

GUIDELINES FOR SAMPLING ORPHANS AND OTHER VULNERABLE CHILDREN

To Estimate the Size and Characteristics of OVC Populations

Anthony G. Turner, Sampling Consultant, May 2003

Introduction

National estimates of the population of orphans and other vulnerable children (OVC) in developing countries are imperfect. The methodology for those estimates often relies upon demographic models (Bicego et al 2003) rather than censuses or surveys. Where the latter have been used, usually in conjunction with household surveys – such as DHS (Demographic and Health Survey) and MICS (Multiple Indicator Cluster Survey) – conducted for broader purposes, the results re orphans have been inconsistent and widely disparate, while there are no known national efforts to measure the population size of other vulnerable children. One source of and reason for poor estimation is that there has not been a focused effort to develop specific guidelines on how to survey OVC – and, in particular, how to sample them. This is partly due to the fact that they still comprise a tiny portion of the populations of most countries, notwithstanding the rapid growth occurring from the HIV/AIDS epidemic; and for that reason there has not been, to date, a strong push to study OVC in a concerted effort.

What the manual includes

This manual is intended to help correct the gap in methodological approaches for surveying/sampling orphans and other vulnerable children. It provides detailed and explicit guidelines on sampling approaches (a) to estimate the size of the OVC population, or its prevalence and (b) to study the characteristics of OVC, particularly as a target population of special interest in the campaign against HIV/AIDS. The latter entails gathering detailed information on OVC (demographic, socio-economic, and – for HIV/AIDS – health status, behavior, knowledge and/or attitudes).

The manual is oriented toward developing countries but the methods proposed are equally applicable in developed nations.

The manual also emphasizes sampling methods for surveys intended to study the *entire population* of OVC, which may be considered as consisting of three sub-populations: (1) OVC in households, (2) OVC residing in institutions and other group quarters and (3) homeless OVC including, especially, street children.

As each of the three sub-populations requires essentially independent sample surveys, it is possible to use the guidelines for each one as a separate endeavor in any country that wishes to survey only one of the sub-groups in a specialized study. In that context the presentation distinguishes between *linked* surveys, when all three components are conducted in tandem, and *stand-alone* surveys restricted to one of the components.

Scientific, probability sampling is taken as a given in these guidelines, because without it there is no statistical justification for drawing inferences about the population that the sample is supposed to represent. Probability sampling is especially crucial for national-level surveys intended to study the total population of OVC or their characteristics. It is not only the theoretical basis for drawing scientific conclusions but the only technique that actually provides a justifiable way of measuring the margin of error in the results.

What the manual does not include

Partly because of the need to emphasize formal, probability sampling, informal methods such as convenience, judgmental or snow-ball sampling are not included in the guidelines. But such techniques are also excluded because their application is generally for small-scale monitoring and evaluation of intervention projects in local areas. The manual does not provide guidelines for limited, special-purpose studies, which, while important, are outside the scope of the manual.

Those techniques, along with rapid assessment methods such as focus group discussions and exit interviews (from health facilities) are useful as indicative, “early-warning” tools for rapid identification of both problems and solutions in localized project research and design. Their use, however, is not intended to provide valid or reliable statistical estimates of the various sub-populations they are drawn from. Likewise, the use of sentinel surveillance sites, which is a valuable technique for the longitudinal monitoring and evaluation of individuals and/or communities but not for statistical estimation, is not covered by these guidelines. Researchers and others, therefore, who have different objectives from prevalence estimation and may be interested, instead, in conducting more limited, focused, and localized studies about OVC are encouraged to continue utilizing the various rapid assessment and surveillance methods that are available.

Capture-recapture methodology might be considered a promising method by which to conduct a survey of street children. It is a technique that entails sampling the target subjects on two occasions (“capture” and “recapture”), matching the two samples and then estimating the total size of the population by applying the Chandrasekar-Deming formula, which is based on the match rate (Chandrasekar 1949). It was decided not to promote or include it in these guidelines, however, for several reasons. The theoretical underpinnings of the technique require that the population being studied be closed, in the sense that there must be virtually no subjects either entering or exiting the population during the study period. This key condition would likely be drastically violated in the case of street children. Each survey occasion must be a totally independent operation, a condition almost impossible to attain when dealing with human populations. Moreover, practical implementation of capture-recapture methodology is rarely successful because of the very real difficulties in accurately matching sample persons between the capture and recapture phases. Poor matching essentially invalidates the method. Thus, it is felt that capture-recapture could not provide a robust estimate since three of the most important conditions for its applicability would be seriously violated.

Finally, the manual does not provide sampling guidelines for trend measurements or estimates of change. Again, those topics are most relevant for project evaluation purposes, whereas this document is confined to one-time, cross-sectional surveys – usually on a large scale – to examine the prevalence and characteristics of OVC.

Definitions of orphan, vulnerable child

The internationally accepted definition¹ of an orphan is as follows:

AN ORPHAN IS A CHILD UNDER 18 WHO HAS LOST ONE OR BOTH PARENTS.

At the recent conference in Gaborone (UNAIDS/UNICEF 2003), the working definition – not yet formally adopted - for OVC (including orphan) is as follows:

Child below the age of 18 who has lost one or both parents or lives in a household with an adult death (age 18-59 years) in past 12 months or is living outside of family care (homeless-street children and children in institutions).

The formal definitions are important because they have implications on the overall survey purpose and data collection strategy including the sampling methodology. When the survey objective is to estimate prevalence (the size of the OVC population) the entire age group under 18 is obviously targeted. On the other hand, a survey focused on HIV/AIDS may, in some country applications, be confined to 10-17 year-old children, especially when sexual issues are emphasized. Or, the primary age group in another survey might be OVC under the age of 10 when health status and socio-economic impact of HIV/AIDS make up the key content of the questionnaire. Regarding sampling, the target age group is an important issue because sample sizes for household surveys are larger, in terms of the number of households that must be screened to locate OVC, for children under 10 years old or 10-17 year-olds compared to all OVC under 18.²

In this manual it is assumed that all OVC under 17 are included in the survey target group. An implication of this assumption, for the survey design as opposed to the sample plan, is that different questionnaire modules would seem to be in order – one focusing on health status and impact of HIV/AIDS for OVC under 10 and the other emphasizing sexual behavior, knowledge and attitudes for OVC 10-17.

¹ See “National AIDS Programmes, a Guide to Monitoring and Evaluation,” UNAIDS/USAID, Geneva, June 2000, p.130, which uses under 15 as the target age group; the planned revision of this document in 2004 will show a change from under 15 to under 18. Individual countries vary somewhat in defining orphanhood with some using death of the mother only as the criterion.

² If n_h is the sample size – number of households – to achieve a certain level of precision for a survey of OVC age 0-17, the sample size for 10-17 would be approximately $2.25n_h$ and for 0-9 approximately $1.8n_h$.

For use in household surveys, the standard survey questions that have been developed and widely used to ascertain orphan status inquire about whether each parent of the child is still alive and living in the household. This would seem to exclude formally adopted children who have lost one or both parents, but it is not clear how such children are treated in specific survey applications.

Survey coverage of the target population

According to the Monitoring and Evaluation Guide, page 138, the indicator for survey measurement, as distinguished from the definition, is “Percent of children under 15 (now revised to under 18) in a household survey whose mother, father or both parents have died.” It is important to recognize that this indicator is not an estimate of the percent of orphans in the population. The target population – orphans and/or OVC – is found in a variety of residential settings. These include traditional households, communal group care quarters, orphanages and other institutions. In addition there are homeless youth living on the street and elsewhere in non-fixed places of residence.

While it is thought that the great majority of orphans can be captured (in the parlance of survey-taking) in households (as opposed to other vulnerable children which are more likely to be living in group arrangements or on the street), a household survey alone is a necessary but insufficient condition to obtain unbiased estimates of either the prevalence of OVC or their characteristics. Estimates of prevalence would be *under-stated* through a household survey alone and characteristics would be *biased* to the degree that those living in households are fundamentally different from those in institutions or on the street.

Consequently, a survey intended to study the total population of OVC must include those living outside traditional households including those that are homeless. Each of these three sub-populations must be *covered* by the survey process in order to achieve unbiased estimates of the number, distribution and characteristics of OVC.

Another aspect of survey coverage is geographic. This manual emphasizes national level coverage in the survey application. Many countries, however, will be interested in more limited geographical study of OVC such as those living in a particular region, province or selected cities. *Note that the guidelines in this manual apply not only to national surveys of OVC but also to large, geographical sub-populations such as regions, designated provinces or cities.* It is important to bear in mind the geographical coverage area when considering the sample size, since sample sizes discussed in the guidelines apply irrespectively of whether coverage is at the national level or at the level of a sub-national domain such as province.

Sample size

The sample size is a key parameter of sample design for OVC, whether estimating the size (prevalence) of the population or its distribution and characteristics. The size of the

sample differs depending on whether linked or stand-alone surveys are planned. When the purpose of the survey is to look at the total population of OVC, then the 3 sub-components (household residents, institutional OVC and homeless) must be combined in *linked* surveys. A study may also be designed for the purpose of examining only one of the components as a stand-alone survey.

The sample size to estimate prevalence is somewhat smaller than that needed to study characteristics, because the latter requires detailed analysis of sub-categories of OVC (gender, age groups, substantive groupings, etc.). The factors that affect sample size are many and include precision requirements, confidence level needed, magnitudes of the characteristics being estimated, sample design effect and others. Appendix A shows the estimation details for calculating the sample sizes, along with the choices for the various parameters that make up the mathematical formula.

We will look at sample size along two dimensions. One is in terms of sample size needed for prevalence estimation versus characteristics. The other is sample allocation among households, institutions and homeless. Chart 1 summarizes the various possibilities, which are further elaborated in the remainder of this section.

Households – prevalence estimation

It is useful, first, to ascertain sample size requirements for the household component of OVC coverage, since a large majority of OVC are orphans that reside in households of surviving parents or other relatives. We speculate that the percentage is 75 and that is the basis for the sample size figures that appear in Table 1 below, although a more precise sample size can be calculated (see Appendix A) for any country that has more accurate data on the percentage of OVC in households. There are no known estimates of OVC at the national level but “Children on the Brink” (see exact reference in Appendix D) does provide estimates of orphans and those are used to help guide sample size calculations.

