DESIGN, IMPLEMENTATION, MONITORING, AND EVALUATION OF CROSS-CULTURAL HIV-RELATED MENTAL HEALTH AND PSYCHOSOCIAL ASSISTANCE PROGRAMS:

A USER'S MANUAL FOR RESEARCHERS AND PROGRAM IMPLEMENTERS

(ADULT VERSION)

MODULE 3:

POPULATION MEASUREMENT

Applied Mental Health Research Group
Center for Refugee and Disaster Response
Johns Hopkins University Bloomberg School of Public Health

The USAID | Project SEARCH, Task Order No.2, is funded by the U.S. Agency for International Development under Contract No. GHH-I-00-07-00032-00, beginning September 30, 2008, and supported by the President's Emergency Plan for AIDS Relief. The Research to Prevention (R2P) Project is led by the Johns Hopkins Center for Global Health and managed by the Johns Hopkins Bloomberg School of Public Health Center for Communication Programs (CCP).

ACKNOWLEDGEMENTS

The development of the original version of this manual was supported by a series of grants from the Victims of Torture Program (VOT) at USAID to The Applied Mental Health Research Group (AMHR) at Johns Hopkins University (JHU). VOT also provided support for the use and refinement of the approach described here in field-based studies with implementing organizations assisting populations affected by torture and violence.

This version is an expansion and adaptation of that original manual with specific reference to research on psychosocial and mental health problems related to HIV. This adaptation process was implemented by AMHR through Project SEARCH | USAID: Research to Prevention (R2P), funded by the United States Agency for International Development (USAID) through task order # GHH-I¬00-07-00032.

An adaptation of the original VOT Module 1 for children has also been written. This was supported by World Vision USA under their Children in Crisis initiative. Under that same initiative, World Vision USA has also supported collaborations between JHU and World Vision national offices in the use of these methods to inform programming for street children and those sexually abused in multiple countries. World Vision USA also supported previous work with adults that also contributed to the development of the methods originally described in the adult version of the module.

Without the support of both USAID/VOT and World Vision USA, the development of the methods described here and the production of this manual would not have been possible.

CONTENTS

ACRONYMS

AIDS	Acquired Immunodeficiency Syndrome
AMHR	Applied Mental Health Research
ART	Antiretroviral therapy
BA	Behavioral activation
CBI	Components based intervention
CBT	Cognitive Behavior Therapy
CD4	T-helper cell targeted by HIV
CDC	Centers for Disease Control
CPT	Cognitive Processing Therapy
CSA	Child Sexual Abuse
DHS	Demographic health survey
DIME	Design, implementation, monitoring and evaluation
DRC	Democratic Republic of Congo
EBT	Evidence based treatment
FG	Focus Group
FL	Free List
GBV	Gender based violence
HIN	Health information network
HIV	Human immunodeficiency virus
IDU	Injecting drug user
IPT	Interpersonal Therapy for Depression
IRB	Institutional Review Board
JHU	Johns Hopkins University
KAP	Knowledge, attitudes and practices
KI	Key Informant
LGBT	Lesbian, gay, bisexual, transgender
LMIC	Low and middle income countries
MEMS	Medication Event Monitoring System
MI	Motivational interviewing
MOH	Ministry of Health
MSM	Men who have sex with men
NGO	Non-governmental organization
OVC	Orphans and vulnerable children
PE	Prolonged exposure
PLWHA	People living with HIV/AIDS

POFO	Positive Outcome for Orphans Study
PPS	Probability proportional to size
PRA	Participatory Rural Appraisal
PTSD	Posttraumatic Stress Disorder
R2P	Research to Prevention
RCT	Randomized Controlled Trials
REC	Research Ethics Committee
SEARCH	Supporting Evaluation and Research to Combat HIV/AIDS
SES	Social economic status
SMS	Short Message Service
SOW	Scope of Work
SRP	Stress related response
STI	Sexually transmitted infections
SW	Sex worker
USAID	United States Agency for International Development
TFCBT	Trauma Focused Cognitive Behavior Therapy
VCT	Voluntary counseling and testing
VOT	Victims of Torture Program
WHO	World Health Organization

INTRODUCTION TO THE MANUAL

The Manual for Design, Implementation, Monitoring, and Evaluation of Cross-Cultural HIV-Related Mental Health and Psychosocial Assistance Programs: A User's Manual for Researchers and Program Implementers has been written to assist researchers and organizations developing and implementing programs in HIV affected populations to 1) identify and measure the impact and prevalence of mental health and psychosocial problems in the populations they seek to serve; 2) to develop or adapt appropriate interventions to address these problems; and 3) to measure the impact of these interventions. The Manual consists of 6 modules. Collectively, the modules describe a process of program **d**esign, **i**mplementation, **m**onitoring, and **e**valuation (DIME) that has been developed and used by the authors since 2000. The modules may be used in sequence, to follow the life of a project, or as stand-alone units to address a specific project need.

- **Module 1** describes procedures for a qualitative assessment to identify priority problems from the local perspective.
- **Module 2** provides guidance in the development and validity testing of tools to measure these priority problems.
- **Module 3** describes population-based assessments to gauge prevalence and severity of the priority problems using the instrument developed in Module 2.
- **Module 4** describes a process for overall design of a program to address the priority problems, including design of program monitoring and evaluation.
- **Module 5** outlines the selection, adaptation, and implementation of interventions.
- **Module 6** describes procedures for assessing intervention impacts.

Definition Box

Intervention(s): Service(s)/activity(ies) directly benefitting the client

Program: The intervention(s) and all ancillary activities necessary to support the intervention(s): logistics, finance monitoring and evaluation, etc.

LAYOUT OF THE MANUAL

Modules are presented in narrative form, with extensive use of subheadings. With the exception of text boxes, each section and each paragraph is meant to be read sequentially. Additional material that is useful as examples of concepts or expansion on subjects discussed in the text has been included in text boxes. Examples of study materials which may be adapted for use in an actual study are placed separately as appendices.

 This symbol indicates that what follows is a critical requirement or constraint.

INTENDED USERS

This manual is primarily intended for researchers and groups responsible for mental health and psychosocial interventions for HIV affected populations, such as government providers and non-governmental organizations (NGOs).

The methods described in each module are intended to be within the typical budget, resources, and time constraints of organizations that normally focus on implementation rather than data collection. The approach is designed to be used in a limited area among a population with a homogenous language, culture, and similar circumstances. In areas containing populations with a variety of languages, cultures, and environments, the approach described in this manual should be used separately with each group. For this reason, the authors have focused on developing a process that is rapid and relatively inexpensive.

This is meant as a **'user'** manual rather than a training manual. It is intended for use in the field by those who have previously received field-based training in its methods (or have similar training experience) and are now leading teams in their own sites. Such persons should either have some prior experience in qualitative and quantitative data collection methods (depending on the module being used) or lead teams with persons who have such experience.

 THIS MANUAL IS NOT APPROPRIATE FOR 'OFF THE SHELF' USE WITHOUT PRIOR ON-THE-GROUND TRAINING OR SIMILAR EXPERIENCE. THOUGH WHAT IS PRESENTED HERE REPRESENTS WHAT THE AUTHORS HAVE FOUND TO WORK WELL TO DATE, FIELD SETTINGS VARY. USERS OF THE METHODS PRESENTED HERE NEED FIELD EXPERIENCE TO INTERPRET AND ADAPT THESE METHODS TO DIFFERENT SITUATIONS.

The authors have found that even with prior experience in data collection, individuals and organizations attempting to use the methods described here for the first time will have many important questions during the process that cannot be addressed in the manual itself. Answering these questions as they arise—and developing the skills required for using the approaches in different settings—is best done in a field-based training situation, with direct instruction in the course of supervised use of this approach among a local population. Even after training, organizations using this approach may want guidance and ad hoc assistance.

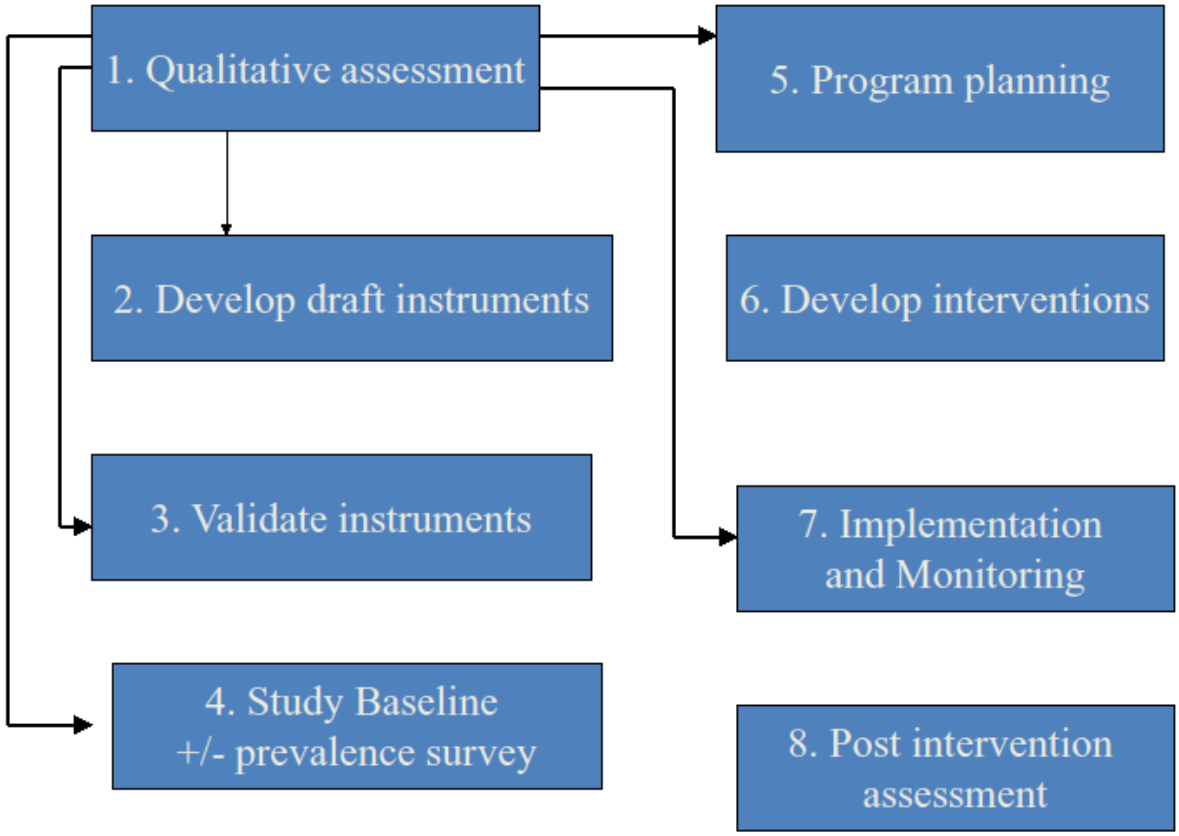
The authors would be pleased to discuss training and technical assistance with any interested organization or individual.

The manual does not contain detailed descriptions of commonly done research activities, such as quantitative interviewing, partly due to the expectation that organizations have persons experienced in these activities and partly because there are many other manuals available that describe these activities. Instead, the manual focuses on research activities or methods that are different from commonly used approaches. For example, Module 1 contains much more information on interviewing than the other modules because the qualitative methods used in Module 1 are less commonly used than quantitative methods.

