bias. This can be achieved in two ways: either all nets at distribution points in a defined area are tagged (in which case one could proceed to sample as described for the prospective, longitudinal design above only that the sampling frame is limited to areas with tagged net distribution), or a systematic approach is used to distribute a sub-sample of tagged nets in which the estimated total of nets to be distributed at a specific point are used as a basis for systematic sampling and every kth household is given tagged nets.
2. Is the sample of tagged nets that was identified in the field as the sampling frame representative of all tagged nets distributed? If not all tagged nets can be identified this could introduce another bias depending on whether or not the households not included in the sampling frame are different in relevant ways from those that were included.

Neither of these two potential biases can be captured in the collected data of the sample and, therefore, cannot be excluded or quantified.

Figure : Sampling in prospective sub-population design: A) tagged nets (blue circles) are distributed ensuring that all reach the population of interest; B) after distribution all households with tagged nets (yellow squares) need to be identified and form the sampling frame from which a representative sample of households (read squares) and nets (red circles) are included for monitoring.



### Single versus all nets sampled at household level

In the past, it has been common practice in some durability studies to only sample one net from each household in order to ensure that household characteristics such as socio-economic status are present in the sample as they are in the population. However, as shown in Figure 4, other factors such as sleeping place or users are specific to each net and others (washing and handling) are determined by both household and net factors. In addition, all these factors may vary over time and nets may change sleeping place or user within the house.

Depending on which net in the household is selected and the household’s characteristics the selection of just one net can introduce a bias (see Figure 5) that can be avoided if all nets are included. Generally a truly random sampling within households with multiple nets should compensate for this variation, but the collected data in a “one net only” approach would not allow excluding any bias or capturing its magnitude. In contrast, multiple measures within the same household (more than one net) can be controlled for in a statistical analysis of all net using e.g. a generalized estimation equations model (GEE) and using the household as random effect. Furthermore, as the interest in LLIN monitoring – as mentioned earlier – is to obtain the most accurate picture of all distributed nets, multiple exposure within the same households is a reality and, therefore, including all nets in the household is the better solution.

Figure : Factor determining the durability outcome of nets. Solid circle= campaign net; open circle= non-campaign net



Figure : Potential bias of selecting only one net per household. Solid black circle= campaign net not sampled; solid red circle= sampled campaign net in serviceable condition; red patterned circle= sampled campaign net in torn condition; open circle= non-campaign net



## Conclusion and Recommendations

1. Based on the considerations presented above it is clear that a prospective, longitudinal (cohort) design with a two stage cluster sampling presents the best option for an accurate representation of the population of campaign nets to be monitored as it includes the least potential biases. This is therefore the generally recommended approach for the monitoring of LLIN durability. There are only two exception to this recommendation:
   1. The campaign nets in sampled households cannot be easily identified by brand label, size, color, and reported age of net because nets with the same specifications have been distributed through other channels previously. In this case a “tagged net” prospective sub-population design can be used provided the monitoring starts before the actual campaign distribution, the tagged nets are truly randomly distributed through the campaign, and at least 80% of the tagged nets can be included in the sampling frame.
   2. The time since distribution is too long (more than six months) to allow capture of nets from the campaign as a cohort before significant loss due to “wear and tear” sets in. In this case a retrospective, cross-sectional design can be used, provided campaign nets can be easily identified.

Figure 6 shows a decision tree to determine which design should be used.

1. In all any designs all campaign (or tagged) nets in a household should be included in the monitoring as this best reflects the population of interest and repeat observations in the same household can be controlled for (if this presents an actual confounder) using statistical approaches. In addition, all other, non-campaign nets in the sampled households should be captured in the surveys at least with respect to source, age and users (not necessarily physical integrity).