Chart 1. Sampling Coverage for Various Types of OVC Surveys and Recommended Minimum Sample Sizes

| Target Population | Type of Estimate | Linked or Stand-alone Survey | Survey Type(s) | Sample Size ³ |
|-------------------------------------|-------------------------------------|------------------------------|------------------------------|---|
| All OVC 17 or younger | Prevalence (size of OVC population) | Linked | Household | 1000-2100 households ⁴ |
| | | | Group quarters | [Census] |
| | | | Homeless sites | 400-600 homeless persons (to yield 100-150 OVC) |
| All OVC 17 or younger | Characteristics | Linked | Household | 1500-3100 households ⁵ |
| | | | Group quarters | 100 OVC |
| | | | Homeless sites | 400 homeless persons (to yield 100 OVC) |
| OVC in households | Characteristics | Stand-alone | Household | 1500-3100 households |
| OVC in group quarters, institutions | Characteristics | Stand-alone | Group quarters, institutions | 400 OVC |
| Homeless OVC | Characteristics | Stand-alone | Homeless sites | 1600 homeless persons (to yield 400 OVC) |

Table note: In the last 2 rows the figures of 400 and 1600 are regarded as bare minimums; where budgets permit it is recommended that 800-1000 OVC be targeted.

A key factor in the construction of Table 1 is that orphans (and OVC by extension) comprise a small percentage of the total population of most developing countries. Even in countries most affected by AIDS, such as in sub-Saharan Africa, the percentage is only about 4.4. Consequently, household sample surveys designed to estimate, with moderate

³ The sample size is doubled when *separate, equally reliable* data are wanted for boys, girls, in which case the latter are regarded as two *estimation domains*.

⁴ See Table 1.

⁵ See Table 2.

reliability, the number or proportion of OVC require sample sizes of about 1000-2100 households. The smallest sample size needed is in the sub-Saharan African countries, where the average is about 1000 households. In the average Latin American country and in China, which have, proportionately, much smaller OVC populations about twice as many households are needed for sampling (see Table 1).

Table 1. Estimated Percentage, p , of OVC Living in Households to Total Population and Approximate Sample Size (Number of Households) Necessary to Measure p ⁶

| Area (1) | Estimated percentage, p , of OVC 0-17 living in households to total population (2) | Sample Size (Households) (3) | Estimated number of OVC in sample (4) |
|--------------------------|---|---------------------------------|--|
| Sub-Saharan Africa | 4.37 | 1027 | 269 |
| Asia, except China | 3.65 | 1241 | 271 |
| China | 2.19 | 2100 | 276 |
| Latin America, Caribbean | 2.35 | 1952 | 275 |

The last column of the table is computed as (column 2)/100 x column 3 x 6.0, where 6.0 is the expected household size on average.

In Table 1 the estimated percentages of OVC were derived from “Children on the Brink” (UNAIDS/UNICEF/USAID 2002) figures, using various assumptions (refer to Appendix A for details). They are not intended to be accurate (and certainly will vary by country) but merely serve as a starting point from which to calculate the necessary sample size.⁷ The sample size requirements for household surveys discussed above are in the context of national surveys, but the sample size figures in Table 1 apply for sub-national areas (regions, selected provinces or cities) as well, provided the estimated percentage of OVC in the designated sub-national area is roughly those of column 2. The reason it is applicable to sub-national areas is because the reliability of a sample estimate depends on the *size of the sample*, not the size of the population. Nor does it depend on the proportion of the population that the sample represents, unless the proportion is large – say, over 5 percent (not generally the case for OVC).

As noted in a footnote, the sample sizes in Table 1 apply to total OVC. When separate estimates are wanted with equal reliability for boys and girls, the figures of the last two columns would be doubled.

⁶ This is the estimate, p , of the proportion of OVC in the population measured with precision of $\pm 0.15p$ and at the 95 percent level of confidence. See Appendix A for the calculation formula and assumptions made.

⁷ In the formula for calculating sample size, it is necessary to use an *estimate*, however approximate, of the percentage, p , of the total population one is trying to ascertain more precisely from the sample survey.

OVC in institutions and other group quarters – prevalence estimation

To account for the institutional component of OVC prevalence it is recommended that a *census* of institutions be conducted, unless this sub-population is very large. It is thought that few developing countries have large numbers of institutions that house OVC. A recent, comprehensive study in Rwanda (Fernwell 2002), for example, showed there to be a total of only 24 government-registered institutions (“unaccompanied children centres”) that house orphans (81 percent) and other vulnerable children (19 percent). Once the institutions have been identified, however, it is simple and inexpensive, relative to a sample survey, to enumerate all the OVC living in them. There may of course be practical difficulties with resistance by center directors suspicious of the purpose of the census. Steps are necessary to gain working access to centers, such as obtaining active support by the ministry with responsibility for monitoring institutions for children and/or the organization which funds each institution’s operations. More details on these points are provided in sections on sample frame development and in step-by-step sampling procedures later in this manual.

Thus, sampling and, by inference, sample size play no role in countries. As such, the estimated number of OVC obtained from the household survey component, discussed above, would simply be added to the count of OVC obtained in an institution census (plus the homeless component discussed below) to obtain the estimated size of the OVC population.

It is of course crucial – whether for a census or a sample survey - that all of the institutions housing OVC be identified, to the extent possible, and this is discussed later in the section on sampling frames.

Street children and other homeless OVC – prevalence estimation

As previously noted, in the absence of better information we have assumed that 75 percent of OVC live in households.⁸ By that assumption, the remaining 25 percent live in group quarters or are homeless. As we would plan to conduct a census for OVC in institutions it remains to establish a plausible sample size for estimating the homeless.

The methodology for sampling and surveying homeless OVC is complicated, with the main issues being how to locate them and to conduct interviews. Fortunately, since the number of homeless OVC is comparatively small, it is only necessary to sample about 100-150 in order to round out the estimate of total prevalence. In that regard we might speculate that a properly designed and executed survey of homeless people would yield, say, 25-35 percent that are OVC under 18 years old (the others would be adults). Thus, we would attempt to design the OVC component of a survey on prevalence by sampling 400 to 600 homeless people and administering a screening interview to ascertain age and perhaps parental status (see Chart 1.)

⁸ A good estimate of this percentage is not available and is only speculated here. Also, whatever the average, the true figure is likely to vary considerably country-to-country, which strongly suggests pilot-testing in a few locations to obtain a better figure before designing and launching a full survey.

It must be recognized that the ratio of total homeless persons to homeless OVC can vary drastically from country to country or city to city. We have assumed throughout this document that the ratio is on the order of 4 to 1, an assumption which governs the sample size calculations and the illustrations that are given with respect to surveying street children. In some places there may be more homeless youth than adults. In those cases the ratio is actually less than 1 and the number of homeless that would have to be screened in order to identify the OVC is reduced dramatically. Clearly, in the application of the techniques described in this manual in a particular country, it is necessary to estimate, as carefully as possible in advance of the survey, the proportion of homeless persons that are likely to be OVC in order to calculate the appropriate sample size.

Estimates of OVC characteristics – households, institutions, homeless

In contrast to a prevalence survey where only reliable estimates of the *number* of OVC and their proportion of the total population are wanted, here the aim is to survey OVC as the focal population to learn about their demographic and socio-economic characteristics. When the objective of the study is related to HIV/AIDS, health status, at-risk behavior, knowledge and attitudes are studied as well.

It is generally accepted that the minimum number of survey subjects needed for fairly reliable detailed analysis – that is, where cross-tabulations are needed by gender, age groupings and substantive categories - is about 400. Obviously, reliability improves the larger that number becomes, and it would be better to sample 600, 800 or 1000 OVC when resources permit.

Table 2 shows the required number of households to be sampled to obtain various numbers of OVC under 18 years old.

Table 2. Approximate Number of Households to Sample in Order to Locate OVC Age 0-17 for Detailed Study of Characteristics⁹

| Area | 400 OVC | 600 OVC | 800 OVC | 1000 OVC |
|--------------------------|---------|---------|---------|----------|
| Sub-Saharan Africa | 1525 | 2275 | 3050 | 3800 |
| Asia, except China | 1825 | 2750 | 3650 | 4575 |
| China | 3050 | 4575 | 6100 | 7625 |
| Latin America, Caribbean | 2850 | 4250 | 5675 | 7100 |

The figures are rounded to the nearest 25; see Appendix A for calculation formulas.

The figures in Table 2, as in Table 1, apply to the 75 percent of the OVC population that would be expected to live in households. Note that the sample sizes shown would

⁹ For countries that might design an OVC survey confined to a different age group, the approximate number of households needed for OVC age 0-9 or 10-17 are, respectively, 1.8 times and 2.25 times the numbers shown in Table 2.

also be applicable in cases where the survey team was studying OVC in households *as a stand-alone target population*, that is, without intending to combine the results with other data from institutions or the homeless.

When linked data are wanted, that is, combined results from household, institutional and homeless surveys, the sample size for OVC in institutions and homeless need be only about 100 from each source. That implies selecting 100 OVC from institutions but about 400 homeless persons, the latter of which would have to be screened to ascertain their age group and, possibly, parental status. Note that the comparatively small sample sizes for the institutional and homeless components pertains because the objective is to obtain enough cases to round out the estimates for *total* OVC (not just those in households). If the study objectives included making reliable *comparisons* among the three sub-components and analyzing their distributions in various ways, then we would be dealing with “characteristics” (as opposed to prevalence) in the same context as if each component was a stand-alone survey. Reference to Chart 1 shows how the sample sizes would be larger accordingly.

For surveys intended to examine the institutional or homeless population as a stand-alone target in its own right, the sample size should be 400 in each case – that is, 400 OVC from institutions or 1600 homeless persons to be screened for OVC status.

Note that from Chart 1 a census is cited as the methodology to use for prevalence estimation for children in institutions (unless the number of institutions is very large), while samples are implied for characteristics. Depending on circumstances in a given country, this may imply either (a) sampling both the institutions and their inhabitants in two stages or (b) sampling the inhabitants in *all* of the institutions. This point is elaborated further in the section, *Sampling design and approaches*.

With respect to boy-girl domain estimates, the numbers in Table 2 and those discussed in the previous paragraphs of this subsection would be doubled for separate, equally reliable data.

Sample frame construction

As implied previously, tapping OVC for survey-taking requires that a typology of OVC be articulated. The typology is in terms of the locations where OVC can be found for purposes of conducting survey interviews. We have identified those venues in the previous sections as traditional households, group living quarters including institutions and various places where homeless children sleep (or congregate).

A separate and independent sample frame is required for each of these three venues. For the household component, development of a sample frame is not necessary, as will be seen below. For institutions, frame construction is straight-forward in concept but not necessarily in implementation. For homeless children, development of a feasible sample frame is complicated both in concept and implementation. In addition, a list frame of

known OVC might also be usefully employed in countries where lists of good quality exist. The list frame, however, would be used in conjunction with the other three, because all its members would be expected to be included, by definition, in at least one of the three.