THE DIME MODEL

The diagram below outlines the steps of the **d**esign, **i**mplementation, **m**onitoring, and **e**valuation (**DIME**) process described in this manual. Qualitative data collection (Module 1) is the first step in the process and the diagram indicates which of the subsequent steps (2-8) are informed by qualitative data. A brief description of each step follows.

Figure 1: Steps of the DIME Process



1. **Qualitative Assessment to identify and describe priority HIV-related mental health and psychosocial problems: (Module 1)**

   Variations in culture and environment affect how people understand the mental health and psychosocial problems related to HIV. By *understand,* we mean how these problems are described, how they are prioritized, their perceived causes, and how people currently cope with them. This information is vital in selecting problems that are important to local people, accurately communicating with them about these problems, and identifying interventions that are likely to be acceptable and feasible for local people and therefore effective and sustainable.

2. **Develop draft instruments to assess priority HIV-related mental health and psychosocial problems: (Module 2)**

   Having decided which problems the program will address, we then draft quantitative assessment instruments to address these problems. These instruments have various uses, depending on the program: conducting community or clinic-based surveys; screening persons for inclusion in a specific intervention program (for programs in which not all people will be served); identifying those with severe problems who may need specialized services including referral; and monitoring and evaluating the effectiveness of services by tracking changes in severity and/or prevalence of the problems identified.

   The process of drafting appropriate instruments includes reviewing the published literature for measures that have already been developed for the selected problems and comparing available measures with the qualitative data to select the measure or measures that best match how local people describe the problem. These measures are then adapted to better fit local concepts.

   Drafting includes translation. Terminology suggested by translators often differs from that used by local populations, particularly by poor and uneducated people. Therefore, qualitative data is preferred as the best source for translating key concepts. Employing the words and phrases that local people actually use (as identified in the qualitative data) will improve the clarity of the instruments, thereby improving their acceptability and accuracy. The translators are instructed to utilize the qualitative data to directly translate all signs, symptoms, problems and topics in the instruments that were mentioned by interviewees in the qualitative study using the same words found in the qualitative data. Only where concepts are not mentioned in the qualitative data do the translators themselves choose the appropriate terms.

3. **Validate draft instrument(s): (Module 2)**

   Once translated, the draft instrument(s) must be piloted and tested for ease of use, clarity, acceptance (both by interviewers and interviewees), and accuracy in the field. Accuracy refers to reliability and validity, which in turn refer to whether the instrument gives the same result with repeated use or use by different interviewers (reliability), and whether it measures what it is supposed to measure (validity). Testing involves interviews with members of the target population using the assessment instrument and analyzing the results.

   Validity and reliability testing are particularly important with psychosocial and mental health measures, where assessment is based on the interview alone (i.e., there are no laboratory or other tests). A tool that is not accurate can lead to inappropriate inclusion/exclusion of intervention participants and also provide incorrect conclusions about need and program impact.

4. **Study baseline +/-prevalence surveys: (Module 3)**

   Both baseline assessments and prevalence surveys are based on the instruments developed in steps 2 and 3. Baseline assessments refer to interviews done using the instrument in order to establish the eligibility of individuals for participation in an intervention program. Prevalence surveys perform the same function at the population level to measure the percentage and numbers of eligible (i.e., affected) persons in the population as well as giving some indication about the variation in severity of problems at the population level.

5. **Overall program planning: (Module 4)**

   This includes planning the program goals and objectives and the strategy and the type of intervention(s) for achieving these. It also includes the development of process and impact indicators, and the overall program work plan.

6. **Develop interventions to address the identified HIV-related mental health And psychosocial problems: (Module 5)**

   The qualitative data on the perceived causes of problems and how those affected cope with the problems are critical to intervention design. Interventions need to address the perceived causes of priority problems (or explain why they do not) in order to make sense and therefore inspire both confidence and cooperation. The more closely interventions can match the ways in which people currently think about and address the

selected problems, the more likely the interventions are to be acceptable to them. Where there are differences, they need to be explained and agreed upon by the local population. For example, using counseling to address a problem that is thought to be caused by poverty will take some explaining.

7. **Implementation and monitoring: (Modules 4 and 5)**

   This refers to the implementation and monitoring of the intervention and the overall program. It includes procedures for iterative changes in the planned activities as needed, according to the monitoring data.

8. **Post intervention assessment: (Module 6)**.

   Upon completion of the intervention, participants are interviewed using qualitative methods to identify potentially important unexpected impacts of the program. They are also re-interviewed using the baseline quantitative instrument, to measure changes in the outcome indicators such as problem severity and function. Where possible, the amount of change is compared with the amount of change experienced by a control group, to determine the true program impact.

MODULE 3:

POPULATION MEASUREMENT

A. INTRODUCTION TO MODULE 3

A.1 PURPOSE AND OVERVIEW OF MODULE 3

Module 3 describes population-based prevalence assessments using the instrument developed in Module 2. By following the steps in this module, trained program staff should be able to:

- Select a sampling design, appropriate for the population of interest and context
- Implement a sample survey in the field
- Analyze the data to estimate the prevalence of the problem and/or function
- Disseminate results

A.2 PREVALENCE STUDIES IN STIGMATIZED POPULATIONS

Although the HIV epidemic is more than 30 years old, HIV is often a stigmatized or "hidden" disease associated with deviant sexual behaviors, injecting drug use, and marginalized populations. While it is true that HIV is treatable with antiretroviral therapy (ART) and manageable like many chronic illnesses, stigma still surrounds HIV. In past and recent times, people have been disowned, divorced, and/or physically and emotionally harmed resulting from disclosure of their HIV status. As knowledge of how HIV is transmitted is limited in many communities, rumors or false information regarding HIV often propagate HIV stigma and misunderstanding. Quantifying the prevalence of mental health concerns within an HIV-affected population may be difficult due to the double stigmatization of mental health and HIV. Additionally, if populations included in a prevalence survey do not accurately and adequately represent the target population of interest, prevalence estimates can be biased. For example, if a study seeks to understand mental health problems in sex workers (SW), but only female CSWs who work in brothels participate in the survey, then the prevalence does not represent **all** CSWs, but rather **only** CSWs working in brothels. The prevalence of mental health problem in CSWs observed in the study may therefore not provide an accurate estimate of the prevalence in the total population.

Methods to ensure accurate prevalence studies include:

- Defining the target population
- Selecting an appropriate sampling method
- Meeting in a comfortable, private place
- Building rapport between interviewer and interviewee
- Ensuring that interviewers are culturally appropriate, for example:
    - Peers interview peers
    - Age of interviewers is appropriate to age of interviewees

    - Age and gender power dynamics are considered
    - People living with HIV/AIDS (PLWHA) who are open about their status are considered as interviewers
- Ensuring interviewers are trained to maintain confidentiality
- Asking most sensitive questions towards the end of the interview
- Pilot testing the survey (Module 2) with appropriate representatives of the target population

At times, marginalized populations, such as prisoners, CSWs, injecting drug users (IDU) and men who have sex with men (MSM) are the populations of interest. Special protections for these populations are important as security and confidentiality issues may arise, particularly in cases where a certain risk behavior is illegal (such as illegal drug use). It is essential that in-country teams have access to these populations in a way that does not put the interviewee in danger or at risk of persecution or incarceration. Sampling methods can be utilized to protect these populations through anonymity. Furthermore, training interviewers on the importance of confidentiality and human subject protections, especially for these populations, is essential.

B. SELECTING A POPULATION SAMPLE

Once the study instruments have been developed, tested and refined (see Module 2), they can be used in a prevalence survey among the target population. This section of Module 3 begins with a multi-step process for choosing a random sample of the target population for such a survey. As this manual is intended to be used in the context of HIV-infected populations, the sample selection process described here assumes that the survey will take place among a **defined** population of PLWHA. Potential sources of HIV-infected populations include clinic registers of HIV-infected patients, human rights organizations, and support groups for PLWHA). Some prevalence studies aim to assess the prevalence of mental health problems in HIV-affected households or villages. In such contexts, sampling frames might not be complete lists of everyone in the population, but would instead use some of the sampling schemes below to select households, villages or districts. In some areas, a defined population of persons living with or affected by HIV is unavailable for a variety of reasons. This module addresses these populations with non-probability sampling (Section B.5).

B.1. SELECTION OF HIV-INFECTED OR HIV-AFFECTED POPULATIONS

B.1.1. POPULATION SELECTION

Understanding mental health in HIV-infected or HIV-affected populations requires a definition of the population of interest. At times, finding and defining populations of HIV-infected persons can be difficult. There are some epidemiologic methods and logistical strategies that may make populations easier to find and define for sampling purposes. Both probability and non-probability sampling strategies are discussed in Sections B.2 through B.5.

An overall population must be determined prior to determining sampling methods (probability versus non-probability). For example, is the study working to understand prevalence of mental health conditions in:

- All HIV-infected persons in a village?
- Women accessing ART in a clinic?
- HIV-affected families using government-supported cash transfer schemes?

Once an overall population is determined, the study researchers must find out whether a listing of all the people in this overall population is available, or whether the population is unknown.

If a population is defined or completely understood, then probability sampling can be used. A *defined population* requires that there is a list or roster in existence of the people in the population (HIV-infected persons or HIV-affected persons) and those persons can be contacted

from this list. A defined or understood population of HIV-infected persons could include, but is not limited to, the following:

- Clinic register of HIV infected persons (e.g., VCT center or ART clinic)
- Government roster of HIV-infected persons
- Outreach program rosters
- Community support organization rosters
- Orphan and vulnerable children rosters from schools

A defined population of HIV-affected persons or regions could utilize:

- Community (e.g., village, district, etc.) rosters using HIV prevalence rates from a Ministry of Health (MOH) or other reliable source
- Family members of participants in a service delivery program or group as listed above

B.1.2. HIDDEN POPULATIONS AND NON-PROBABILITY SAMPLING

Sometimes a population of HIV-infected persons is an unknown or hidden population, or else privacy concerns do not allow for a roster of PLWHA. For example, understanding mental health among HIV-infected CSWs or IDUs may be difficult as sex work and drug use are often illegal behaviors, and fear of persecution keeps the population relatively hidden. Marginalized or stigmatized groups, such as MSM, CSWs, former prisoners, or IDUs, are examples of hidden populations. In order to find these hidden populations, non-probability sampling can be used. A well-known technique is called *snowball sampling* or *respondent-driven sampling*. This method is discussed in Section B.5.

B.1.3. LOCATION OF POPULATIONS: ISSUES WITH URBAN AND RURAL SITES

Both probability and non-probability methods can be used in rural and urban settings. In an urban setting, HIV-related surveys may be easier to complete due to larger populations with resulting increased access to HIV-related services and ability to access those services anonymously (health clinics, support groups, NGOs). However, a large population size does not mean that non-probability sampling methods are not indicated. Although populations are large in urban settings, social networks, especially within marginalized groups, may be very tight-knit or small.

In rural settings, the populations may be more hidden due to stigma associated with HIV or concerns about confidentiality. It is essential to recognize that stigma associated with HIV may prevent people from participating in a very small town or rural setting if the study seems to identify persons with HIV as infected to their communities.

B.1.4. OVERCOMING BARRIERS

Whether using probability or non-probability sampling methods, study staff should work closely with in-country program staff that is knowledgeable about: (1) their HIV-infected or HIV-affected clients, (2) cultural norms relevant to finding and interviewing PLWHA, and (3) methods of finding hidden populations who may be affected by HIV. Planning with in-country programs and partners that can navigate the social networks of HIV-infected and affected populations will help ensure that a survey is more acceptable to the community and more likely to be successful. It is important to recognize that hidden social networks may be very tight-knit; rumors or messages about the study may spread quickly in the community. This may be very helpful if the research is seen as helpful to the community. However, rumors may be damaging to a study if false or unfavorable information is spread. It is important to have the staff report rumors and problems encountered to the study supervisors and research team. As soon as problems arise in the community, especially negative rumors or mistrust of the study, meetings should be held between stakeholders and the study team to work to clear the misperceptions.