Figure : Decision tree to assist in determination of best monitoring design



## Sample size

*What should be the criteria to calculate sample size considering funds available and objectives of study, how much precision do you need? How to decide how many clusters and how many households to use? What are the implications of that?*

While the sampling strategy (discussed above) is critical to achieve a high level of accuracy or representativeness, the sample size is critical for the precision of the results, i.e. essentially the width of the resulting confidence intervals for estimates.

Sample size is usually a compromise between desired precision and available budget. On the one side, we have the sample size needed to achieve a pre-defined precision sufficient to make programmatic decisions, and on the other side, we are constrained by the available budget. The factors that need to be included into the statistical calculation of sample size are the following:

* The minimum difference to be detected (statistically) in order to still make the study worthwhile and provide actionable results. In our case it would be the question: “what difference in estimated LLIN durability or survival do I need to be able to detect between two different LLIN products at the same place or between the same LLIN product at two different locations to say that they are programmatically different?” As a general orientation it is thought to be about 10-12 percentage-points difference in “% survival in serviceable condition” after three years or a 0.5 year difference in median LLIN survival.
* The level of confidence that is to be used (alpha error); this is usually the 95% level.
* The anticipated power to find the desired difference (beta-error); usually at least 80%.
* The anticipated design effect for cluster sampling designs. This is the ratio between the within-cluster and between-cluster variation of the key outcomes. For a design with 15 clusters, we assume a design effect of 2.5.
* The non-response rate and loss to follow-up (in our case households from our cohort that move to another area during the three year monitoring phase), usually assumed to be 5%.
* The average number of nets received from campaign per household, which depends on the mean household size and the allocation strategy of the campaign, e.g. 2 nets per household, 1 net for every two persons rounding up (or rounding down), etc.
* The anticipated attrition rates for a) all-cause attrition and b) attrition due to “wear and tear” (resulting in discarding and/or re-purposing of nets). A rough orientation from existing data is about 35-40% all-cause attrition over three years and 20% attrition due to “wear and tear”, leaving 15% of nets being given away
* The expected median survival of the LLIN over three years. Here it is best to assume 50% (equivalent to a three-year survival) if no previous data exist, as this gives statistically the most conservative estimate of sample size.

The file 2h. DM Sample Size Calculator provides a simple tool for calculating both the margin of error or the sample size based on the above parameters.

While the exact figures will always vary a bit based on the values used for each of the factors, a general orientation is that per site, a cohort of 250-275 campaign **nets** is needed at the start of the monitoring to obtain the ±10-12 percentage-points difference after three years mentioned above. In the typical case of a household size of about 5 people and a standard “universal coverage” campaign this will require 150 **households** per site at the start (depending on assumed loss-to-follow-up rate).

Given that in our case we will need a cluster-sampling design (see sampling strategy above) to obtain representative data of the area of interest (the domain), the next step in sample size determination is how the needed number of households is split up between the clusters. The following factors need to be considered:

* The number of households per cluster should be small enough to ensure that they can be identified in a single community, i.e. it will depend on the size of the most remote communities, which in some countries can be as low as 20-30 households. On the other hand, the number of households per cluster should be as large as possible to be most cost-effective in survey implementation.
* The number of clusters should not be too low, as a very small number of clusters will always result in a higher design effect. While it is generally true that the higher the number of clusters the lower the design effect (at the same number of households per cluster), modelling has shown that gains in design effect are small above 30 clusters and tend to get large below 20. This has lead to the general use of 30 clusters as “optimal” in the EPI 30x7 cluster survey design for immunizations.

Since one can generally expect after a mass campaign that coverage of villages is reasonably homogeneous, meaning that design effects are not extreme, it is plausible to use the minimal number of clusters, 15, as a reference for durability monitoring in order to minimize cost. This would lead to a recommended 15x10 design (15 clusters and 10 households per cluster) to achieve 150 household and ~250 campaign net cohort per site at the start of the monitoring under the assumptions outlined above.