The sample plan for an OVC survey intended to study *all* OVC would thus be based on a *multi-frame* design. In cases where a research effort is focused on only one of the component parts (for example, OVC living on the street), a single frame approach suffices. Each of the different frames is discussed below.

Households

Consistent with our conjecture that, on average, 75 percent of a country's OVC are housed in traditional households, the *principal* frame for sampling that component is obviously a household frame. The features of household sampling frames are well known. Moreover, it is recommended that when household surveys are used for OVC measurement they be done as so-called *riders* to existing household surveys rather than mounting an independent survey from scratch. The cost of a separate survey dedicated exclusively to OVC is likely to be beyond the means of most countries. Furthermore, when a rider module is used the household frame and other parameters of the sample design are pre-determined by the designers of the host survey to which the rider is attached. For these reasons, we will not discuss the general parameters of sample frame construction with respect to households.

It must be pointed out that there are significant methodological considerations, which go beyond the sampling issues, on the substance of what might be added to a broader household survey. For example, guidance is needed on how to avoid pitfalls that could yield inaccurate information such as misrepresentation of orphans as non-orphans and vice versa. Asking the wrong questions or the right questions in the wrong way generates distorted results. Practical advice informed by field experience gathering this type of data is needed of how to avoid the significant risks of over and under-counting orphans. These broader methodological issues are beyond the scope of this manual.

There are, however, two sampling-related aspects specific to the measurement of OVC that surveys which use household frames are likely to encounter.¹⁰ They are important for survey coverage and steps must be taken to avoid bias from under-estimation.

The first concerns child-headed households, which are often ruled out-of-scope in traditional household surveys. It is perhaps apparent that this survey practice would have to be avoided in a survey intended to estimate OVC prevalence or characteristics. Thus, when using an OVC module as a rider, it is important for the survey team to review the procedures for the host survey to ensure that child-headed households are not excluded.

¹⁰ These are not so much problems with the household frame as such or even with sampling but rather with the *survey* methods generally used in implementing household surveys in developing countries.

The second aspect concerns the coverage of children in communal care at the village level, a cultural practice often found in African villages (and perhaps elsewhere) and, in particular, for children orphaned by AIDS. Household surveys are not designed to cover such children, as the survey instruments seek to identify and include only *members* of households who *usually* eat or sleep there. Any child under general, jurisdictional care of the village as a whole and not attached to a *specific* household would be missed in a traditional household survey. *This type of under-coverage may in fact account for a significant part of the under-estimation of orphans that has plagued some surveys.*

There is a fairly simple solution to correct for the under-coverage of orphans in particular and perhaps other vulnerable children as well, described in the previous paragraph, in household-based surveys in general, and especially, when the household survey is used as a rider for OVC measurement and follow-up. Most household surveys in developing countries consist of a stage of selection in the rural areas involving villages or parts of villages. Villages are often defined as the first or second stage sampling units. For OVC measurement, it would be recommended that the chief or head of each sample village be contacted and asked to supply specific survey information about the OVC under communal care by the village but not members of particular households. An example of such communal arrangements might be “drop-in” centers for OVC of the type known to exist in Botswana.

Each OVC identified, or a sub-sample of them, would be eligible for the survey interview – in person or by adult proxy, depending upon age. The survey weight that is appropriate to produce the OVC estimates for that portion of the sample picked up by “communal” interviews is the inverse of the probability of selecting the village (generally the weight associated with the first-stage unit of selection), modified, if necessary, by a weight for sub-sampling whenever a sample of OVC so identified is interviewed.

It is important to note that this type of survey inquiry at the village level has most likely not been done. It is recommended therefore that a community level questionnaire to administer to the village head be developed and pre-tested.

Orphan lists

This sub-section applies mainly in cases where a country (or province, city, etc.) would plan to conduct a *dedicated* OVC survey, without benefit of the rider approach. It can also be applied in rider surveys, but to a lesser extent.

List sampling depends on whether a suitable list or set of lists of known OVC (most likely orphans only) exists in a given country or jurisdictions within countries. Where lists do exist, they should be explored for their potential use as a list frame to supplement the household frame. If it is thought, for example, that a significant portion of known orphans is listed on an available set of rosters, sampling from those lists would reduce the size of the sample needed from the household component. In countries where ministries of social welfare or human resources maintain administrative rosters of orphans, those rosters should be reviewed for use as a sampling frame. The review would ascertain

whether they are sufficiently current, accurate and whether names and locatable addresses of the registrants are available. Such rosters need not be complete to be useful, because a household-based frame would also have to be used to cover OVC not listed on official registers, especially non-orphaned, other vulnerable children.

A survey complication in surveying OVC sampled from two different, overlapping frames (list frame and household frame) is that un-duplication procedures have to be implemented to avoid over-estimation by double counting. The methodology for un-duplication requires a time-consuming and error-prone matching operation¹¹ to find out which interviewees selected from the household frame also appear on the list frame. Such cases must either be eliminated or “down-weighted.” What is sometimes not completely understood is that this is a requirement irrespective of whether the case in question happened to be selected on the list frame. The mathematical probability and concomitant weight depend on the *chance* of being selected, not whether actually selected.

Another important disadvantage of using a list frame of known OVC, whether alone or in combination with a household frame, is that the children selected are likely to be randomly scattered about in the population rather than clustered (except when the list includes OVC living in institutions). Interviewing costs on a per case basis are much higher for OVC selected from the list as opposed to those selected in the household frame, the latter of which is clustered.

Although list sampling may have some advantages, especially if a very high percentage of OVC is listed on official registers, the disadvantages of cumbersome un-duplication requirements and higher per case costs lead us *not* to recommend list sampling in general. Note that list sampling is not strictly necessary anyway, since the household and institutional frames will cover all the registered OVC with known addresses. This document does not include list sampling as an option in the stepwise procedures given further on. It is recommended, however, that the sample design specialist in any country contemplating an OVC survey explore the feasibility for list sampling in terms of costs, complexity and applicability.

Institutions, other group quarters

In this section we speak of the *sampling frame* of institutions and other group living quarters, although in general it is recommended that a census of such places be conducted for OVC prevalence estimation (in which case no sampling is involved). The requirements for compiling a list of non-household residential establishments for OVC are the same, however, irrespective of whether a census or a sample is undertaken.

Of the three venues where children may be found, those living in institutions that house OVC should be the easiest to identify. Lists of institutions that are run by the government, religious or non-governmental organizations must be compiled. This,

¹¹ The chief difficulty in matching operations of this type is in comparing names and deciding when a match occurs.

however, should not be especially difficult, as such institutions are likely to be well-known and compiling a list of them easily accomplished. In order for the list to be acceptable as a frame it should be, as much as possible, complete, accurate and current. The contact persons for identifying institutions include officials at ministries of social welfare or human resources, religious leaders, NGOs that focus on youth and any organization that works on HIV/AIDS issues. Such officials should be sought out not only at the national level but also at the levels of province, district and large city, in order to ensure complete, or maximum, coverage.

Orphanages and temporary foster care facilities obviously belong on the list, but also any other institutions that house children for one reason or another. Juvenile detention centers, jails and even adult correctional facilities that also have young inmates should be included. By the same token, military barracks/quarters ought to be included, even though the majority of their inhabitants would likely be older than 17.

An important consideration is that most institutions would likely be separated by gender. In that case, sampling would be done independently in each set, thus resulting in a separate set of sample institutions for boys and for girls.

Homeless shelters require special treatment, even though they may be regarded as “group living quarters.” It is recommended, however, that homeless shelters not be included on the institutional frame because of the transient nature of their occupants. Instead, shelters would be part of the sample frame for dealing with homeless children, discussed in the next sub-section.

Information to be obtained from each residential establishment for constructing the list includes only its name, location and the number of resident OVC. The latter, which is the so-called “measure of size,” is needed for three reasons. First, if the overall number of institutions is small enough that a census of OVC is to be conducted, the measure of size informs the survey team about the workload for interviewing in each institution. Second, where the number of institutions is so large that a sample of them must be selected, the measure of size is used to administer the sampling procedure – systematic probability-proportionate-to-size (*pps*) selection process. Third, it is used to establish the sample probabilities and weights.

It is important to be aware that in those institutions that house other persons as well as OVC the measure of size should include only the OVC sub-population. It is also important to note that the OVC count need not be perfect in order to serve its purpose as a measure of size for *sampling* (though it should be a fairly close approximation). The exception is that the measure of size should be as accurate as possible when a census of institutions is to be conducted.

Street children and other homeless OVC

It is perhaps ironic that the component of OVC populations that is most likely the smallest, at least in many countries, is the most difficult to survey. The principal

challenge for sampling OVC that live on the street is in locating them. Meeting that challenge is largely a matter of frame development. First, it should be pointed out that any organization planning a survey effort intended to cover street children will have to carefully define, operationally, what is meant by a street child. In some contexts, for example, it might include children who spend most of their time on the street as truants but nevertheless sleep in the home of their parents each night. The task of defining street children for survey-taking is beyond the scope of this manual, which is focused on the sampling methods, and belongs instead to the survey manager working in collaboration with government authorities and professionals who are developing its overall concepts, objectives, design. The procedures for sampling including development of the frame, however, are the same no matter what definition of street children is used.

Examples of “floating” populations (no fixed residence) include orphaned or abandoned children, runaways, persons involved in prostitution and/or drugs. The latter two groups include persons of all ages, though a significant proportion would be OVC and other young people.

To locate OVC for survey-taking requires going to the sites where they congregate, visit frequently or sleep. The number of persons to be found at a particular site varies from time to time and this variation must be taken into account in sample design. These floating populations of youth must be sampled using the concept of *time-location* sites, a method of covering mobile youth populations that is unbiased and adheres to the tenets of probability sampling. Each time-location site identified is defined as a separate primary sampling unit (*PSU*), all of which taken together then comprise the sample frame. *PSUs* are thus defined and created as frequented sites in combination with time intervals. Accordingly, the same *physical* site is usually included in the sampling frame more than once, depending on different times of the day, or week, that sampling is to take place.

On the grounds that everyone sleeps somewhere, it is recommended that the *sampling of street children be confined to places where youth are known to sleep* such as railway stations and other transportation terminals, under bridges, abandoned buildings, vacant lots and any others that community leaders know about and can pinpoint.

Construction of the frame of *PSUs*, using the time-location definition, must begin with community experts and key informants who are knowledgeable about the behavior of homeless or street youth. These experts are asked to identify (so-called *ethnographic mapping*) all the known locations where the youth may sleep. For the sites identified, calendar segments of discrete time intervals are then associated with each one and the site-time intervals thereby become separate *PSUs*.