B.2. INTRODUCTION TO PROBABILITY SAMPLING

There are two main types of sampling methods: *probability* and *non-probability* samples. *Probability samples* employ some form of random selection mechanism to control for subjective bias. Each unit in the population has a known, equal, and non-zero probability of being selected. Probability samples enable researchers to measure the uncertainty (sometimes referred to as *confidence limits*) in making inferences from sample data to the population of interest.

Non-probability samples "are selected based on the judgment of the researchers to achieve particular objectives" (Henry, 1990). Because selection is non-random, the likelihood of being selected among members of the target population is not equal. Subjectivity is either tolerated or, in some cases, intended in the method. With non-probability samples, uncertainty (as a statistical matter) cannot be measured.

This module describes types of probability sampling only as these are approaches that will be most appropriate to measures of prevalence in an HIV-infected or -affected population. These are *simple random sampling*, *systematic random sampling*, and *cluster sampling*.

B.2.1. SIMPLE RANDOM SAMPLING

Using this method, individuals are randomly selected from a list that includes each member of the population. The list of the population from which the sample is drawn is referred to as the *sampling frame*. Following are the steps in simple random sampling:

1. Find or generate a *sampling frame*, which comprises any complete, up-to-date listing of individual population members in the target population.

2. Calculate a *sample size*. This is done using formulae that require a) the desired level of precision for measuring a particular indicator or characteristic; b) the estimated population prevalence of that characteristic; and c) the desired level of confidence in the results.

3. Draw *random numbers* from a random numbers table (see Appendix E) or a computer program (Excel, Epi-Info, SPSS, etc.). Each random number drawn would be matched to an individual in the sampling frame and that case would be located and interviewed.

4. Repeat random numbers drawing until the desired sample size is reached.

Simple random sampling is not simple to implement because of the need for a complete list of individuals in the population. But it is simple in theory and design. Results from a simple random sample can be analyzed without adjustments, recalculations, or qualifications. Real life, however, rarely is so simple. Available population listings are usually incomplete, unreliable, or unavailable, and the logistical constraints in finding selected individuals can be highly challenging and time-consuming in many settings, particularly in rural areas in lesser-developed countries. In HIV-infected populations, it may be possible to obtain a complete listing from a clinic program, support group, or government listing.

B.2.2. SYSTEMATIC RANDOM SAMPLING

Systematic random sampling is similar to simple random sampling except that only the first individual is chosen at random, while each subsequent individual is selected based on a systematic sampling interval. A systematic sample includes the following steps:

1. Find or generate a *sampling frame*, as for simple random sampling. The total number of persons in the sampling frame (i.e., the total number of the population from which the sample will be drawn) is also calculated.

2. Calculate a *sample size* appropriate to the study.

3. Calculate a *sampling interval*. This will be represented by i such that $i = N/n$ where N is the total number of persons in the population and n is the sample size. In a population of 15,000, for example, if the desired sample size is 300, then $i = 15{,}000/300 = 50$.

4. Select a single random number between 1 and i. This will be represented by r. In this example, let $r = 28$.

5. Select the first sampling unit and each subsequent unit. These will be r, r + *i*, r + 2*i*, r + 3*i*…until the sample size is reached. In this example, the first sampling unit would be the 28th name on the population listing or household number 28, the second would be 78 (28 + 50), the third would be 128 (28 + 100) and so on.

The main advantage of systematic sampling over simple random sampling is that it ensures that selection is equally distributed throughout the list, rather than having higher selection in some sections than others, as might happen with simple random sampling. This may be desirable when the list is ordered on the basis of an important variable, such as where people live, HIV related characteristics (i.e., risk group: injected drug use, sex work, etc.), and the researchers would like to ensure that all sub-populations are represented.

B.2.3. CLUSTER SAMPLING

As the name suggests, *cluster sampling* refers to the random selection of groups rather than individuals. Cluster sampling is most useful when a population is geographically dispersed or when a sampling frame is not available. As such, it has become widely used in crisis settings and developing countries. In the context of HIV-affected populations, cluster sampling may be used to select households in an HIV-affected village or villages in a district. Although cluster sampling may involve several variations, its most common form is *multi-stage* with *probability proportional to size* (PPS), which has the following steps (each of which will be considered in more detail in the next section):

1. Identify an approximate *sampling frame*, which, in the first stage, would comprise an estimate of the total population in a given geographic area as well as the populations of the sub-areas from which the sample will be drawn. These sub-areas are known as *primary sampling units*. In the example we will provide in the module, say that the sampling frame is a district of 50,000 people, comprising approximately 10,000 households across 30 villages, with the number of households in each village ranging from 50 to 1,200.

2. Calculate a *sample size* appropriate to the study. In cluster sampling, this includes consideration of population size, desired confidence limits, desired precision, and estimated prevalence of the characteristic or indicator of interest. In addition, cluster sample size calculations must take into account the fact that populations living in proximity to one another (that is, clusters) tend to be more alike than populations living farther apart. This correlation within clusters can produce estimates that are more variable than those produced using simple random or systematic sampling.

A *design effect* of 1 means that there is no correlation between individuals (or households) within clusters. A design effect greater than 1 implies positive correlation within clusters. One rule of thumb suggested for cluster sample size calculations is: lacking more specific information, use a design effect of 2. That is, double the sample size from what it would be if simple random or systematic sampling were used.

NOTE:

If the outcome of interest is more likely to occur within certain clusters than others, the estimates of prevalence will be less precise than if individuals were selected at random. This is because there is a certain probability of selecting some very high- or low-prevalence clusters for any given sample. This leads to an overestimate or underestimate of the true population prevalence for that given sample, making the estimate less precise. For the same reason, the sample size required to achieve a certain level of precision must be larger for cluster sampling than simple random sampling This inflation factor is the ratio of the variance obtained with cluster sampling to that obtained with simple random sampling and has been termed the 'design effect' (Katz & Zeger, 1994).

3. Determine the *number of clusters* to be selected. In this example, with a sample size of 600 households, there are no fixed criteria to say whether to sample 60 clusters of 10 households each, 10 clusters of 60 households each, 45 clusters of 13 households each, etc. Evidence from previous epidemiologic surveys, however, shows that around 30 clusters should yield acceptably valid and precise estimates and that, if resources are limited, it is preferable to decrease the number of basic sampling units per cluster rather than decrease the number of clusters (Binkin et al., 1995). In this example in the module, the sample is to be drawn in 30 clusters of 20 households each.

4. Calculate the *cluster selection interval*. In the example, there are 10,000 households in the sampling frame (N) and 30 clusters (*c*) in the sample so the selection interval (*i*) = N/c = 10,000/30 = 333.

5. Identify a *random start number* (*r*) between 1 and *i*. In this example, let *r* = 124.

6. Select the location of the first cluster and each subsequent cluster—r, r + *i*, r + 2*i*, r + 3*i*…—until the desired number of clusters is reached.

In the following two sections, the module will explore how clusters are selected from *principal sampling units* in multi-stage cluster sampling, and how households and individuals are selected from *basic sampling units* at the village level.

B.3. SELECTING PRINCIPAL SAMPLING UNITS

For any given probability sample of a population, several questions need to be answered:

- What is the target population (or population of interest)?
- What is the geographical distribution of that population?
- What information is available about that population in terms of estimated total size as well as size per a particular geographic area or administrative level (for example, country, province, district, township, village, etc.)?

Whatever the distribution of a target population, the principal sampling unit in multi-stage cluster sampling represents the first administrative level or geographical area for which there are multiple units, each with an estimated population size. If one were conducting a national survey, for example, the principal sampling unit might be provinces; if conducting a province-wide survey, the principal sampling units might be districts, and so forth. For studies of HIV-affected communities, sampling typically takes place in a relatively limited area among a population with a relatively homogenous language and culture. The example here assumes that the target population lives in a single district and the principal sampling units are villages (see Table 1 below).

Table 1: Cluster Sampling (Multi-Stage with Probability Proportional to Size)

a. Village Number	b. Number of Households per Village	c. Cumulative Total of Households	d. Selection Intervals in Village	e. Number of Clusters per Village
1	100	100		0
2	1,000	1,100	124, 457, 790	3
3	300	1,400	1123	1
4	200	1,600	1456	1
5	500	2,100	1789	1
6	100	2,200	2122	1
7	50	2,250		0
8	500	2,750	2455	1
9	300	3,050	2788	1
10	800	3,850	3121, 3454, 3787	3
11	100	3,950		0
12	50	4,000	4120	1
13	200	4,200		0
14	100	4,300	4453	1
15	250	4,550	4786	1
16	350	4,900		0
17	1,200	6,100	5119, 5452, 5785	3
18	100	6,200	6118	1
19	200	6,400		0
20	50	6,450		0
21	50	6,500	6451	1
22	300	6,800	6784	1
23	100	6,900		0
24	300	7,200	7117	1
25	400	7,600	7450	1
26	1,200	8,800	7783, 8116, 8449, 8782	4
27	100	8,900		0
28	800	9,700	9115, 9448	2
29	100	9,800	9781	1
30	200	10,000		0
Source: Adapted from (Brown et al., 1999).				

Following the steps for multi-stage cluster sampling with probability proportional to size (PPS), we will examine each of the six steps to selecting a 30-cluster sample of 600 households (because only one adult is selected from each household, this equates to a sample size of 600 people).

1. Identify a *sampling frame*. In this example, say that the sampling frame is a district of 50,000 people, or approximately 10,000 households. To create the sampling frame,

generate a list of all the villages and towns in the study area, their approximate populations, and the cumulative total population. (Though the order of villages or towns in the list is not particularly important, we recommend following listings that do not sequence by size). In the example provided, 30 villages are listed with household populations ranging in size from 50 to 1,200. Note that the population estimates are for numbers of households, **not** numbers of people. This is done because population listings in rural areas in developing countries rarely are based on estimated numbers of people and, if they are available at all, are more likely based on estimated numbers of households.

In order to select a cluster sample from among these villages, the list must include not only an estimated population per village, but also a cumulative population. In Table 1, Column (b) provides the village population figure while Column (c) provides the cumulative population figure, which is simply the given village population figure added to the cumulative sum of all previously listed villages.

2. Calculate a *sample size* appropriate to the study. In simple random or systematic random sampling, the formula for calculating a sample size to estimate prevalence.

In cluster sampling, in addition to the desired confidence limits, desired precision, and estimated prevalence of the characteristic of interest, we must take into account the "design effect." Because subjects within the same cluster are generally more similar to each other than to members of different clusters, there is a net loss of independent data, which results in decreased precision. The imprecision of cluster sampling is compensated for by multiplying the sample size calculated for a simple random sample by an inflation factor called the "design effect," which can range from 1.1 or 1.2 to 2 or more.