# Part 2: Implementation

## Organization of field work

*How many interviewers/supervisors, optimal team size considering the tasks such as hole assessment in the field, critical aspects of organizing training*

In determining the size of the field team the following factors need to be considered:

* The team should be large enough so that a single cluster can ideally be done in one working day.
* The team should be small enough so that it can be transported in a single vehicle.
* The team members attending to each household should be sufficient to allow a high-quality of the hole assessment of the nets.

From past experience a team of three interviewers plus one supervisor for each or every two teams (depending on proximity of communities) has shown to be optimal in many cases. The three interviewers can move separately during the initial phase of preparing the household lists for sampling and then move together to each of the sampled households. For the hole assessment of nets two interviewers hold the net on its four corners while the third makes the assessment.

A more detailed description of the field procedures and roles and responsibilities will be given in the field manual as part of the Durability Monitoring package.

## Budgeting

Please refer to document 1.d “Sample Budget” for a detailed description of budgeting considerations for all stages of fieldwork and analysis, including personnel, fringe, travel, equipment and supplies, fieldwork, training, insecticide testing, and other laboratory testing.

## Data collection tools

*What needs to go into the questionnaire, what is essential and what is “nice to know”?*

The data collection tool (questionnaire) should always be as much standardized as possible in order to allow consistency and comparability across countries, but also needs to allow adjustment of responses (e.g. for household assets) to the local culture and conditions.

The following sections or modules should be part of the standard questionnaire:

1. Identification
   1. Location, cluster and household number, result of interview, date and consent
2. Household member roster (to be updated at each monitoring round)
   1. Age, gender and relation to head or each person present in the house
   2. Identification of whether a person is a “usual” household member and whether a person was present in the house the night before
3. Household characteristics
   1. House characteristics (roof, floor, windows) and assets owned. This section also includes a number of questions that refer to potential threats for nets such as cooking in the sleeping room, storage of food or crops in sleeping rooms and observed presence of rodents in the last six months
   2. Ownership of mosquito nets and their source
   3. Exposure to BCC messages around net use, care and repair. This includes a number of Likert score questions (level of agreement with certain statements) that are used to establish summary scores of household attitudes, knowledge and perceptions around nets
4. Net care and repair behaviors
   1. Recall of damage mechanisms in currently owned nets
   2. Reasons for repairing/not repairing
   3. Type of repairs done
5. Nets from the campaign that have left the household since the distribution (this section is only needed for the initial establishment of the cohort)
   1. Number of nets ‘lost’, fate of lost nets (what happened to them) and reasons for their loss
6. Assessment of all campaign nets currently owned by the household (campaign net roster)
   1. During the first survey (establishment of cohort) this includes identifying the campaign net based on brand/label and age of net and marking it with a unique identifier (ID number label or bar-code)
   2. During follow-up surveys this includes establishment of presence of each cohort net and if absent, fate and reasons for loss
   3. Location of net (hanging or not, type of sleeping place etc.)
   4. Use of net (and list of net users from household member roster (line number)
   5. Washing practices
   6. Result of the hole assessment
7. Assessment of any other nets (not from campaign) in household
   1. Includes information on type, source and age of net, use and washing patterns. It does **not** include a hole assessment of these nets

A questionnaire for the monitoring activity is provided as part of the Durability Monitoring package (2b Survey Questionnaire).

## Data capture, management and safety

*Digital versus paper… how to decide, what needs to be considered for capacity, IT back-up, training etc.; aspects of databases, data safety and how to address these vis-à-vis ethical review boards*

Many surveys are now done using digital devices for data capture and storage like PDAs, smartphones or tablets, and transferring data directly into a web-based database using telephone networks. While this approach has some clear advantages such as incorporating data quality checks at point of entry and increased speed of raw data availability, some aspects need careful consideration if such a system is not yet well established with the survey implementers:

* The type of device used. Consider the operating system, battery life, robustness in the environment (heat, dust etc.), access to technical backstopping in case of failure, and options for charging batteries in the field (solar etc.)
* The software to be used for the programming of the questionnaire on the device. There are many free “questionnaire tools” available for phones and tablets but not all of them are suitable. In the case of Durability Monitoring the software needs to be able to establish relational databases (the household, household member, cohort LLIN, other nets) and establish cross checks between them (e.g. pull up all cohort nets in the household for assessment, provide a list of all members when identifying net users assuring that each user can only be entered once etc.)
* Programming of the devices needs considerable time and must start at least 2-3 months before the survey with multiple test runs to make sure all checks and skips are correct and any bugs are fixed
* At least 3-4 additional days need to be added for the training and practicing of the survey team
* Great care must be taken to create daily back-ups of data in case of interruptions of network connections and other potential failures
* Cost of digital data collection can be higher compared to paper if it has to be built from scratch and no previous experience exists.

It must be remembered that a well working paper-based system (with good field QA and double entry with validation) may be better than a poorly working digital system.

Regardless of the type of data collection system used (electronic or paper), data safety and confidentiality is a key consideration. At no time should data be revealed outside the study team. There are many good standard procedures for ensuring data safety in the field and once collected, and these are detailed in the study protocol. Ethical review boards will require thorough description of how information will be kept private throughout the study. Briefly, these measures include:

During fieldwork:

* During the interview itself, conducting the interview away from other household members so that answers are not overheard and the respondent is not influenced by presence of others
* Fieldworkers are trained not to share information amongst themselves or with other study participants or family members
* Immediately filing checked questionnaires in appropriate envelopes and packaging to keep out moisture and dust (if using paper)
* Supervisor responsible for ensuring that completed questionnaires are securely transported (paper) or that PDAs/tablets are securely managed throughout data collection
* Electronic backups are made often (at least daily), with password protection

# Frequently Asked Questions

1. *What happens during the baseline survey if the fieldworkers find that nets are still in their packaging? Are they still tagged?*
   1. Yes, all campaign nets in the household are still tagged, even if they have not yet been unpacked or used. The study team should gently encourage the household to use the nets, but it is ok for households to ‘save’ their nets while they are still using their previous nets. This will be taken into account during the analysis phase. Holes should still be counted in unused nets – this is usually zero, of course, but depending on how long the net has been stored, it is possible that some holes have developed, so each net should still be examined.
2. *What happens if households have fewer campaign nets than expected – can this affect our sample size calculations? What do we do?*
   1. There are a number of reasons why households may have fewer campaign nets than expected:
      1. Campaign was not implemented well, and households were missed,
      2. Fewer nets were given to households because of a shortage of nets in that area
      3. Households may have given away or sold their campaign nets
   2. The sample size calculator tool (file 2h) will show what the detectable difference will be.
   3. When do I need to add more households, and how do I do that?
3. *Are we supposed to replace the LLINs sampled for bioassays with the same ones (same brand, same batch, same lot no., etc) or any LLIN will do?* 
   1. No, it’s not required to replace the sampled nets for bioassays with identical, original campaign nets, although it may be easiest in many circumstances to request nets from the campaign for this purpose. Ideally, the replacement nets will be the same shape and the same material, for minimal disturbance for the household.
4. *If a household is randomly sampled for the 12M bioassay, and randomly sampled again for the 24M bioassay, do we replace the household?*
   1. First, this is likely to be a very rare event.
   2. This will depend on the number of campaign nets in the household. If the household received only one campaign net (or has kept only campaign net) at the 12M bioassay, then they cannot be resampled again for the 24M bioassay, and should be replaced, if selected.
   3. If the household received two campaign nets, and one is randomly selected for bioassay at 12M, the household receives a replacement net. If the household is again randomly selected at the 24M bioassay, as long as the household can confidently identify a remaining original campaign net, it can be sampled for bioassay. If for example the replacement net is identical in shape, color, material, etc, and the household is unable to confidently distinguish it from the original campaign net, then the household should be replaced. The goal is to avoid using a replacement net for bioassay, since the replacement nets have not been used for as long as the original campaign nets.