Use of the sleep-site concept seems to be the most feasible for establishing a closed, non-overlapping frame of time-location *PSUs*. Such an approach would diminish considerably the estimation problems that would arise from duplicate coverage – the statistical chance of including the same street child from both her venue for sleeping and a site where she performs various waking activities. To apply the approach accurately, however, requires that the sleep sites be identified with near 100-percent perfection. And

it is recommended that all time intervals be used in the *PSU* construction, since sleeping would take place during daylight hours as well as night.

Further details on *PSU* construction for homeless children are provided in the section on sampling approaches and in the step-by-step procedures of Appendix C.

Sampling design and approaches

Households

For the household component the most promising methodology in terms of cost effectiveness is the so-called *double sample* approach – also referred to as post-stratification. An initial large sample¹² of households is used to conduct simple screening interviews to identify and locate households with OVC. The screening question would be of the type mentioned in the UNAIDS Monitoring and Evaluation Guide, that is, whether “the household is currently caring for any children under the age of [18] whose father, mother or both parents have died.”

In practice, the screener should be applied through use of the rider approach, that is, by appending it to another household survey that has a sufficiently large sample. Timing of course is crucial and the rider approach must be discussed and approved for the host survey during the early planning stages of the latter. In countries that intend to conduct a DHS in a time frame commensurate with planning for an OVC survey, the former - with its rigorous sampling methods and typical sample size of about 6000 households - would be a suitable vehicle in which to use the rider for initial screening to identify households with OVC.¹³

Note that because the host survey will have been designed by others, we will not discuss such matters as geographic stratification, cluster sizes, stages of selection for the OVC module. It is incumbent upon the sampling technician, however, to ensure that the treatment of those matters in the host sample design meets the needs for the OVC survey. He/she must also ensure the sample size for the host survey is big enough to comply with the requirements for OVC measurement as depicted in Tables 1 and 2. If the host survey has a sample size that is too big, it is reasonable to select only a sub-sample for the OVC module, which may be done in one of two ways.

To illustrate, suppose 3000 households are needed from the host survey in order to find 400 OVC (see Table 2), but the host survey uses a sample of 6000. In that

¹² The initial screening may also be done in a decennial census provided the timing is right and agreement can be reached to add to the census a minimum set of questions that identify orphans, although identification of other vulnerable children may require more questions than a census can easily accommodate.

¹³ The sample design for a host survey is generally pre-determined but the sampling practitioner in charge of the OVC survey should ensure herself that the former is suitable with respect to such points as the use of probability methods, geographical coverage, stratification criteria, number and size of clusters and, of course, sample size.

case, 800 OVC are likely to be identified if the rider is appended to every host survey questionnaire. The survey team may decide either (1) to apply the screening rider to only 1 in every 2 of the host survey households in order to keep to the original 400 cases or (2) to apply it in all 6000 of the host survey households. In option 2, the decision could then be taken later whether to conduct the second-phase, detailed questionnaire on OVC in all 800 households (budget permitting), or sub-sample those households by half and thus revert to the originally intended 400.

To estimate OVC prevalence it is a simple matter of merely tabulating the results from the screening interview.

To study OVC characteristics, however, a second-phase survey would most likely be needed, because it could be too burdensome to have a detailed questionnaire re OVC appended to the DHS (or other household) survey that is used for the rider.

Thus, for surveying OVC characteristics a second survey (second phase of sampling) is done, in which follow-up interviews in all¹⁴ the households identified as housing OVC or other vulnerable children are conducted for the focused study. The detailed questionnaire on OVC is only administered at the second phase sample.¹⁵

Group quarters and institutions - prevalence

First, we discuss the sample/census plan to estimate the prevalence of OVC living in institutions. As mentioned already, it may not be necessary to use sampling at all. In many developing countries where orphans are cared for by a surviving parent or other relatives, there are few formal institutions for housing them or other vulnerable children. In such countries it is recommended to conduct a census of the comparatively few residential institutions that do exist. This scenario is very likely to apply for any survey confined to geographical sub-areas of a country such as particular provinces or cities, if not for a national survey.

To control costs, a sample of institutions would, however, be necessary when there are large numbers – say, 150 or more - of institutions and other residential facilities that house OVC. A sample of about 40 to 50 institutions is recommended. Note that 40 is the recommended minimum, but this number would apply no matter whether there are 150 institutions in the universe or 1000.

To estimate prevalence, a *census* of the OVC in the *selected* institutions would be conducted, that is, a simple count of the residents under 18 with collection of only minimal detailed data other than age. Note, for this reason there is only a single stage of

¹⁴ It is of course possible to choose a sub-sample of the households, as described in the illustration, especially if substantially more OVC were identified than expected in the first-phase, host survey.

¹⁵ The detailed OVC module may, instead, be administered during the initial (and only) visit to the household whenever an OVC is found. This would likely be the preferred approach if all such OVC were to be interviewed without any further sub-sampling at a second phase.

sample selection – the institutions themselves and they would thus be selected on the basis of size strata rather than by probability proportionate to size (*pps*). More complicated stratification criteria such as cost per child per year are not recommended for the reason that information necessary to create such strata are not likely to be feasibly obtainable.

To create size strata means it is necessary to obtain figures on size (number of OVC) from institution officials. Obtaining such figures is somewhat less problematic in facilities where only OVC are housed, since officials may have current counts or rosters available. In other facilities such as detention centers and military barracks both OVC and adults would be residents. In those cases, administrative records on age may not be readily available and it is necessary to work with institutional officials to obtain a count of the OVC. Such a count, for use as the measure of size in each institution, can be approximate.

Approximate numbers of OVC would then suffice for assigning an institution to its stratum (and the design would not be ruined if a few institutions are assigned to the wrong stratum). The exact definition of what constitutes small, medium or large would no doubt vary by country, but a rule-of-thumb might be as follows:

Chart 2. Size Strata for Institutions, Other Group Quarters

| Stratum | Size (Number of OVC) | Illustrative Sample Selection ¹⁶ |
|---------|----------------------|---|
| Small | Under 20 | 1 in 10 <i>PSUs</i> |
| Medium | 20-99 | 1 in 5 <i>PSUs</i> |
| Large | 100 or More | Select all |

In creating the strata the institutions should be arranged in a geographic fashion within each stratum by urban-rural and by province prior to sample selection, in order to ensure geographic spread of the ensuing sample. Selection within each stratum would be systematic according the selection rates illustrated in Chart 2.

The second-stage census counts would be inflated by the weight associated with the sample institution, depending upon its stratum. In the above example, the weights would be 10, 5 and 1, respectively, for small, medium and large institutions. All OVC in the sample institutions from the small stratum would have a weight of 10, for example.

In countries where boys and girls are housed in separate institutions, the sample plan should call for each to be treated separately and independently. This is especially important when equally reliable data are wanted for boys and for girls, in which case the sample size is the same for both.

¹⁶ Sampling fractions, in practice, depend on the total number of *PSUs* wanted for the sample and their average sizes.

Group quarters and institutions – OVC characteristics

Second, we consider sampling plans for estimating characteristics. Recall from Chart 1, we need a minimum of 100 cases when this is a linked survey with a household and homeless component; 400 when a stand-alone survey. There are two cases to consider for selecting the appropriate sample of OVC. One is the situation where the institutions in the universe are so few that a census of them is taken. In that case, to illustrate, suppose there are 18 institutions in which the census was carried out, and a total of 900 OVC are housed in them. A complete roster of the OVC from each institution must be compiled. Then, the sample of, say, 400 OVC should be selected systematically from the combined list of all 900. The sampling interval is equal to $900/400$ (in this illustration), or 2.25; a random starting number is chosen, using a table of random numbers between 0.01 and 2.25. See the step-by-step procedures in Appendix C for an illustration of systematic sampling from a list.

The second case occurs when the sample of OVC (again, say, 400) is taken from a sample of institutions. Selection in this case would be based on a two-stage sample design. The first stage should be a sample of institutions selected systematically using probability proportionate to the size of the institution. Again, as above, this means we must obtain figures on size (approximate number of OVC) from institution officials.

A *fixed* number of OVC from each sample institution would be selected to participate in the interview for the OVC survey. Again, the number of institutions to select should be, at minimum, 40. Suppose in this case we intend to select 400 OVC; with a minimum of 40 institutions, the cluster size would be 10, that is, 10 OVC would be chosen in each sample institution, using systematic selection from a complete roster of OVC in that institution.

As mentioned previously, independent sampling of boys and girls in separate strata will likely be necessary in countries where they are housed in institutions separated by gender.

Procedures for implementation of *pps* sampling of institutions and systematic sampling of residents are given in Appendix C in the stepwise methodology.

Homeless children

In many countries, the number of homeless children may be considered so few that it would not be feasible to survey them because of the technical complexity of doing so. We suggest sample sizes of 1600 homeless persons in order to locate 400 street OVC, for example; clearly, this does not make sense if the total number of such persons is only a few hundred. In other countries, however, where there are large numbers of street children in big cities, their exclusion would produce (1) significant under-estimation with respect to OVC prevalence and (2) serious bias with respect to their characteristics, especially for surveys focused on HIV/AIDS.

Related to the last point is the question of the appropriate survey coverage area for homeless youth. As suggested, this sub-population of youth is likely to be heavily concentrated in large cities. The comparatively small numbers of them to be found in rural areas, towns or small cities probably does not justify conducting national or regional surveys for homeless youth.¹⁷

Finally, locating street youth for surveys will likely require that some interviewing take place in the middle of the night, which can be intrusive and dangerous. It is recommended that pilot tests be carried out on a small number of cases to ascertain the problems associated with such interviewing conditions, as well as overall feasibility, before a full-scale survey is mounted.

When the decision is taken to survey homeless youth, a two-stage sampling plan should be implemented. The first stage is a selection of the time-location *PSUs* that were described in the section on sampling frames. The second stage of sampling consists of posting interviewers at the site for the time interval designated and interviewing all (take-all approach) the youth that are present or arrive at the site during that period. A key feature of the design is ensuring that sampling takes place over a fixed time interval which is the same for every sample *PSU*.

Because of the unusual definition of the *PSUs* and the way they are constructed, it is not practical to sample a fixed number of street OVC or other homeless at each selected site. The number of homeless is apt to vary considerably from one day to the next at a given site. And even if an approximate measure of size can be assigned to a site it is likely to be unstable in certain ways: the majority of homeless persons may not be within the survey age group, may not be OVC, etc. As a result, the ultimate number of street children in the sample cannot be controlled precisely, although that number can be targeted and, hopefully, closely approximated.