Using the formula for cluster sampling, we calculate a sample size by first setting a desired confidence limit of 95% (plus or minus 2 standard deviations or Z=1.96). To be conservative and because prevalence of depression and other mental health measures are so varied, we will estimate a prevalence of p=0.5. For desired precision, we will accept a margin of error of ±5%. Design effects for cluster samples of depression are not well established so we will accept a suggested cluster design effect of 1.3 (Aday, 1989). Therefore, sample size= $1.96^2[0.5*(1-0.5)]/0.05^2$ = 381.4, assuming a simple random sample. Rounding up and multiplying by a design effect of 1.3= 497. Finally, estimating that 10% of households initially contacted will refuse to participate and another 10% of households will fail to be found or will lack an eligible respondent, we would increase the sample size estimate by 20% to 596.4 rounded up to 600. In order to control for household clustering, we will employ a protocol of sampling only one eligible adult member per household. Therefore the sample of 600 households will equate with a sample of 600 adult individuals.

3. Determine the *number of clusters* to be selected. As noted previously, it is recommended that at least (and not necessarily more than) 30 clusters be selected (Binkin et al., 1995). Thus, for a sample of 600 households, the sample would be drawn in 30 clusters of 20 households each.

4. Calculate the *cluster selection, or sampling, interval*. The total population (that is, the cumulative population in the last row of the list in Table 1) is 10,000. Therefore, the sampling interval for a 30 cluster sample is 10,000/30=333.

5. Identify a *random start number* (*r*) between 1 and the sampling interval number (333). This can be done by using a random number table (see Appendix E). In this example, let *r* = 124.

6. Select the location of the first cluster and each subsequent cluster—r, r + *i*, r + 2*i*, r + 3*i*…—until the desired number of clusters is reached.

    - The first cluster is located in the town/village where the cumulative population of the village includes the random number. The cumulative population of the village where the first sampling site is located is the number closest to the random number **without being greater than** the random number. Note that in the example in Box 5.1, Village 1 (population 100) is skipped and the first cluster is located in Village 2 (population 1,000; cumulative population 1,100)

    - The second cluster is located in the town/village where the cumulative population of the village is equal to the sum of the *random number* and the *sampling interval.* The cumulative population of the village where the second sampling site is located is the number closest to this sum **without being greater than** the sum. Note that in the example below, the second and the third cluster are located in Village 2.

    - Identify the remaining clusters up to the 30th and final cluster. This is done by continuing to add the *sampling interval* to the sum obtained in locating the previous cluster until 30 sampling sites are identified.

This approach to selecting clusters is called sampling with *Probability Proportional to Size* (PPS), meaning that villages with larger populations are more likely to be selected (in some cases, more than one cluster will fall in the same village) and villages with smaller populations are more likely to be skipped. The PPS approach provides for a self-weighted sample and facilitates post-sampling data analysis. Note too, that cluster sampling (as opposed to simple random or systematic random sampling) has the advantage of reducing the total number of villages included in the sample (20 out of 30 total villages), thereby facilitating survey logistics. The sample is 'weighted' in that towns or villages with larger populations are more likely to be chosen. The probability of a town or village being selected is proportional to the relative size of its population.

The next, and final, step in cluster sampling is to decide on a sampling strategy within each village to identify the 20 households per cluster.

B.4 SELECTING BASIC SAMPLING UNITS

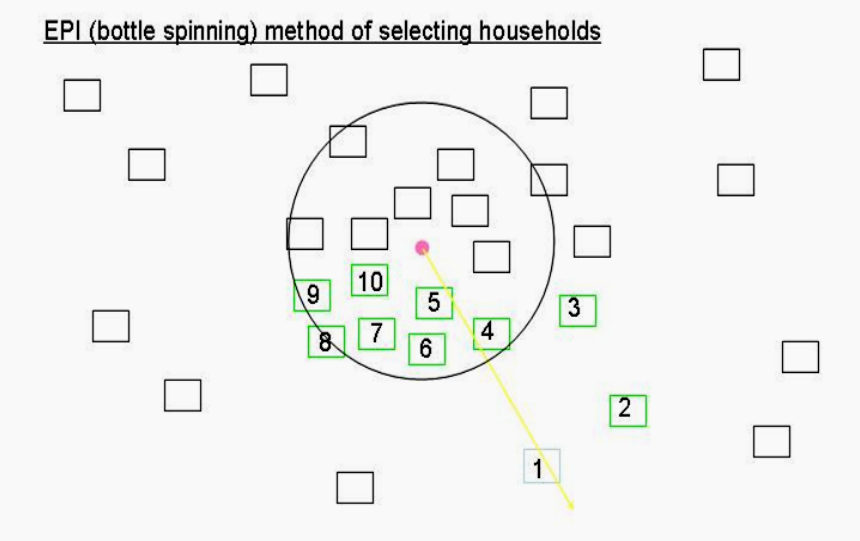
In multi-stage cluster sampling, the *primary sampling unit* may be as large as a province or as small as a village. The cluster sampling process may involve only two stages or it might involve three or more. In the final stage, the cluster sampling process comes down to selecting the households or individuals to be interviewed. These households or individuals are referred to as

the *basic sampling unit*, or sometimes also called the *ultimate sampling unit* or *listing unit* (Bennett et al., 1991). In nutritional surveys, the basic sampling unit might be a child (or all children) between the ages of 6-59 months. In mortality surveys, the basic sampling unit is the household, including all the members of that household who were alive during all or a portion of the recall period. In a prevalence survey of mental health problems or function among adults, the basic sampling unit would be an adult member of the household (if looking at mental health in villages) or the individual infected with HIV.

Field methods for selecting the designated number of basic sampling units in each cluster in the final stage of sampling include the following:

The EPI (Expanded Program on Immunization) method. Following the EPI method (sometimes called the "random walk" method) interviewers first choose a central location in the cluster site (village, camp section, etc.), choose a random direction from that location (by spinning a pen or bottle, for example), count the number of households between the central location and the edge of the site in that direction, select one of the households at random to be the starting point, then interview that household, then the one nearest, and so forth until reaching the designated number of basic sampling units for that cluster. The method is simple, widely known, easy to train users in, and likely results in a biased sample.

Figure 2: EPI Method of selecting households (Source: Woodruff)

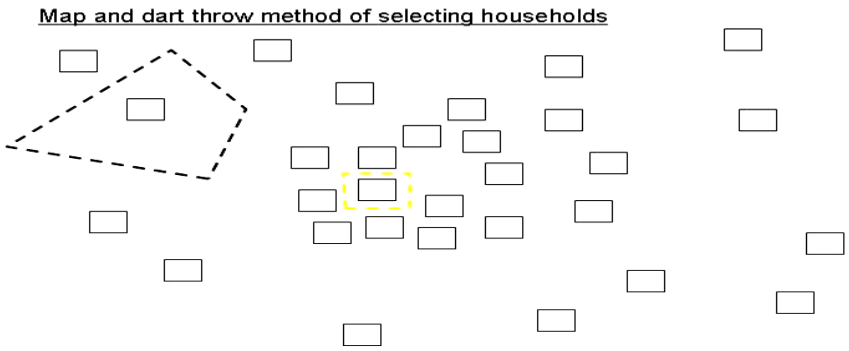


As the figure above suggests, not only are households closer to the center more likely to be selected initially but subsequent proximity selection is likely to include more households in the center than on the periphery. Simulations and field comparisons have demonstrated that the EPI method oversamples households at the center of the community and may be particularly

biased for variables that are likely to cluster in communities. Because of theses biases, the EPI method is not recommended (Bennett et al., 1994).

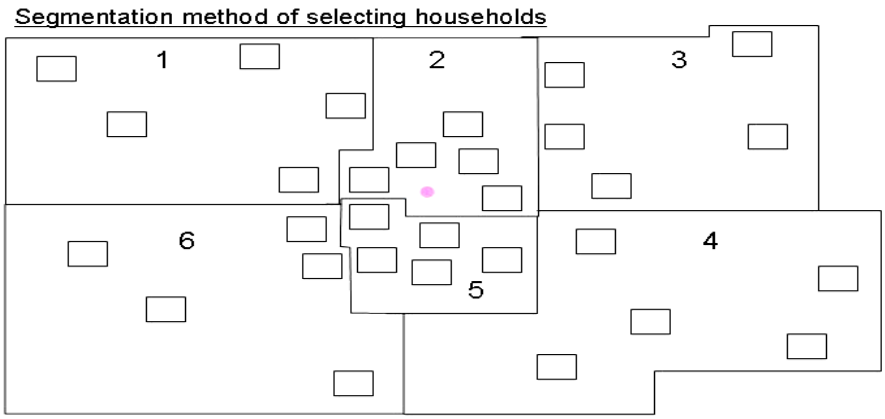
The "map and dart" method. With this method, the survey team creates a map of the cluster site (village, camp section, etc.) and then tosses a dart at the map and moves to the household closest to where the dart has landed. This method, however, biases selection in favor of households in less densely settled areas. As Figure 3 shows, the two households circumscribed by dotted lines do not have an equal probability of being selected since the two "selection" areas within which the dart might fall are not of equal size. A more "high-tech" alternative to "map and dart" is to select random coordinates within the cluster site (village, camp section, etc.) using GPS handheld receivers and then move to the household closest to those coordinates (Kaiser et al., 2003). This alternative, however, also gives a household in a less densely settled area a greater probability of being selected than a household in a more densely settled area (though this bias might be greater in rural areas than urban areas) (Grais et al., 2007).

Figure 3: The "Map and Dart" Method. (Source: Woodruff)



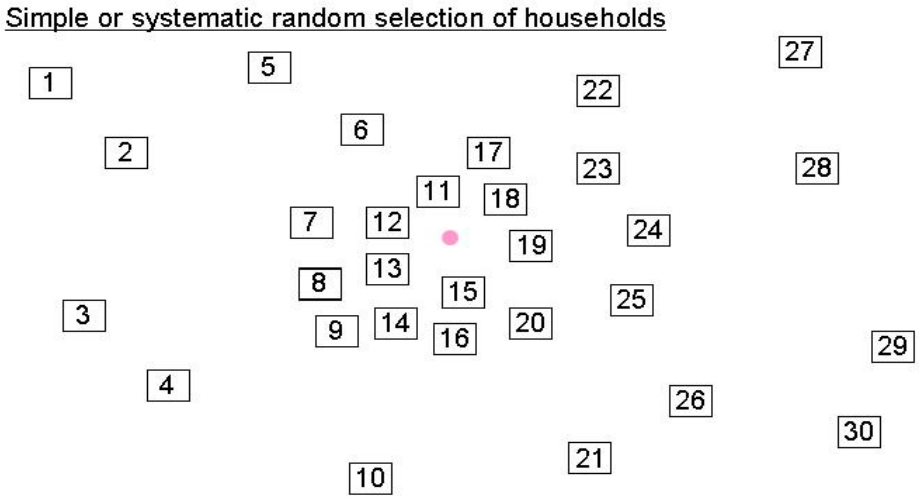
The "segmentation" method. Using this method, the survey team uses a local map to create "segments" with an approximately equal number of households in each segment. One or more of these segments would be randomly selected and all households would be interviewed in each segment until the designated number of households was reached. The disadvantage of this method is that it requires time to carry out local mapping and counting of households at the cluster site. The advantage is that obtaining up-to-date local information on household numbers and locations allows survey teams to check these numbers against the estimates used in the initial sampling process (Milligan et al., 2004).

Figure 4: The "Segmentation" Method. (Source: Woodruff)



Simple or systematic random selection method. If it is possible to number all of the households in the cluster site, then simple or systematic random sampling could be used to select the households for the cluster sample. This method is the least biased but logistically the most complex in that it requires local mapping and enumeration of households in the cluster site. As with the "segmentation" method noted above, this local mapping and enumeration can add time to the survey process but it also provides valuable, current information about the local population.