As recommended previously, the *PSUs* should be defined in terms of sites where street children sleep, excluding other sites where they may congregate for their waking activities. As such, each sleep location should be divided along the time dimension in 4 to 6 hour segments for *PSU* construction. For example, four *PSUs* might be formed as follows:

- Under City Bridge – 6 a.m. to noon
- Under City Bridge – noon to 6 p.m.
- Under City Bridge – 6 p.m. to midnight
- Under City Bridge – Midnight to 6 a.m.

Similar *PSUs* would be constructed for all other sleep locations identified by key informants. See Form S1 in Appendix C on step-by-step procedures for the questions to

¹⁷ An exception is countries where homeless youth are known to live in mining sites or quarries which are typically located in rural areas; a case can be made for conducting an OVC homeless component in such countries (or provinces affected).

use in inquiring of officials and other key informants about places where street children sleep.

In all *PSUs*, so constructed, a measure of size must be determined. This is an approximate count of the number of homeless persons expected to be at a particular location during the time interval specified. It is a tedious but necessary step in order to be able to impose as much control as possible over the ultimate sample size for the survey. To establish the measure of size requires pre-survey field work involving a visit to each site and counting, roughly, the number of homeless people that come and go during the specified time interval. As implied, the count need not be exact and it will be unstable anyway, but still it is necessary to know whether a *PSU* should be assigned a measure of, say, 5 or 50.

PSUs would be stratified according to size categories based on their measures of size – small, medium and large, similar to the institutional strata described above. Again, the exact size categories would be defined at the country level, but a rule-of-thumb is given in Chart 3.

Chart 3. Size Strata for Time-Location *PSUs* Where Homeless Persons Sleep

| Stratum | Size (Number of Homeless Persons) | Illustrative Sample Selection |
|--------------------|-----------------------------------|-------------------------------|
| Small <i>PSUs</i> | Under 5 | 1 in 10 <i>PSUs</i> |
| Medium <i>PSUs</i> | 5-25 | 1 in 5 <i>PSUs</i> |
| Large <i>PSUs</i> | More than 25 | 1 in 2 <i>PSUs</i> |

Other parameters of sampling to be considered are sample size, number of *PSUs*, cluster size. For stand-alone surveys targeted solely on homeless youth, sample sizes should be a minimum of 1600 homeless persons, on the grounds that only about 400 of them would be OVC in the age group (Chart 1). When the survey is intended to estimate homeless youth as part of a larger study to estimate the overall size of the OVC population, the sample size necessary is 400 homeless persons (again, 100 expected to be OVC under 18).

Cluster size and number of *PSUs* are intertwined and involve the number of homeless youth expected, on average, to show up at a given site during the time interval. The cluster size has two dimensions to consider. One is the number of homeless people at the *PSU* site and the other is the number of those that are OVC. In any case, a short set of screening questions is necessary when conducting the interview in order to eliminate any homeless person who is out-of-scope for the survey. There is the issue of whether to screen for age by questioning each person in the sample *PSU* or, instead, relying upon observation by the interviewer. The first is more expensive and reliable, while the second is the reverse. A compromise procedure would be to forego questioning adults who are clearly older than 17 but administer a simple question about age for those whose appearance is more ambiguous.

Administration of the screener should last only two or three minutes. The main interview questionnaire intended for the (OVC) target group should determine the cluster size. And this is done in terms of how many OVC interviews can be conducted at a given site in the allotted time. That number will vary of course, dependent on such factors as interview length and number of interviewers to be posted at the site. See illustration for calculating average *PSU* size in the step-by-step procedures of Appendix C.

As already mentioned the interviewers should be posted for a fixed interval of time. That interval must be the same as the interval used to define the *PSUs*, which is recommended at 4 to 6 hours. At each selected *PSU* an interview is sought from *all* OVC present or arriving during that time. In cases where there are expected to be a large number of OVC present, the interview team should be expanded accordingly. For example, if two interviewers can interview 15 OVC in 6 hours and 30 OVC might be expected at a given site, the interview staff should be doubled for that site.

It is reasonable to question why not impose a maximum number of interviews. Estimation becomes problematic, in terms of accuracy, when there is a cut-off. By contrast, an unbiased estimate can be made (for both prevalence and characteristics) when all interviewers, in each *PSU*, are posted at the site for a fixed interval of time and interview all in-scope persons who show up. Sub-sampling the OVC at a site is also problematic with its implications for listing them first and then applying systematic selection procedures on the spot. Moreover, during a given interval, new arrivals would be expected throughout so that their total number, which is needed to calculate the sampling interval, cannot be known until the time interval expires – a paradox.

Some youth would be expected to sleep at the same site or different sites within the duration of the survey. Questions must be added to the survey instrument to identify such cases in order to “down-weight” the results for accurate estimation.

To illustrate: It is quite likely the same site will fall into sample more than once, since the site will comprise 4 *PSUs* based on time of day if 6-hour time intervals are used or 6 if 4-hour intervals are used. Any youth accustomed to sleeping in the same place has a good chance of showing up at the location at different time intervals, thus giving him/her multiple chances of selection. When that occurs, the youth should be interviewed only once. A somewhat more difficult duplication problem to sort out concerns those youth that sleep in *different* sites from one day to the next. To overcome this problem we need to take into account the length of the survey period – one week, two, etc. Let us assume 2 weeks for the example. It is necessary to ask each respondent how many other places he/she usually sleeps during a 2-week period. He would also be asked to identify those locations. Those locations would be matched – back in the office – to compare against the full list of sites that make up the *PSU* frame. For each match, the survey weight for that respondent must be down-weighted by the factor, $1/t$, where t is the number of matching sites.

If the survey of homeless youth is being done in combination with another that covers OVC in households or institutional facilities, duplication of coverage can also occur. In that case, it is necessary to ask each youth interviewed at the sample site whether he/she has a usual place of residence – again in order to down-weight the results.

Documentation and evaluation

The sampling technician should take necessary steps to carefully document not only the sample plan for the OVC survey, whether linked or stand-alone surveys, but also its implementation. Sample designs often require adaptation at some stage of the field work, because of unforeseen situations that arise in the conduct of the survey. It is important to record - step-by-step - all the procedures used in carrying out the sample plan to make sure the implementation is faithful to the design. When it is not, it is even more important to document all the departures from the design, even minor ones. This information is necessary later at the analysis stage, in case any adjustments need to be made; but it is also indispensable for planning future surveys.

To evaluate the results of the survey, sampling errors should be estimated. The sampling error, or *standard* error, is the measure that allows the confidence interval to be constructed around the estimate, so that users can evaluate how reliable the data are. Standard errors are estimated from the survey data themselves, whenever the design adheres to probability sampling methodology, by using fairly complex mathematical formulas and procedures. The procedures used must reflect the actual design employed in the survey. The latter is often a difficult task to perform and is frequently done only with the assistance of a sampling expert.

There are computer software packages that may be used to calculate the standard errors, in lieu of designing a dedicated variance¹⁸ estimation program. Information about many of the packages can be found on the Internet including “Wesvar” from Westat Corporation and (Westat.com) “VPLX” from the U.S. Bureau of the Census (Census.gov). Those packages, like most others, are suited especially to household surveys. They would have to be adapted, again with the help of an expert, for application to institutional or homeless surveys; otherwise, dedicated computer programs may be necessary for those components.

¹⁸ The term “variance” refers to the square of the standard error.

Appendix A. Parameters for Figures of Tables 1 and 2 of Text

This appendix shows the estimation formulas used to derive the figures pertaining to household survey sample sizes appearing in Tables 1 and 2 of the text, including the parameters used and assumptions made. For OVC in institutions or homeless the sample size is not calculated per se, but rather is presented as the minimum necessary to produce fairly reliable cross-tabulations for analysis in the case of measurement for characteristics or the minimum necessary to round out a household survey when total prevalence is the objective.

Estimating the parameter, p

In order to calculate sample sizes, it is first necessary to have an approximate estimate of the parameter, p , (defined below) that one is actually trying to estimate from the survey. The estimation formula for p in Table 1 is

$$p = (0.54) p' (0.75), \text{ where} \quad [1]$$

p is the parameter to be calculated and is the estimated proportion of OVC 0-17 living in households to total population;

p' is the estimated proportion of children 0-17 who are *orphans*, taken from the estimates for each continent provided in “Children on the Brink” (UNAIDS/UNICEF/USAID 2002); those estimates are for orphans 0-14 years old and it is assumed here that the estimated proportion for 0-17 is the same;

0.54 is the average estimated proportion of children 0-17 in the total population, where the assumption is made that .03 is the proportion for each single year of age in developing countries; thus $0.54 = 18 \times .03$; and

0.75 is the *assumed* proportion of OVC *that are living in* households; in this context it is thought that most of the OVC in households are orphans and that the remaining 25 percent will be both orphans and other vulnerable children living in group quarters or homeless.

Estimating sample sizes

The estimation formula for the sample size, n_h , in Table 1 is

$$n_h = 3.84 (p) (1-p) (f) (1.1) / 6 (e^2), \text{ where} \quad [2]$$

n_h is the parameter to be calculated and is the sample size in terms of households;

3.84 is the square of the z -statistic (1.96) which defines the 95 percent level of confidence;

p is from equation [1];

f is the sample design effect, assumed to be 1.5 – a value which implies moderate clustering because although OVC may concentrate in certain neighborhoods or villages their overall proportion is comparatively small;

1.1 is a multiplier to account for an expected rate of non-response of 10 percent;

e is the margin of error to be attained, set at a moderate level of precision, which is 15 percent of p ; thus $e = 0.15p$; and

6 is the average household size in developing countries.

The estimation formula for the number of households necessary to locate varying numbers of OVC age 0-17, found in Table 2 is

$$\eta = \alpha / 6p, \text{ where} \quad [3]$$

η is the number of households to be calculated;

α is the desired number of OVC to be sampled and interviewed and it varies from 400 to 1000; and

p is from equation [1].

When the assumptions do not apply

For countries in which one or more of the assumptions discussed above do not hold, simple substitutions may easily be made in the formulas to arrive at more accurate figures on sample size. For example, the average household size may be larger or smaller than 6.0; non-response may be expected around 5 percent instead of 10; and the value of p' for a particular country can be more precisely computed than by using the continental average which is what these formulas above assume.

Each country would be expected to have exact figures from its census on the percentage of the population that is under 18, so that 0.54 may be substituted for. Some countries may have more precise information on the proportion of OVC that reside in households, in which case the value of 0.75 in formula [1] would change. It is recommended however that no change be made for the z -statistic value of 1.96, which is the international standard. The design effect, f , should also be left at 1.5 as the default value. It is also recommended that e be defined as $0.15p$ except in cases where budgets would permit a more stringent margin of error, in which case it may be reduced to $0.12p$ or $.1p$. Such reductions in the margin of error will yield dramatically higher sample sizes however.

Appendix B. Calculating Survey Weights

Analysis of the survey data conducted according to the procedures described in this document will generally require weighting – inflation of the raw data by the inverse of the probabilities of selection for the sample cases. Weighting is necessary because, with few exceptions, the sample plans discussed are not self-weighting.

Household Component

Use of the rider approach, in which OVC survey instruments are attached to an existing host survey, means that the weights for the OVC identified and surveyed are the same as those for the host survey. Hence, it is only necessary to obtain the weights associated with the host survey and apply them. Those weights will be based on the design of the host survey, which will no doubt differ in every country that applies the OVC sampling guidelines. Frequently, however, the host survey design may be a self-weighting one in which each sample household and/or person has the same weight. Even in that instance, however, it is necessary to use that weight when combining the OVC results with surveys from institutions and homeless children.

There are two special situations that need attention for weighting in the household component – situation in which the host survey weights cannot be used as is. First, the guidelines mention that sub-sampling of OVC identified from the host survey may be used whenever the number is much larger than expected. In that case the host survey weight is multiplied by k , where k is the inverse of the sampling fraction used to sub-sample for the OVC interviews. For example, if only 2 in 3 of the host survey households containing OVC are interviewed for the OVC survey, then the appropriate weight would be $w \times 1.5$, where w is the host survey weight and 1.5 is the inverse of the sub-sample fraction, 2/3. Note that w may be variable, depending on whether the host survey is self-weighting; if it is not self-weighting but varies say, by strata or urban-rural, then there would be a different value of w for each such stratum.

The second situation concerns the case of village sampling where heads of the villages selected in the (usually) first stage of the host survey are asked about OVC living in communal arrangements. OVC identified in this case would receive the weight associated with selection of the village. Again, it is necessary to obtain this weight from the host survey, but in many household surveys it is equal to:

$$w_i = M/(a m_i), \text{ where}$$

w_i is the weight of the i^{th} sample village determined from the host survey,

M is the total number of households in the sampling frame for the host survey and is equal to the sum of the m_i values,

a is the number of first-stage units (villages) selected in the host survey,

m_i is the measure of size (number of households) in the i^{th} sample village.

If the OVC identified from the communal source are sub-sampled before interviewing, the host survey weight, w_i , is multiplied by k (described above).

Institutional Component

When a census of institutions is taken to obtain the count of OVC for prevalence estimation, no sampling is involved nor is weighting.

If the institutions are systematically sampled for OVC prevalence estimates, each OVC in a given sample institution receives a weight of I_s , which is the sample interval for the s^{th} institutional stratum (small, medium or large). For example, selection of institutions in the small stratum at 1 in 15 would result in a weight of 15 for each OVC in that institution (recall that for prevalence, *all* the OVC in the sample institutions would be included in the survey interview).

When males and females are selected either in separate strata or as separate domains, the weight must be calculated separately for each.

For OVC characteristics, two cases occur. First, when institutions are few and a census of them is taken, the sample weight is calculated on the fraction of OVC selected (recall this is a constant fraction over all institutions combined). The weight is simply the sampling interval, I ; there will be two such intervals, however, when boys and girls are sampled separately.

In the second case, involving a large number of institutions, a two-stage sample is done in which institutions are first selected systematically with *pps* and a fixed number of OVC is sampled systematically within each sample institution. The weight is then equal to:

$$w_j = N/(b n_j)(n'_j / c), \text{ where}$$

w_j is the weight of the j^{th} sample OVC,

N is the total number of OVC in the sampling frame and is equal to the sum of the n_i values, that is, sum of the institution measures of size,

b is the number of institutions selected in the survey,

n_j is the estimated measure of size in the j^{th} sample institution,

n'_j is the actual count of OVC in the j^{th} sample institution,

c is the fixed number of OVC selected in each sample institution.

Again, the above would be repeated for boys and girls separately when they are sampled from two independent sets of institutions. For a given stratum (that is, male institutions or female institutions) the weight is approximately the same for all OVC in the sample, to the extent that the estimated measure of size is the same as the count of OVC in a given institution. Hence, the sample is approximately but not precisely self-weighting. Occasionally, the total number of OVC residing in a given institution may be less than the value, c , in which case all the OVC at that institution are interviewed. Calculating the weight for that institution then requires substitution of the actual value of c rather than the pre-determined, fixed value – another reason why the sample will not be precisely self-weighting.

Homeless Component

The basic weight for the homeless component is very simple because of the way the time-location *PSUs* are constructed and sampled (even though that process is laborious). There are three weights, one for each of the three strata (small, medium, large). Each is given by I_s , or simply the sampling interval for each stratum. Each OVC in a given *PSU* receives the same weight for that *PSU*.

Some sample OVC will need to have their weights modified by a factor to correct for multiple sleeping sites they will have used during the survey period. In that case, the final weight is equal to I_s multiplied by $1/t$, where t is the number of sites he/she may have used.

Combining results for estimation

Weighting will almost surely be required when the three components are combined in a linked survey for estimates of total OVC or their characteristics. This is because each component entails an independent sampling frame and design, and the probabilities of selection will necessarily differ among the three components. The weights are calculated separately, in accordance with the description above, for each component and simply applied to the data files to produce the combined estimates.

Appendix C. Stepwise Procedures for Sample Selection in OVC Surveys

This appendix provides step-by-step procedures for implementing the sampling methodologies described and proposed in the text. Less attention is devoted to household sampling than to the other components because it is recommended that the household component of OVC surveys be done through the method of appending a rider questionnaire module to an existing household survey, the design of which is more or less pre-determined.

The reader is referred to the main text to find the context and the conditions under which the various methodologies given here, including the appropriate use of linked surveys of the three components versus stand-alone surveys, should be applied.

Household Component

Because the rider approach is the recommended strategy to be used for the household component, the steps involve sampling-related activities in the context of the host survey.

STEP 1 – Review sample design of host survey for adherence to measurement objectives of OVC survey.

STEP 2 – Examine design to ensure that probability methods are used at all stages of selection.

STEP 3 – Check that sample size is large enough to obtain required number of households (see Tables 1 and 2); note, especially, the larger sample sizes that are needed for boy-girl estimates as separate, equally reliable *domains*.

STEP 4 – Ensure that cluster sizes are tolerably small such that sample design effect for OVC estimates is about 2.0 or less.

STEP 5 – Examine design for completeness of coverage and be prepared to describe in survey documentation any important groups that may not be covered by survey (for example, nomadic households, boat people, areas excluded for security reasons, etc.).

STEP 6 – Examine design to determine if any sub-populations, such as urban areas, are over-sampled and be prepared to adjust sampling accordingly for OVC rider.

STEP 7 – Work with host survey team to ensure that child-headed households are not excluded from coverage.

STEP 8 – Work with host survey team to identify all first stage sampling units (host survey *PSUs* - villages, urban neighborhoods, etc.):

- (a) Administer special questionnaire to head of each sample village or neighborhood to find out whether orphans or other children outside family

care may reside in communal living arrangements rather than in particular households.

- (b) Identify all such children under 18 years old and tally them (with *PSU* weight) for prevalence estimation; interview them (or sub-sample) for detailed characteristics.

STEP 9 – Administer screener questionnaire as rider to host survey, “Is there any child under 18 living in this household whose mother is dead?” [Repeat for “father”]. *Note that the screener will identify orphans but not other vulnerable children who may be living in households.*

STEP 10 – For all households identified as having at least one orphan, record on the screener questionnaire appropriate identification facts (*PSU* number, province, district and other administrative codes, address/location of household, name of household head), in order to be able to find the household again for the OVC follow-up survey.

STEP 11 – Tally the screener questionnaires with their appropriate weighting factors from the host survey (see Appendix B) to estimate prevalence or size of OVC population.

STEP 12 – For studies of OVC characteristics, conduct follow-up interviews using OVC module:

- (a) With *all* households where orphans are found through rider questionnaire
- (b) Or, systematic sub-sample of those households if their number is substantially larger than expected.

STEP 13 – If sub-sampling is used in step 12, record sub-sampling fraction so that proper weighting can be implemented at data analysis phase of operations.

STEP 14 – Maintain detailed documentation of the sampling operations and procedures that are used, noting especially any implementation features that depart from the design.

Institutional Component

STEP 1 – Develop sampling frame by contacting government officials, NGOs, religious leaders and others who are knowledgeable about institutions and other group living quarters where OVC reside.

STEP 2 – Make appropriate contacts at national level such as ministries of health, social welfare or others that deal with youth; national headquarters of NGOs, national headquarters of religious groups.

STEP 3 – Make similar contacts also at province level and at district level.

STEP 4 – Through contacts above, compile comprehensive list of all institutions and group living quarters that house OVC.

Following is a suggested questionnaire, Form GQ1, to use when contacting government officials and other experts for this purpose.

Illustrative Form 1. Illustrative Data Collection Form to Compile Group Quarters

| | | |
|---|---|--|
| <p>Form GQ1 – Compilation of List of Institutions and other Group Living Quarters Known to House Orphans, other Vulnerable Children</p> <p>[Introduce yourself and explain the purpose of the survey].</p> | <p><i>Administer this form to officials of the ministries of health, social services and others working with youth; to religious leaders and to NGOs that work with youth. Complete a separate form for each contact. Continue on additional forms as necessary.</i></p> <p><i>This form requires obtaining an estimate of the number of OVC under the age of (____) in each residential facility. The estimate may be an approximation if exact figures are not readily available from the respondent.</i></p> | |
| <p>We must compile a list of all the orphanages, institutions and other group living arrangements where orphans and other children outside traditional family care are known to reside so that we may conduct the survey.</p> | | |
| <p>1. Are there any government-run orphanages in the (country, city, province, district)?</p> | <p>List them with name, address/location.</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> | <p>Number of OVC</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> |
| <p>2. Are there juvenile centers for youth in trouble with the law?</p> | <p>List them with name, address/location.</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> | <p>Number of OVC</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> |
| <p>3. Are there institutions for adults, such as jails, that may house youth including OVC as well?</p> | <p>List them with name, address/location.</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> | <p>Number of OVC</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> |
| <p>4. Are there informal group homes that provide foster care for OVC that you know about?</p> | <p>List them with name, address/location.</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> | <p>Number of OVC</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> |
| <p>5. Are there any orphanages that are run by religious organizations?</p> | <p>List them with name, address/location.</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> | <p>Number of OVC</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> |
| <p>6. Are there military barracks or camps where youth including orphans may live?</p> | <p>List them with name, address/location.</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> | <p>Number of OVC</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> |
| <p>7. Are there any other types of institutions for housing young people including OVC that you know about?</p> | <p>List them with name, address/location.</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> | <p>Number of OVC</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> |
| <p>8. [Ask especially in cities, rural communities] Are there “drop-in” centers where orphaned children may live?</p> | <p>List them with name, address/location.</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> | <p>Number of OVC</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> |

Form GQ1 is illustrative and would be adapted as necessary to fit a particular country’s conditions. The age group of OVC is not specified in the form, since it may

vary from country-to-country from under 18 to under 15. But in adapting Form GQ1 for your country the defined age group should be specified in the form so that respondents can answer relevantly.