Figure 5: Simple Random or Systematic Selection Method. (Source: Woodruff)



While it is reasonable to assume that, of all the methods described above, simple or systematic random sampling would be the most effective in reducing bias in the final-stage sampling, it is not necessarily the most efficient method in all settings. Context matters, and organizations must decide what is the most appropriate method for their particular survey context, taking

into account such factors as security, population settlement patterns, survey resources (human, financial and logistical), and time available (Grais et al., 2007). All things considered, we recommend that the EPI method be avoided if possible and that a method be used (segmentation, for example) that incorporates local mapping and enumeration of the cluster site population. It is also recommended that, whatever the method chosen for final-stage sampling, the same method should be applied with reasonable consistency across all clusters.

Following are a set of steps to be taken in implementing the segmentation sampling method to select basic sampling units in each sample village:

1. A survey supervisor and interviewer should visit each one of the selected villages (20 in this example).
2. The supervisor and interviewer find someone knowledgeable about the area, such as a village leader.
3. This person is asked to draw a rough map of the village or settlement. It need only provide basic landmarks, such as roads, streams, churches, schools, and other important features. The map-maker should try to mark out sectors (segments) of dwellings delineated by these landmarks and record on the map the approximate number of houses in each sector. Sectors are usually described according to existing divisions (roads, streams, etc.) and therefore the number of houses may vary. In some cases, the number of houses in some sectors may be more than 100 or less than 20. The number of houses listed for each sector need not be exact but should be a reasonably accurate estimate.
4. For each village, make a numbered list of all the sectors, the approximate number of houses in each sector, and the cumulative total number of houses.
5. Chose a random number between 1 and the total number of houses in the village/town using a random number table (see Appendix E).
6. Check this number against the cumulative total to determine the sector in which the number occurs. That sector is then chosen. Select one sector for each cluster identified for that village/town. Note that large villages and towns may have more than one cluster located within them. Therefore, two or more sectors are selected within those.
7. Enumerate all the houses within the chosen sector(s). This is usually done by placing a chalk number on every house. Either simple random or systematic random sampling is then used to identify the 20 households per cluster.
8. At each chosen house, the interviewer asks any person living there to list all the people living in that house aged 18 years or more, in any order.
9. The interviewer records the names as a numbered list, then uses the random number table to choose one person. This person is the respondent and they are interviewed on the spot if available and willing.

10. If the chosen person is not present or does not want to be interviewed at that time, the interviewer makes an appointment to come back later. If the chosen person refuses to be interviewed at all, the interviewer does not interview anyone else in that house, but proceeds to the next house allocated to her/him.

B.5 NON-PROBABILITY SAMPLING

Respondent-driven sampling or snowball sampling is a non-probability sampling method used when the whole population or sampling frame is unknown. As the name suggests, invoking the image of a snowball rolling down the hill, snowball sampling starts with one or two persons who are known in a population and then accumulates others as it develops. The initial participants contact others they know who meet study eligibility criteria and invite those persons to participate. Those new persons then invite others to participate, etc. This method is especially useful for hidden populations, where social networks are known to those in the network, but not to the outside world.

Some examples of hidden populations include SWs, MSM, or IDUs. Use of non-probability sampling would be appropriate in the following examples:

- CSWs in Vietnam may know each other, but there may not be a roster of CSWs as it is officially an illegal activity.
- Accessing populations of lesbian, gay, bisexual, transgender (LGBT) communities may be difficult due to stigma. Accessing these populations through social networks and social scenes (bars, etc.) may allow them to be reached more easily.

In these hidden populations, snowball sampling could be utilized, including the following steps:

1. Determine the programs in which the population of interest would participate
2. Approach stakeholders or members of the program for contacts
3. Approach the contact and ask their participation; these contacts will act as seeds (Heckathron, 1997).
4. The seeds hand out participation invitation slips to their contacts that meet eligibility criteria; seeds may receive an incentive for successful invitations for new contacts. New contact members who participate will now invite their contacts, and so on.

Snowball sampling continues until the population is saturated or until the sample size is met. It is important to recognize that snowball sampling or respondent driven sampling will produce biased results, as the population is not randomly selected.

It is important to ensure that the points of access to the hidden populations are not only the standard or assumed locales (i.e., MSM in bars, SW in brothels, etc.). Using the in-country program staff to understand the social networks, support structures and community groups that are used by persons in these hidden populations will provide more reliable information on prevalence in the community.

C. SURVEY PROCESS

C.1 HIRING AND TRAINING INTERVIEWERS AND SUPERVISORS

If the interviewers trained for the qualitative assessment/instrument development are not engaged in this phase, it would be useful to engage community leaders and stakeholders in thinking through who would be appropriate candidates for interviewers**.** If interviewers and supervisors are different from those in the pilot and validity study, then both should receive the same interview training and orientation to the instrument.

These roles may be filled by staff from the implementing organization, by outside hires, or a combination of both. Staff may be used if there is an interest in building capacity (particularly if future studies are anticipated) and/or in order to save costs. Usually, however, there is not sufficient staff to cover all positions, so interviewers and supervisors are typically a mix of staff and outside hires. It is important to ensure that the selected interviewers are acceptable to the population being interviewed, particularly when the target population, such as individuals or groups generally stigmatized in their communities, may be mistrustful. In past studies we have responded to this sensitivity by using former drug users as interviewers in a study of HIV-related behaviors related to current drug use. Though not all of the interviewers in that study were former drug users, all had experience working with that population.

Potential interviewers and supervisors must be informed of the importance of working every day during the survey. Once hired, they will be expected to prioritize the training and survey over other work (an issue that often comes up when interviewers come from the implementing organization and are pulled in many different directions). However, it is understood that emergencies and/or unexpected but important events can occur that can oblige those involved to miss a day or more. Under such circumstances, an interviewer can leave briefly and return to the survey as soon as possible. In addition, potential interviewers should arrange childcare, as it is inappropriate to have infants and children present in the interview.

For a given site, between ten and twenty interviewers and half that number of supervisors are needed to test the instrument. Each supervisor has responsibility for two interviewers. If the study is to be done in multiple sites with different characteristics (i.e., urban vs. rural; different

languages; etc.), it will be necessary to multiply the number of interviewers and supervisors by the number of different types of sites to ensure that a full representation of data is gathered.

If available, the qualitative study interviewers can be used since they have prior training in general interviewing techniques. Supervisors and interviewers are trained together. Training consists of 2-3 days of didactic training including standardized interviewing methods and procedures as well as specific orientation and practice with the draft instrument among themselves.

C.1.1 INTERVIEWER QUALIFICATIONS

Each interviewer should have the following qualifications:

- ✓ Fluent and literate in the language(s) of the local population where the study will be conducted
- ✓ Able to commit to the study timeline (full time for approx. 12 working days)
- ✓ In good health and able to walk long distances if needed
- ✓ Acceptable to the target population (in terms of reputation, where they are from, gender, age, ethnicity, etc.)

While the interviewers need to be acceptable to the sample being interviewed, they should not be personally known by the interviewees. This is important because the goal is for the interviewees to provide all the information they can on the topics being studied. If they know the interviewer they may not feel comfortable answering certain questions.

C.1.2 SUPERVISOR QUALIFICATIONS

Each supervisor should have the following qualifications:

- ✓ Fluent and literate in language(s) of the local population where the study will be conducted
- ✓ Fluent and literate in language of the program/study director (in order to act as a liaison between study director and interviewers where they do not share a common language)
- ✓ Able to commit to the study timeline (full time for approx. 12 working days, plus any additional preparation time)
- ✓ In good health and able to walk long distances where needed
- ✓ Acceptable to local people (in terms of reputation, where they are from, gender, age, ethnicity, etc.)

Supervisors provide the link between the study director and the interviewers. Like the interviewers, they need not have interviewing experience although prior experience working on a study of any type is helpful. As a supervisor, they may need to conduct some interviews and/or sit in when an interviewer becomes unavailable or additional supervision is needed, as well as re-interview some respondents to assess reliability of findings. Thus, they share the same qualification requirements as the interviewers and undergo the same training, with the additional requirement that they are able to communicate verbally with both the interviewers and the study director.

C.1.3 STUDY DIRECTOR QUALIFICATIONS

The study director should have the following qualifications:

- ✓ Preferably team leader or manager for the implementing organization, or someone with similar experience
- ✓ Available to direct pre study planning and activities and post study use and sharing of data
- ✓ Available for duration of study itself (full time for approximately 12 working days) plus additional part time for study preparation over the preceding month or two)
- ✓ Can also be the trainer or otherwise speaks the same language as the trainer and the supervisors (and interviewers if possible)

C.2. ROLE OF INTERVIEWERS

- Correctly identify respondents by random selection.
- Obtain informed consent prior to each interview.
- Interview respondents according to the standard procedures and principles outlined in the training.
- Maintain confidentiality of participant's HIV status.
- Keep an ongoing record of each house/participant assigned to them, and whether the interview for that house/participant had been completed, not yet completed, refused, or no eligible person lives in the house.
- Record and report progress and any problems to the supervisor. If, during the interview, it is apparent that the person is (or becomes) severely distressed, this is reported to the supervisor.

C.3 ROLE OF SUPERVISORS

- Monitor the performance of the interviewers and provide assistance and feedback. This includes each day observing one interview by each interviewer to maintain quality and correct mistakes. It also includes answering questions and addressing problems raised by the interviewers. If there are problems that the supervisor cannot address, these are passed on to the program director for resolution.

- At the end of each day collect all completed interviews and review them for completeness, correctness, and clarity.

Definition Box:
Completeness: Refers to recording responses to all the questions.
Correctness: Refers to checking that recorded answers are consistent with each other and the program design.

Clarity: Refers to records responses clearly.

Below is a list of common data errors on data collection forms that supervisors should review:

- Dates are inconsistently entered as DD-MM-YY or MM-DD-YY
- Respondent IDs are not unique (i.e., two different participants have the same ID as each other)
- Missing data is noted incorrectly according to the protocol. Some standards of data collection are placing one of the following to note missing data: **., 99, or -.**
- Missing respondent ID or interviewer ID numbers

- Review each interviewer's records of completed interviews, appointments and refusals. In all cases in which an interviewer failed to obtain an interview at a selected house/participant (refusal, the house is abandoned, or no adults live there), the supervisor visits the house/participant to confirm the problem.

WHY?

There are several checks that are done to prevent fraud by the interviewers: 1) the supervisors return to the homes of refusers and non-respondents to confirm their non-participation; and 2) the supervisor may repeat some of the interviews to confirm the results. The latter is normally done when there is suspicion that an interviewer is entering fake data.

NOTE:

It is important to try to complete the survey as quickly as possible. Where data gathering extends longer than one week, significant reactivity begins to emerge. This refers to local discussion and rumors based on the experiences of those who have already been interviewed, and which can affect the responses of those not yet interviewed. The most efficient way to complete the survey will vary, based on transportation availability

Example: Survey Implementation Timeline

One approach, which we have used successfully, is to allocate three settlements to each interviewing team (a supervisor and two interviewers). On the first day of the survey, each team proceeds to one of the settlements, enumerates the houses in the chosen sector(s), and then begins interviewing. Interviewers try to contact all of the 20 chosen houses within the sector(s), either completing interviews, making appointments to come back (usually the next day) or recording refusals. The following day the team proceeds to the next settlement and enumerates all the houses in the chosen sector(s). One interviewer then begins to work in that sector while the supervisor takes the second interviewer back to the settlement they visited the day before. Here the second interviewer follows up on all the appointments made the previous day, while the supervisor checks on all refusals and repeats some of the interviews. When completed, they return to the other settlement and assist the other interviewer. This process is repeated daily and teams can usually complete interviewing in all their assigned settlements in five working days. In cases where appointments for interviews cannot be made for the following day, the supervisor usually returns to the settlement in question alone on a later day and does the interviews her/himself.