[OVC Prevalence]

STEP 5 – If compilation of complete list results in less than 150 institutions prepare to conduct complete census of OVC to estimate size, or prevalence, of OVC population.

- (a) Visit every institution.
- (b) Obtain only basic information from each young person, such as name, age, gender and whether one or both parents are still alive.

STEP 6 – If compilation of complete list results in 150 or more institutions, prepare to select sample of 40-50 institutions. *This step would be repeated separately for boys and girls in countries where they are institutionalized separately.*

- (a) Assign institutions to small, medium or large stratum on basis of their estimated size in terms of number of OVC (see Chart 2).
- (b) Prior to sample selection arrange institutions in geographic fashion within each stratum by urban-rural and by province.
- (c) Figure sample selection interval, I – per Chart 2 – for each stratum to yield 40 (or 50) institutions.
- (d) For each stratum, select a random start using a random number table and select 1 in every I of the institutions. For example, if there are 180 institutions in the small stratum and you decide to select 18 of them, then $I = 10$ and the random start would be any randomly chosen number between 1 and 10.
- (e) Visit every sampled institution but conduct *census* of OVC in those institutions.
- (f) Obtain only basic information from each young person, such as name, age, gender and whether one or both parents are still alive.
- (g) Prepare prevalence estimates using, as weights for each stratum, the sampling interval, I ; apply appropriate weight to every OVC in a given institution according to the weight for that institution. For example, if institution A is sampled from the small stratum at the rate of 1 in 10, all OVC in that institution are included in the census and each one receives a weight of 10 to produce the prevalence estimates.

[OVC characteristics]

STEP 7 – Decide on sample size – minimum of 100 OVC for linked survey and 400 for stand-alone OVC survey (see Chart 1).

STEP 8 – Double the sample size if estimates are wanted for male-female separately.

STEP 9 – When the total number of institutions is so few (under 150) that all of them are included, prepare to select a systematic sample of OVC.

- (a) Visit every institution.
- (b) Obtain a roster of all the OVC at each institution from the official in charge, with only basic information for each young person, such as name, age, gender.

- (c) Arrange the rosters (male-female separately when required) in a continuous stream (as though from a single source).
- (d) Calculate the sampling interval, I , equal to N/n , where N is the total number of OVC in all the institutions combined and n is the sample size (for example, 100 or 400). Calculate separate intervals if male-female are to be separate domains. Find a random start in a table of random numbers between 1 and I . An illustration follows of systematic sampling (the illustration is for only 20 male OVC in a population of 72, to enable it all to fit on two pages).

Chart 4. Illustration of Systematic Selection of 20 Male OVC

| Name of Facility | Male OVC | No. | Selection Random Start = 0.4 | Name of Facility | Male OVC | No. | Selection |
|------------------|----------|-----|--|------------------|----------|-----|-----------|
| Institution 1 | Name | 01 | | | Name | 37 | |
| | Name | 02 | | | Name | 38 | |
| | Name | 03 | | | Name | 39 | |
| | Name | 04 | 4.0 | | Name | 40 | 40 |
| | Name | 05 | | Institution 6 | Name | 41 | |
| Institution 2 | Name | 06 | | | Name | 42 | |
| | Name | 07 | | | Name | 43 | |
| | Name | 08 | 7.6 | | Name | 44 | 43.6 |
| | Name | 09 | | | Name | 45 | |
| | Name | 10 | | | Name | 46 | |
| | Name | 11 | 11.2 | | Name | 47 | 47.2 |
| | Name | 12 | | Institution 7 | Name | 48 | |
| Institution 3 | Name | 13 | | | Name | 49 | |
| | Name | 14 | | | Name | 50 | |
| | Name | 15 | 14.8 | | Name | 51 | 50.8 |
| | Name | 16 | | | Name | 52 | |
| | Name | 17 | | | Name | 53 | |

| | | | | | | | |
|---------------|------|----|------|---------------|------|----|------|
| | Name | 18 | 18.4 | | Name | 54 | 54.4 |
| | Name | 19 | | | Name | 55 | |
| | Name | 20 | | | Name | 56 | |
| | Name | 21 | | Institution 8 | Name | 57 | |
| | Name | 22 | 22 | | Name | 58 | 58 |
| Institution 4 | Name | 23 | | | Name | 59 | |
| | Name | 24 | | | Name | 60 | |
| | Name | 25 | | | Name | 61 | |
| | Name | 26 | 25.6 | | Name | 62 | 61.6 |
| | Name | 27 | | | Name | 63 | |
| | Name | 28 | | | Name | 64 | |
| | Name | 29 | 29.2 | Institution 9 | Name | 65 | 65.2 |
| | Name | 30 | | | Name | 66 | |
| | Name | 31 | | | Name | 67 | |
| | Name | 32 | | | Name | 68 | |
| Institution 5 | Name | 33 | 32.8 | | Name | 69 | 68.8 |
| | Name | 34 | | | Name | 70 | |
| | Name | 35 | | | Name | 71 | |
| | Name | 36 | 36.4 | | Name | 72 | 72.4 |

(e) Each OVC selected is then eligible for the detailed interview.

STEP 10 – When the number of institutions is large enough to require sampling, that is, 150 or more, prepare to select a two-stage sample.

- (a) Prior to sample selection arrange all institutions in geographic order by province and within province by urban-rural.
- (b) Cumulate the measures of size consecutively and select a systematic, *pps* (probability proportionate to size) sample of 40 to 50 institutions. An illustration is given in Chart 5.

Chart 5. Illustration of Systematic *pps* Selection of 45 Institutions, Other Group Quarters

| Administrative Area, Name of Facility | | | Measure of Size (no. of OVC) | Cumulative | Sample Selection | | |
|---|-------|-------------|---------------------------------|-------------|------------------|-----|-------|
| Province 01 | Urban | Facility 01 | 22 | 22 | | | |
| | | Facility 02 | 7 | 29 | | | |
| | | Facility 03 | 9 | 38 | 31.1 | | |
| | | Facility 04 | 23 | 61 | | | |
| | | Facility 05 | 12 | 73 | | | |
| | | Facility 06 | 6 | 79 | | | |
| | | Facility 07 | 10 | 89 | 87.2 | | |
| | Rural | Facility 07 | 16 | 105 | | | |
| | | Facility 01 | 20 | 125 | | | |
| | | Facility 02 | 20 | 145 | 143.3 | | |
| Province 02 | Urban | Facility 03 | 8 | 153 | | | |
| | | Facility 04 | 5 | 158 | | | |
| | | Facility 01 | 25 | 183 | | | |
| | | Facility 02 | 17 | 200 | 199.4 | | |
| | | Facility 03 | 10 | 210 | | | |
| | | Facility 04 | 9 | 219 | | | |
| | | Facility 05 | 12 | 231 | | | |
| | | Facility 06 | 8 | 239 | | | |
| | | Facility 07 | 9 | 248 | | | |
| | Rural | Facility 08 | 6 | 254 | | | |
| | | Facility 09 | 22 | 276 | 255.5 | | |
| | | Facility 01 | 14 | 290 | | | |
| | | Facility 02 | 20 | 310 | | | |
| | | Facility 03 | 21 | 331 | 311.6 | | |
| | | Province 03 | Urban | Facility 01 | 16 | 347 | |
| | | | | Facility 02 | 27 | 374 | 367.7 |
| | | | | Facility 03 | 21 | 395 | |
| | | | | Facility 04 | 23 | 418 | |
| Facility 05 | 7 | | | 425 | 423.8 | | |
| Facility 06 | 7 | | | 432 | | | |
| Facility 07 | 11 | | | 443 | | | |
| Facility 08 | 12 | | | 455 | | | |
| | * | | * | * | * | | |
| | * | * | * | * | | | |
| | * | * | * | * | | | |
| Province 20 | Urban | Facility 01 | 17 | 2447 | 2443.4 | | |
| | | Facility 02 | 18 | 2465 | | | |
| | | Facility 03 | 12 | 2477 | | | |
| | | Facility 04 | 10 | 2487 | | | |
| | Rural | Facility 01 | 8 | 2495 | | | |
| | | Facility 02 | 9 | 2504 | 2499.5 | | |
| | | Facility 03 | 12 | 2516 | | | |
| | | Facility 04 | 8 | 2524 | | | |
| SAMPLE PARAMETERS: Selection Interval = 2524/45, OR 56.1; Random Start = 31.1 | | | | | | | |

Note that the same facility may be selected more than once if its measure of size exceeds the sampling interval, *I*. If that should happen, the number of OVC to select in such a facility is double for two “hits,” triple for three “hits” and so forth.

(c) In each sample facility prepare to select a fixed number of OVC – say 5 males and 5 females – but the number should be equal to n divided by the number of institutions selected (for example, if n is 400 and 40 facilities are chosen, then the fixed sample size in each facility is 10). When there are fewer OVC than the required fixed number in a given facility, interview all of them.

(d) Obtain a roster of the OVC in each sample institution and select a systematic sample from each one in the same way as described above in Steps 9 (a) through (e).

STEP 11 – Maintain detailed documentation of the sampling operations and procedures that are used, noting especially any implementation features that depart from the design.

Homeless Youth Component

STEP 1 – Decide on important survey, as opposed to sample, parameters: precise operational definition of street child or homeless child; coverage only in large cities where street children concentrate or elsewhere.

STEP 2 – Prepare to construct sampling frame of time-location *PSUs*.

STEP 3 – Develop sampling frame by contacting government officials, NGOs, religious leaders and other key informants who are knowledgeable about locations where homeless youth are known to sleep.

STEP 4 – Make appropriate contacts at national level such as ministries of health, social welfare or others that deal with youth; national headquarters of NGOs, national headquarters of religious groups.

STEP 5 – Make similar contacts also at province level and at district level (or city government level if survey is restricted to cities).

STEP 6 – Through contacts above, compile comprehensive list of all locations where OVC sleep.

Following is a suggested questionnaire, Form S1, to use when contacting government officials and other experts for this purpose. The form is illustrative and would be adapted as necessary to fit a particular country's conditions.

Illustrative Form 2. Illustrative Data Collection Form to Compile Sleep Sites

| | |
|---|---|
| Form S1 – Compilation of List of Sleep Sites for Homeless Youth <i>[Introduce yourself and explain the purpose of the survey].</i> | <i>Administer this form to officials of the ministries of health, social services and others that work with youth; to religious leaders and to NGOs that work with youth. Complete a separate form for each contact. Continue on additional forms as necessary.</i> |
| We must compile a list of all the places where street children and other homeless youth are known to sleep, so that we may find them to conduct the survey. | |
| 1. Are there any shelters ¹⁹ for homeless persons, including young people, in the (country, city, province)? | List them with name, address/location. 4. _____ 5. _____ 6. _____ |
| 2. Are there abandoned buildings used by youth for sleeping? | List them with name, address/location. 1. _____ 2. _____ 3. _____ |
| 3. What about other enclosed structures such as shopping malls, transportation terminals? | List them with name, address/location. 1. _____ 2. _____ 3. _____ |
| 4. Any other enclosed structures that you know about? | List them with name, address/location. 1. _____ 2. _____ 3. _____ |
| 5. What about outside? Under bridges? | List them with name, address/location. 1. _____ 2. _____ 3. _____ |
| 6. Underground, including sewage tunnels? | List them with name, address/location. 1. _____ 2. _____ 3. _____ |
| 7. Squatter areas? | List them with name, address/location. 1. _____ 2. _____ 3. _____ |
| 8. On the street, especially out-of-the-way spots? | List them with name, address/location. 1. _____ 2. _____ 3. _____ |

¹⁹ Homeless shelters are included as “sleep sites” for street children on the grounds of their transient nature, as opposed to being considered group quarters or institutions.

STEP 7 – On basis of responses to Form S1, create list of time-location *PSUs*, using 6-hour segments for time dimension.²⁰ An example of four of the *PSUs* might be as follows:

Under City Bridge – 6 a.m. to noon
Under City Bridge – noon to 6 p.m.
Under City Bridge – 6 p.m. to midnight
Under City Bridge – midnight to 6 a.m.

STEP 8 – Prepare to conduct field work to obtain measure of size for each *PSU* in the entire universe of *PSUs*. Figure workload in terms of number of *PSUs* in universe and field staff available. Note that each *PSU* may require 6 person-hours of work since staff will have to remain at site for 6 hours to obtain count.

Illustration: Suppose there are 300 *PSUs* and you intend to use 15 field personnel. Then, 20 *PSUs* ($300 \div 15$) would be assigned per person; the total time required, per worker, is 120 hours (20×6), or the equivalent of 3 person-days for each. Altogether, it would be 15×3 , or 45 person-days of work (in this illustration) to compile the measures of size. Alternatively, the number of field workers required can be calculated instead through a corresponding process if the number of person-days is fixed.

STEP 9 – Visit each *PSU*²¹ to obtain approximate count of homeless persons present or arriving during time interval; do not attempt to screen for OVC status or age group (this is done later).

STEP 10 – Establish 3 strata of *PSUs* according to size categories based on their measures of size – small, medium and large; see Chart 3 for rule-of-thumb, though exact size categories would be decided differently for each country.

STEP 11 – Decide upon target sample size. See Chart 1: minimum of 400 OVC for characteristics in a stand-alone survey, 100 for prevalence in a linked survey.

STEP 12 – Multiply sample size by 4 to determine number of homeless persons to be screened (only about 25 percent of which are expected to be OVC). *Note that in the particular application for a country the ratio of homeless to OVC needs to be more carefully determined, rather than just assuming it is 4. In that case, the multiplier would be different of course.*

STEP 13 – Calculate average size, \bar{n} , of *PSU*, in terms of number of homeless persons. This equals the total estimated measure of size over all *PSUs* divided by the number of *PSUs*. For example, if total number of homeless persons is 8000 and there are 320 *PSUs* in universe, then \bar{n} is 25 ($8000 \div 320$).

²⁰ Alternatively, 6 *PSUs* of 4-hour time intervals may be constructed.

²¹ Note that this implies visiting the same *location* 4 times, in order to establish the measure of size for each time interval.

STEP 14 – Calculate number of *PSUs* to select to reach sample size wanted (Step 11). For example, if 400 OVC is the sample size, then 64 *PSUs* would be needed, calculated as $400 \div (0.25 \times \bar{n})$.

STEP 15 - Calculate the average *PSU* size in each stratum (same as Step 13 but separately by stratum).

STEP 16 – Distribute the number of *PSUs* to select from each stratum according to the average size, from Step 15, and Chart 3.

Illustration: suppose the values of \bar{n} for each stratum are 4, 16 and 40, respectively, for small, medium and large. Recall this is the average number of homeless persons, and the average number of OVC is expected to be about 25 percent of those numbers, or, respectively, 1, 4 and 10. Suppose further that there are 160 *PSUs* in stratum 1 (small), 100 in stratum 2 (medium) and 60 in stratum 3 (large). If the sample size is 400, a plausible sample scheme would be to select 1 in 11 of the stratum 1 *PSUs* (yielding 15 or 14), 1 in 4 of the stratum 2 *PSUs* (yielding 20) and one-half the stratum 3 *PSUs* (yielding 30). The total number of *PSUs* would be 65 (or 64) and the *expected* number of OVC would be equal to $[(15 \times 1) + (20 \times 4) + (30 \times 10)] = 395$.

STEP 17 – Select the *PSUs* using systematic selection; in each stratum choose a random start between 0.1 and the sampling interval and successively add the sample interval to designate the selected *PSUs*. The methodology for systematic selection is similar to that illustrated in Chart 4 of Step 9(d) of the Institutional Component section.

STEP 18 – Prepare to conduct the interviews in the selected *PSUs*.

STEP 19 – Set the time interval at 6 hours (or whatever interval that was used to establish the *PSUs*)²² that interviewers are to remain at each selected site for interviewing. Be sure that it is the *same* for every *PSU*.

STEP 20 – Post an interviewer (or team of interviewers) in each sample *PSU* site. Remember that the *PSUs* span all hours of the day and one-fourth of them are from midnight to 6 a.m., so some interviewers must work in the middle of the night. The interview team must remain at the site for the entire 6-hour interval in order to ensure accurate estimation later during the analysis stage.

Step 21 – Screen the occupants of the sleep site for OVC status, especially age, eliminating those who are 18 and older. This may require a screening questionnaire for persons who cannot be eliminated as adults clearly from observation. Perform this activity not only for occupants already present but others who show up during the time interval.

²² If time-location *PSUs*, for example, are established in 4-hour segments, then the interval of time that interviewers must be stationed for the survey at the sample sites is also 4 hours.

STEP 22 – Administer detailed OVC questionnaire to all OVC identified in the *PSU*. Note that the text recommends this questionnaire include questions to ascertain whether the respondent slept in other locations during the survey period – say, “last two weeks” – and, if so, their locations, in order to use in weighting at the analysis stage. See illustration in the text, repeated below.

To illustrate: It is quite likely the same site will fall into sample more than once, since the site will comprise 4 *PSUs* based on time of day. Any youth accustomed to sleeping in the same place has a good chance of showing up at the location at different time intervals, thus giving him/her multiple chances of selection. When that occurs, the youth should be interviewed only once. A somewhat more difficult duplication problem to sort out concerns those youth that sleep in *different* sites from one day to the next. To overcome this problem we need to take into account the length of the survey period – one week, two, etc. Let us assume 2 weeks for the example. It is necessary to ask each respondent how many other places he/she usually sleeps during a 2-week period. She would also be asked to identify those locations.

STEP 23 – [When survey is linked only]. Ask respondent if there is a usual place of residence where he/she normally sleeps, information that is required in order to un-duplicate with the household or institutional frame.

STEP 24 – In office operation, match sites mentioned by respondents from Steps 22-23 against master list of sites that make up the *PSU* frame. For each match, the survey weight for that respondent is down-weighted by the factor, $1/t$, where t is the number of matching sites.

STEP 25 – Maintain detailed documentation of the sampling operations and procedures that are used, noting especially any implementation features that depart from the design.

Appendix D. References and Other Reading

Bicero, G., S. Rutstein and K. Johnson (2003), "Dimensions of the Emerging Orphans Crisis in Sub-Saharan Africa," *Social Science and Medicine*, 56(6): 1235-1247. Rotterdam: Elsevier.

Centers for Disease Control (2002), *Handbook for Conducting Youth Risk Behavior Surveys*. Atlanta: CDC:

Chandrasekar, C. and Deming, W. Edwards (1949), "On a Method of Estimating Birth and Death Rates and the Extent of Registration," *Journal of the American Statistical Association*, 1949: (44), 101-115. [This article is source of capture – recapture theory as applied to human populations].

Family Health International (2000), *Behavioral Surveillance Surveys (BSS)*, Guidelines for Repeated Behavioral Surveys in Populations at Risk of HIV, Research Triangle Park: FHI.

Greenwell, K. Fern (2002), "A Demographic Profile of Rwanda's Children in Residential Care." Rwanda: UNICEF.

Kalton, Graham (1993), *Sampling Rare and Elusive Populations*. New York: UN Statistical Office.

Kish, Leslie (1965), *Survey Sampling*. New York: John Wiley.

Macro International Inc. (1996), *Demographic and Health Surveys (DHS) Sampling Manual*, DHS-III Basic Documentation 6. Calverton, MD: Macro International Inc.

Turner, Anthony G. (2000), Chapter 4: "Designing and Selecting the Sample," Chapter 6: "Conducting the Fieldwork" and Appendix 7: "Sampling Details," *End-Decade Multiple Indicator Survey Manual, Monitoring Progress Toward the Goals of the 1990 World Summit for Children*. New York: UNICEF.

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

Turner, Anthony (1998), "Some Notes on Village Sampling," *Proceedings of the Joint IASS/IAOS Conference on Statistics for Economic and Social Development*. Aquascalientes, Mexico.

UNAIDS/UNICEF/USAID (2002), *Children on the Brink 2002*, Joint Report on OVC Estimates and Program Strategies. Geneva, New York, Washington: UNAIDS.

UNAIDS (2000), *National AIDS Programmes – A Guide to Monitoring and Evaluation*. Geneva, New York, Washington: WHO, UNICEF, USAID.

UNAIDS/UNICEF (2003), Report on “Technical Consultation on Indicators Development for Children Orphaned and Made Vulnerable by HIV/AIDS,” 2-4 April 2003 in Gaborone, Botswana. New York: UNICEF.

World Bank (2000), *Child Needs Assessment Tool Kit*, Task Force for Child Survival and Development. Washington: World Bank.