C.4 INTERVIEWING ON SENSITIVE TOPICS

Interviewing persons about sensitive topics such as mental health or behaviors related to HIV may cause discomfort and make it difficult for people to agree to participate in the survey. There are several techniques that can be used, individually or in combination, to improve the uptake and completeness of the survey:

1. Interviews should be completed in a private space where confidentiality can be maintained.

2. Interviews, especially in snowball sampling populations, should be conducted in a neutral place. For example, surveys of CSWs should not be conducted near a police station or in a government office.

3. It may be possible to have self-administered surveys in populations with sufficient literacy.

4. Reassuring the participants at the beginning in the interview that answers will be confidential, that they may refuse to answer individual questions, and that they may stop the interview at any time may improve confidence and cooperation.

5. Utilizing interviewers who are peers or members of the same community as the target population (but are not personally known to the respondent) can make participants more likely to participate and answer survey questions. It is essential that the target

populations trusts the study interviewers; such trust can be established by having community members choose the interviewers.

C.5 OBTAINING INFORMATION ABOUT HIV-RELATED BEHAVIORS AND OTHER SENSITIVE TOPICS

Although this prevalence study is examining mental health, understanding how HIV-related risk factors are associated with mental health outcomes is important. Therefore, surveys often include questions on risk factors including concurrency of sexual partners, sexual contact with high-risk populations (CSWs, etc.), intercourse without a condom, and sharing needles for injecting drug use. For a list of survey topics, refer to the UNAIDS document, Guidelines for measuring national HIV prevalence in population-based surveys, pgs. 20-21. An additional resource for survey topics and questions of HIV-related risk behavior are the Demographic Health Surveys conducted by Measure DHS. There are several surveys inquiring about HIV-related behavior that can assist researchers in developing HIV behavior surveys, such as the *HIV Prevalence Survey* and the *HIV/AIDS Knowledge, Attitudes and Behavior Survey* (Demographic Health Surveys, AIDS Indicator Surveys).

Obtaining sensitive information about HIV-related risk behaviors can be made easier by:

- Building rapport with interviewer
- Ensuring that interviewers are culturally appropriate, for example:
    - Having peers interview peers (CSW interview CSW)
    - Ensuring that the age of interviewers is appropriate to age of interviewees
    - Considering age and gender power dynamics
    - Consider as interviewers PLWHA who are open about their status
- Ensuring interviewers are trained to maintain confidentiality
- Asking the most sensitive questions towards the end of the interview
- Pilot testing the survey (Module 2) with appropriate representatives of the target population

As can be seen in the DHS surveys or the UNAIDS surveys mentioned above, it is recommended that the study interview starts with simple, non-sensitive information on demographics, such as marital status, education, occupation, etc. After rapport is built between interviewer and interviewee, then questions about sexual behavior and HIV-related risk factors can be asked. Questions must be asked in an open, non-judgmental manner. Non-verbal cues from the interviewer must be observed during the role-play and practice of the survey as well as the survey piloting. Asking the same question in different ways allows researchers to check for consistency of responses from the interviewee.

To gain information about sensitive topics, allowing the participant to speak and using a list on the survey to check off the answers provided can give the study staff more accurate information than specifically requiring answers to every question.

For example, instead of asking,
Tell me yes or no: can HIV be transmitted by the following: (Interviewer reads EACH answer out loud and waits for a yes/no from the interviewee)

- *Hugging*
- *Sharing razors*
- *Sharing needles for drug use*
- *Having sex without a condom*
- *Shaking hands or greeting each other*

Ask:
Tell me all the ways you know of that HIV is transmitted: (Interviewer reads to here and then waits for interviewee to respond with routes of transmission. Interviewer circles each of the answers below that fit what the interviewee says).

- *Hugging*
- *Sharing razors*
- *Sharing needles for drug use*
- *Having sex without a condom*
- *Shaking hands or greeting each other*

Using methods such as these will allow the participant to answer more naturally. Asking about sensitive information can be difficult but can be made easier by appropriate training of interviewers on how to generate rapport between interviewer and interviewee. This includes approaches like:

- Chatting first about mundane topics prior to the interview until the interviewee appears relaxed.
- Showing interest in the interviewee and their responses
- Not showing surprise, approval, or disapproval of responses but projecting an air of acceptance.
- Treating the person with friendliness and respect.

D. DATA ENTRY AND ANALYSIS

As with all studies, data collection leads to the necessary steps of data entry, management and analysis. It is essential that data collection is completed with confidence by the data collectors and supervisors. A common theme in data collection, entry, and analysis, is GIGO (Garbage In, Garbage Out). No analytic tool or cleaning process can fix problems caused by a flawed data collection process.

There are many computer packages available for sale and for free for data entry, management and analysis. Free packages include:

- EpiInfo (www.cdc.gov/epiinfo)
- CSPro (http://www.census.gov/population/international/software/cspro/)
- EpiData (http://www.epidata.dk/)

Packages with a cost associate include:

- Microsoft Access
- FileMaker Pro
- SAS (www.sas.org)
- SPSS (http://www-01.ibm.com/software/analytics/spss/products/)
- STATA (www.stata.com)

Computer packages should be chosen based on the skill set among the study team and the desired complexity of analysis to be completed. This section provides an overview of the types of analyses that are typically conducted with this type of data. The use of these techniques requires persons skilled in data analysis beyond the discussion presented here.

D.1 DATA CLEANING AND ENTRY

As data is collected and reviewed by the data collector and data supervisor, data entry must be completed. As mentioned above, there are several data entry packages. For this module's examples, we use Microsoft Excel as the data entry platform. The principles described below apply generally to data entry and cleaning but how these principles are operationalized will vary by data software and study design.

Using Excel, a basic spreadsheet program, a simple worksheet should be set up. Each column represents one question or variable. Each row is dedicated to a single interview. Using the questionnaire shown in Appendix D Part A as an example (the questionnaire is taken from Module 2 on instrument development), below is a sample of a spreadsheet in Excel with two rows of made-up data to show how data should be entered:

Row	Respondent ID	Interviewer ID	Date of Interview	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
1	123	456	2-Jan-11	0	1	2	3	0	1	2	3	0	1	2	3	0	1
2	987	654	3-Jan-11	3	2	1	0	3	2	1	0	3	2	1	0	3	2
3																	
4																	
5																	
6																	
7																	
8																	
9																	

A simple list of directions for all data entry personnel should be created. The instructions provide guidance on how data should be entered consistently in the spreadsheet. For example, the instruction list may state the following for the spreadsheet above:

1. Each row is for one individual's interview
2. Respondent ID and Interviewer ID should each be at least three digits
3. Respondent IDs should not be repeated
4. All calendar dates should all be entered as day-month-year with a 3-letter code for the month (DD-MMM-YY), i.e., 2-Jan-11
5. Each mental health question (A1-A14) should have a number 0 to 3 entered in the row for that patient corresponding to the response on the data collection tool. If there is missing data, then the column should be left blank for that person.
6. Each number related to the mental health question is a code. (Not at all=0, A little bit=1, Quite a bit=2, and Extremely=3)
7. After entering the data, review the row to make sure the computer records matches the paper record

It is possible to include different questionnaires and demographic data in one spreadsheet **as long as there are different column names for each question**. It is essential that each row of information represent the same interviewee as connected by the respondent ID. If the questionnaire in Appendix D Part B was added to the same spreadsheet as above, the excel sheet would look like the table below:

Row	Respondent ID	Interviewer ID	Date of Interview	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	B01	B02	B03	B04	B05	B06	B07	B08	B09
1	123	456	2-Jan-11	0	1	2	3	0	1	2	3	0	1	2	3	0	1	0	1	2	3	4	0	1	2	3
2	987	654	3-Jan-11	3	2	1	0	3	2	1	0	3	2	1	0	3	2	4	2	3	1	4	3	2	1	0
3																										
4																										
5																										
6																										
7																										
8																										
9																										

Data cleaning consists of range and logic checks. In Excel, it is simple to review the data by visual inspection. Range checks examine the data for allowable responses. For example, question A1 above should not have an answer of 5 or 6; the allowable responses for question A1 are 0 to 3. Logic checks examine data to determine if the appropriate skip patterns were met. For example, Appendix D Part B has a survey that is gender-specific. A logic check will make sure that the appropriate version of Part B was filled out specific to the gender of the participant.

Using the tools built into Excel, data cleaning can be made easier. Some of the tools that should be considered are:

- Conditional formatting: set up range checks in the cells so that out of range values are colored.
- Formulas: =MAX(), =MIN() at the bottom of the column will inform you which values are the maximum or minimum, in other words, a simple range check.
- Logic checks: =IF() formulas allow the user to know if a logic check is met.

In the example table below, there is missing data, data outside the maximum and below the minimum ranges:

Respondent ID	Interviewer ID	Date of Interview	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	B01	B02	B03	B04	B05	B06	B07	B08	B09
123	456	2-Jan-11	0	1	2	3	0	1	2	3	0	1	2	3	0	1	0	1	2	3	4	0	1	2	3
456	789	3-Jan-11	1	3	4	0	1	2	3	4	1	2	3	0	2	0	1	2		4	5	1	2	3	4
789	102	4-Jan-11	2	2	4	1	4	2	4	0	1		3	2	1	1	4	2	0	1	2	3	1	0	4

The table below demonstrates where some of the data errors are contained, noted by the highlighting in red. The data errors include an answer of 4 on question A3 when only 0 to 3 is allowable; there is missing data in question A10 and B03. Question B05 has an answer outside the allowable range. These can all be caught with range and logic checks.

Respondent ID	Interviewer ID	Date of Interview	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	B01	B02	B03	B04	B05	B06	B07	B08	B09
123	456	2-Jan-11	0	1	2	3	0	1	2	3	0	1	2	3	0	1	0	1	2	3	4	0	1	2	3
456	789	3-Jan-11	1	3	4	0	1	2	3	4	1	2	3	0	2	0	1	2		4	5	1	2	3	4
789	102	4-Jan-11	2	2	4	1	4	2	4	0	1		3	2	1	1	4	2	0	1	2	3	1	0	4

For additional sources to utilize Excel for data cleaning, see pages 42-56 in the book entitled, Excel Data Analysis: Your visual blueprint[tm] for creating and analyzing data, charts and PivotTables, by Jinjer Simon (Wiley Publishing, 2005).

Finally, the time required by each staff to conduct accurate data entry and cleaning varies. It is important to allow sufficient time for each staff to do these tasks correctly.

D.2 TYPES OF DATA

There are several types of data and the type of data determines the kind of analysis that can be conducted. The main distinctions are between categorical, ordinal, and Interval data:

Categorical or Nominal data is data that consists of names and are not quantitative. For example, demographic questions on sex (male or female) and where a person lives are nominal data, while age is quantitative and therefore is not nominal. Although it is not numeric, categorical data can be summarized numerically, using frequencies and cross-tabulations (see below).

Ordinal data is also categorical data except that there is an order (i.e., one response is considered greater or lesser than another). But the intervals between the responses vary and may not be known. Therefore the data is not quantitative, even though numbers may be applied to different responses to represent their order. Most of the data in the mental health surveys presented here is ordinal. For example, participants could only answer: Not at all (0), A little (1), Quite a bit (2), and Extremely (3) to questions in the Appendix D Part 2 survey. While there is a definite order here it is not clear that the difference between a little and quite a bit is the same as the difference between quite a bit and extremely. Ordinal data can be summarized and analyzed using frequencies, cross-tabulations and can even be averaged (based on the numeric codes, although this is understood to have a more subjective meaning than averages of quantitative data).

Quantitative data is numerical data. It can be discrete, in the forms of counts with equal distances between each value and no 'in-between' values, such as the number of children (there cannot be 2.5 children for example. Or it can be continuous which means that any amount is possible: for example, it is possible to be 24.45 years of age or any other number. Measurement data such as height, weight, and CD4 count are all examples of quantitative data. Means or averages, standard deviations, and medians are some of the statistics that can be used to describe quantitative data.

D.3 FREQUENCIES AND CROSS-TABULATIONS

Data analysis can be completed using a variety of computer packages. Below we present the concepts of data analysis using frequencies and cross-tabulations in SPSS; however, many programs work just as well. Refer to the user's manual of the data analysis package being used to determine the step-by-step process. For most data packages, the set-up created in Excel for data entry as outlined above will easily import into the data analysis software of choice.

D.3.1 FREQUENCY TABLES

Once data has been entered and cleaned, the first step in basic data analysis is calculating frequencies. *Frequencies* present the number of times an answer is provided by the participants surveyed for categorical variables. Continuous variables, such as age or CD4 count, could be presented with a mean, standard deviation, and range. For example, for question B09 in the table above, two out of the three people interviewed reported a "4" or that they "Often cannot provide community care." A basic frequency table can be built from the data entered into excel with a format similar to the following:

Question	Reponses	N	n	Percentage
A1: Feeling low	Not at all (0)	3	1	33.3%

in energy, slowed down	A little (1)	3	1	33.3%
	Quite a bit (2)	3	1	33.3%
	Extremely (3)	3	0	0.0%
A6: Feeling hopeless about the future	Not at all (0)	3	1	33.3%
	A little (1)	3	2	66.7%
	Quite a bit (2)	3	0	0.0%
	Extremely (3)	3	0	0.0%
B09: Community Care Tasks	Cannot say (blank)	3	0	0.0%
	No more (0)	3	0	0.0%
	A little more (1)	3	0	0.0%
	Some amount more (2)	3	0	0.0%
	A lot more (3)	3	1	33.3%
	Often cannot (4)	3	2	66.7%

Note that the basic frequency table of categorical data presents: (1) the options to each question, (2) the number of people who answered the question overall (N), (3) the number of persons who answered a particular option (n), and (4) the percentage of people who responded a certain way. The percentage is calculated as percentage=n/N*100. To check that all data is included in the table, one can add up the percentages for each question. The total percentage for each question should be 100%.

In SPSS, frequencies are easily built using the “Analyze” tab in the program and clicking “Descriptive Statistics” followed by “Frequency”. Using a point and click system, the user can tell SPSS which variables should be shown in a frequency table. For additional information on using SPSS, refer to the user’s manual appropriate to the version of SPSS being used. The SPSS Tutorials and the “Help” menu in the SPSS program also offer step-by-step instructions to create frequency tables.

Typically, the first data analysis presents the response frequencies for each question in the study instrument. Frequency tables help assess the cleanliness of the data. They allow the user to see whether all the data were within range, the amount of variation in the responses, and whether the patterns of responses are similar to what was expected. More data cleaning will be potentially needed after examining the first draft of the frequency tables. Once frequency tables are finalized, the next step in the process is cross-tabulation.

D.3.2 CROSS-TABULATION

Cross-tabulations are tables that look like basic frequency tables, but involve a comparison between two or more groups. Cross-tabulations should be used to compare groups of interest or to determine whether certain risk factors have an impact on the mental health responses. A

risk factor, sometimes called a determinant, is defined as a variable or characteristic associated with an outcome. Here the outcome is a mental health problem. In HIV-infected populations, potential HIV-specific risk factors for mental health outcomes may include length of knowledge of HIV status, access to and uptake of ART, adherence to ART, disclosure of HIV status to family or friends, or attendance at a support group. Additional risk factors that are not HIV-related might be education, gender, family make-up, marital status, and experience of trauma. It is important that risk factors of interest are included in the data collection.

In our example, if we are interested in the impact of gender on the responses to questions A1, A6 and B09, we should look at the cross tabulation of gender compared to the questions. The tables would now look like this:

		Females			Males		
Question	**Reponses**	**N**	**n**	**Percentage**	**N**	**n**	**Percentage**
A1: Feeling low in energy, slowed down	Not at all (0)	50	10	20.0%	50	5	10.0%
	A little (1)	50	25	50.0%	50	20	40.0%
	Quite a bit (2)	50	5	10.0%	50	30	60.0%
	Extremely (3)	50	10	20.0%	50	5	10.0%
A6: Feeling hopeless about the future	Not at all (0)	50	3	6.0%	50	10	20.0%
	A little (1)	50	17	34.0%	50	5	10.0%
	Quite a bit (2)	50	15	30.0%	50	10	20.0%
	Extremely (3)	50	15	30.0%	50	25	50.0%
B09: Community Care Tasks	Cannot say (blank)	50	0	0.0%	50	10	20.0%
	No more (0)	50	4	8.0%	50	5	10.0%
	A little more (1)	50	5	10.0%	50	10	20.0%
	Some amount more (2)	50	1	2.0%	50	12	24.0%
	A lot more (3)	50	20	40.0%	50	13	26.0%
	Often cannot (4)	50	20	40.0%	50	10	20.0%

Using a cross-tabulation, study staff will be able to compare two groups and their mental health outcomes. Looking at the table above, we see that 20% of women reported feeling extremely low in energy (Question A1), but only 10% of men reported the same thing. However, in question A6, 50% of men reported feeling extremely hopeless about the future, but only 30% of women reported feeling extremely hopeless about the future.

In SPSS, again, using the “Analyze” tab, click “Descriptive Statistics” followed by “Crosstab”. The” Crosstab” requests the information for the row and the column. The Row variable should be the risk factor of interest and the Column should be the outcome of interest. Percentages can be presented (using the Cells button) and statistics of chi-square analysis can be used to compare the two groups (using the Statistics button). A chi-square test is used to compare categorical data and a p-value is produced which describes the likelihood that any differences are statistically significant.

For quantitative variables, such as age or total mental health score, we can compare the mean, standard deviations, or medians to each other between the groups. A t-test can be used to compare the means and standard deviations between the two groups. A p-value provides a measure of statistical significance.

Frequency tables create the base for the analysis and additional analyses can be performed using the next section’s steps.

D.4 PREVALENCE AND RISK FACTORS

Prevalence of an outcome or mental health condition is defined as the number of people with a condition divided by the number of people in the population. It is presented as a proportion or percent.

Utilizing the basic frequency tables, a prevalence, or percentage, of those with a mental health problem can be presented. If it is observed that 20 out of 100 HIV-infected participants in a survey were diagnosed with a mental health problem, we would calculate that the prevalence of the mental health problem was 20% (20/100*100%). In the text of the report, the following statement could be made:

> *Of those HIV-infected participants, prevalence of the mental health problem under investigation was 20%.*

Note that this is overall prevalence. When examining and reporting the cross-tabulations, the statements must be modified to present the correct sub-group the statistic is representing. In the cross-tabulation above, we would report:

> *Of HIV-infected male participants, prevalence of feeling extremely hopeless about the future was 50%. Of HIV-infected female participants, prevalence of the same mental health problem was 30%*.

If a chi-square test has been run on the cross-tabulation of males versus females with respect to feeling extremely hopeless, and the p-value was less than 0.05, the text could be modified to the following:

Of HIV-infected male participants, prevalence of feeling extremely hopeless about the future was 50%. Of HIV-infected female participants, prevalence of the same mental health problem was 30%. There was a statistically significant difference between men and women with respect to feeling extremely hopeless about the future ($p<0.05$).

Chi-square tests that produce a statistically significant result, indicated by a p-value of less than 0.05, should be reported as part of the final prevalence and risk factors report. In addition, those risk factors that were expected to have an impact on the mental health outcome, but did not have a statistically significant impact, should be noted. There may be important findings to be discussed with the community if the hypotheses of researchers are quite different than outcomes observed.

Additional analyses, including logistic regression models, may be utilized to determine the impact of various risk factors on mental health outcomes.

E. DISSEMINATION OF RESULTS

E.1 EXIT MEETING

Soon after data collection and entry is finished (before the director leaves the area) a preliminary analysis is done of the prevalence of the psychosocial issue/problem(s). Local leaders are invited to a meeting along with local staff, other interested parties, and the interviewing teams. The director presents the results to the assembly and leads a discussion of future directions.

A note-taker should be present at the dissemination meetings, as understandings of relationships between risk factors and mental health outcomes may be highlighted. These community insights and the responses of the researchers should be included in the final report.

E.2 FINAL REPORT

After the full analysis has been done, the director and others complete a report of the activities and results. The report should be in hard and soft copy depending on computer accessibility and literacy in the region. The report should be disseminated to the community, local staff, the wider organization, and other interested parties.

REFERENCES

Bennett, S., Woods, T., Liyanage, W.M., Smith, D.L. (1991). A simplified general method for cluster-sample surveys of health in developing countries. *World Health Stat Q.,44,* 98-106.

Bennett, S., Radalowicz, A., Vella, V., & Tomkins, A. (1994). A computer simulation of household sampling schemes for health surveys in developing countries**.** *International Journal of Epidemiology, 23,* 1282-91.

Brown, V., Moren, A., & Paquet, C. (1999). *Rapid Health Assessment of Refugee or Displaced Populations*. Paris: Medicins Sans Frontieres.

Demographic Health Surveys, AIDS Indicator Surveys. (2011). http://www.measuredhs.com/aboutsurveys/ais/start.cfm

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

Heckathron, D.D. (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems, 44(2),* 174-199. Retrieved from: http://www.respondentdrivensampling.org/reports/RDS1.pdf

Kaiser, R., Spiegel, P.B., Henderson, A.K., & Gerber, M.L. (2003). The application of geographic information systems and global positioning systems in humanitarian emergencies: lessons learned, programme implications and future research. *Disasters, 27,* 127-40.

Katz, J. & Zeger, S.L. (1994). Estimation of design effects in cluster surveys. *Annals of Epidemiology, 4,* 295-301.

Simon, J. (2005). *Excel Data Analysis: Your visual blueprint*^{tm} *for creating and analyzing data, charts and PivotTables.* New York: Wiley Publishing.

For all Woodruff references, please include both of the items below:

1. Bradley Woodruff "Review of Survey Methodology," (Powerpoint presentation), CDC, Inter- national Emergency and Refugee Health Branch.
2. Center for International Emergency, Disaster and Refugee Studies (CIEDRS), and the Hopkins Population Center, (2003). *Demographic Methods In Emergency Assessment: A Guide For Practitioners*. Johns Hopkins University Bloomberg School of Public Health Baltimore, Maryland.

http://www.humanitarianinfo.org/imtoolbox/05_Assessments/Reference_Resource_Documents/2003_Demographic_Methods_In_Emergency_Assessment_CIEDRS.pdf

UNAIDS. *Guidelines for measuring national HIV prevalence in population-based surveys*. Retrieved from: http://www.who.int/hiv/pub/surveillance/guidelines measuringpopulation.pdf

APPENDIX A: EXAMPLE SCHEDULE

The following schedule of working days is proposed in the table below. It may be adapted as needed. This schedule **assumes** that Module 2 ends on Day 16 of on-site activities.

PROPOSED SCHEDULE FOR MODULE 3	
Day	**Activities**
17	Randomly select sampling sites where the study interviews will be conducted. Begin process of mapping the villages and towns in which the sampling sites are located.
18	Complete mapping of villages and towns with sampling sites. Divide villages and towns into sectors. Select 30 sectors for interviews (20 interviews will be done in each sector).
19-21	Conduct survey. Begin data entry as soon questionnaires begin to arrive at the training site.
22-23	Complete data entry and cleaning
24	Data analysis
25	Draft Report
26	Exit Meeting
27-30	Complete and disseminate the Final Report

APPENDIX B: RESOURCES

TIME

Preparation usually requires one month on a part-time basis, including:

- Preparations for the assessment
- Meeting with the community
- Preparing logistics
- Arranging personnel
- Adapting and testing the instruments

PERSONNEL

One program director/trainer and 10 supervisors with the following qualifications:

- NGO or MOH local staff or students (e.g., nursing)
- At least a high school education (Preferred = college level education and good knowledge of English)
- Available for duration of study
- Ability to read and write in the local language
- Willing and able to ride a bicycle or motorbike, if needed and available

20 additional interviewers:

- Qualifications are same as above, including:
    - Less important that they be NGO or MOH local staff
    - Capable of walking long distances if needed
    - Understand they will be supervised by the above interviewers

TRANSPORT

The program director and all staff will meet at the training site at the beginning of each day before proceeding to the study site. Therefore, daily transport between the training and study site for all workers (interviewers and supervisors) is required throughout all phases of the program.

NOTE:

During the study, each supervisor will require a bicycle or motorbike or similar individual transportation. These are necessary to supervise both interviewers while also returning to sites visited on previous days to repeat interviews and check on refusals. The lack of such independent transport for each supervisor will considerably slow down the assessment process, particularly during the survey, so they are well worth the trouble to arrange.

TRAINING AND OFFICE MATERIALS

A combined training and office site is needed to complete all modules that is:

- ✓ Large enough to seat all workers
- ✓ Quiet
- ✓ Has power (electric or generator)
- ✓ Available exclusively for use by the team throughout the program
- ✓ Able to be locked at night

It should have the following:

- ✓ Blackboard/whiteboard/easel and chalk/marker pens
- ✓ Overhead or LCD projector
- ✓ Printer and two reams of paper
- ✓ Reliable photocopier capable of printing thousands of pages a day when required (this includes printing copies of the instrument)
- ✓ Three toner cartridges
- ✓ 20 reams of paper
- ✓ Laptop computer with word processing and statistical programs

INTERVIEWER MATERIALS

For completing all modules, each interviewer and supervisor should have the following:

- ✓ Exercise book
- ✓ Piece of chalk if being used
- ✓ Two pencils, eraser and sharpener
- ✓ Waterproof carrying bag
- ✓ Daily portable lunch
- ✓ Copy of a random number table if being used
- ✓ Sufficient copies of instruments, consent forms, and non-verbal response forms
- ✓ Cell phones or other method of communication with study director and supervisors

APPENDIX C: VERBAL CONSENT FORM

VERBAL CONSENT FORM FOR RESEARCH STUDY.

Instructions for the Interviewer:

The following sections written in italics are to be read to the subject prior to the interview. If the subject then agrees to participate, you must sign on the line marked 'Witness to Consent Procedures', at the end of this form. Also mark the date on the appropriate line.

Purpose of the Study

You are being asked to be part of a research study. We want to find out about the problems affecting people in this area. This research is being done by (your organization).

Procedures

To obtain this information we are talking with some people in the community who we selected by chance/because we heard that you are knowledgeable about this topic. This is how we selected you. If you agree to help us, I will ask you some questions. These questions are about your health. We may also want to return and talk with you again later.

Risks and Discomfort

Each interview will take about __ __ minutes. It is possible that some questions may upset you. You may refuse to answer these questions, or any questions, if you wish. You may stop the interview at any time.

Benefits

This information will help (your organization) to provide better programs to improve the health of the people in this area. However, there may be no direct benefit to you personally.

Confidentiality

During the interview I will write down the information you tell me. This is the information we will use for our study. The record of this information will not have any information which can be used to identify you. I will also record your name and address, but this will be stored separately from the record, and will be locked in the program director's office. Only the program director will have the key. Only he will be able to see this information. Every effort will be made to protect the confidentiality of this information as far as is legally possible.

Voluntariness

It is your decision whether or not to be in this study. You can stop being in this study at any time. This will not affect any assistance you get from (your organization) or any other organization.

Whom to Contact

If you have any questions you can ask (program director). He/she is in charge of the study and be contacted through the (your organization) Office in _____, telephone _____. In the future if you have any questions about the study, you should ask (local contact for your organization). He/she and the other researchers will tell you if they learn anything new that they think will affect you.

Do you have any questions?

Do you agree to participate in this study? **Yes** (PROCEED) **No** (STOP)

_____ _____

Signature of Interviewee Date

[I have explained this research study to the subject.]

_____ _____

Interviewer or other witness to consent. Date

(to be signed only if subject has verbally consented).

_____

Signature of Investigator

APPENDIX D: EXAMPLE OF DRAFT ASSESSMENT INSTRUMENT FOR TESTING (KURDISTAN, IRAQ)

Date / /

Site

Interviewer

Age of Interviewee

Gender of Interviewee M F

Name of Interviewee

Section A: Functionality Assessment

There are two versions of this section - one for men and one for women. If the interviewee is a man, use only the version for men. If the interviewee is a woman, use only the version for women.

I am going to read a list of activities and duties. These are tasks and duties that other torture survivors have told us are important to them. For each one I am going to ask you how much more difficulty you are having doing it THAN MOST OTHER MEN/WOMEN OF YOUR AGE. You should tell me whether you are having no more difficulty, a little more, a moderate amount more, or a lot more, or you often cannot do that task.

Now say each task, and after each one say:

Are you having no more difficulty than most other men/women of your age, a little more, a moderate amount more, a lot more, or you often cannot do this task?

Section B: Symptom Assessment Instrument

Hopkins Symptom Checklist -25

Listed below are some problems that people sometimes have. For each one state how often you have felt like this in the last two weeks including today.

Symptom	Never or No	Sometimes	Often	Always
B01. Suddenly feeling scared for no reason	***0***	***1***	***2***	***3***
B02. Feeling fearful	***0***	***1***	***2***	***3***
B03. Feeling faintness	***0***	***1***	***2***	***3***
B04. Nervousness	***0***	***1***	***2***	***3***
B05. Heart pounding or racing	***0***	***1***	***2***	***3***
B06. Trembling	***0***	***1***	***2***	***3***
B07 and B34. Feeling tense	***0***	***1***	***2***	***3***
B08. Headaches	***0***	***1***	***2***	***3***
B09. Episodes of terror or panic	***0***	***1***	***2***	***3***
B10. Feeling restless, can't sit still	***0***	***1***	***2***	***3***
B11. Feeling low in energy, slowed down	***0***	***1***	***2***	***3***
B12 and B44. Blaming yourself for things	***0***	***1***	***2***	***3***
B13. Crying easily	***0***	***1***	***2***	***3***

Symptom	Never or No	Sometimes	Often	Always
B14. Loss of interest in sex or loss of sexual pleasure	***0***	***1***	***2***	***3***
B15. Poor appetite	***0***	***1***	***2***	***3***
B17, B39, and B46. Feeling hopeless about the future	***0***	***1***	***2***	***3***
B18. Feeling depressed	***0***	***1***	***2***	***3***
B19. Feeling lonely	***0***	***1***	***2***	***3***
B20. Thinking about ending your life	***0***	***1***	***2***	***3***
B21. Feeling not free or caught	***0***	***1***	***2***	***3***
B22. Worrying too much about things	***0***	***1***	***2***	***3***
B23 and B38. Loss of interest in things	***0***	***1***	***2***	***3***
B24. Feeling that everything you do is difficult	***0***	***1***	***2***	***3***

Harvard Trauma Questionnaire (Part IV – Symptoms)

The following are problems that people sometimes have after experiencing hurtful or terrifying events in their lives. For each problem please state how often you have had it in the last two weeks including today.

Symptom	Never or No	Sometimes	Often	Always
B26. Recurrent thoughts or memories of the hurtful or terrifying events	***0***	***1***	***2***	***3***
B27. Feeling as though these events were happening again	***0***	***1***	***2***	***3***
B28. Nightmares	***0***	***1***	***2***	***3***
B30. Able to feel emotions	***3***	***2***	***1***	***0***
B31. Feeling jumpy, easily startled	***0***	***1***	***2***	***3***
B32. Difficulty concentrating	***0***	***1***	***2***	***3***
B33 and B16. Trouble sleeping	***0***	***1***	***2***	***3***
B35. Feeling irritable or having outbursts of anger	***0***	***1***	***2***	***3***
B36. Avoiding activities that remind you of the traumatic or hurtful events	***0***	***1***	***2***	***3***
B37. Inability to remember parts of the traumatic or hurtful events	***0***	***1***	***2***	***3***
B40. Avoiding thoughts or feelings associated with the traumatic or hurtful experience	***0***	***1***	***2***	***3***
B41. Suddenly feeling very different emotionally or physically when reminded of the traumatic or hurtful events	***0***	***1***	***2***	***3***

Symptom	Never or No	Sometimes	Often	Always
B42. Feeling that people do not understand what happened to you	***0***	***1***	***2***	***3***
B43. Difficulty performing your work or daily tasks	***0***	***1***	***2***	***3***
B45. Feeling guilty for having survived	***0***	***1***	***2***	***3***
B47. Feeling ashamed of the hurtful or traumatic events that happened to you	***0***	***1***	***2***	***3***
B48. Spending time thinking about why these events happened to you	***0***	***1***	***2***	***3***
B49. Feeling as if you are going crazy	***0***	***1***	***2***	***3***
B50. Feeling that you are the only person who has suffered these events	***0***	***1***	***2***	***3***
B51. Feeling that others are hostile to you	***0***	***1***	***2***	***3***
B52, Feeling that you have no-one to rely on	***0***	***1***	***2***	***3***
B53. Finding out or told by other people that have done something that you cannot remember	***0***	***1***	***2***	***3***
B54. Feeling as if you are split into two people, one of you is watching what the other is doing	***0***	***1***	***2***	***3***
B55. Feeling betrayed	***0***	***1***	***2***	***3***

Traumatic Grief

Symptom	Never or No	Sometimes	Often	Always
B56. Hearing the voice of a deceased person speaking to you	***0***	***1***	***2***	***3***
B57. Seeing a deceased person standing in front of you	***0***	***1***	***2***	***3***
B59. Feeling that you have lost your sense of control	***0***	***1***	***2***	***3***
B60. Feeling that the death of someone close to you has changed your view of the world	***0***	***1***	***2***	***3***
B61. (Feeling that you are) having pain the same part of your body or having the same symptoms as people who have died.	***0***	***1***	***2***	***3***
B62. Feeling that moving on with your life (like making new friends, pursuing new interests) would be difficult.	***0***	***1***	***2***	***3***
B63. Feeling envious of others who have not lost someone close.	***0***	***1***	***2***	***3***
B64 and B29. Feeling like you have lost the ability to care about other people.	***0***	***1***	***2***	***3***
B65. Feeling drawn to places and things associated with people who have died.	***0***	***1***	***2***	***3***
B66.imitating some of the same behaviors or characteristics of people who have died.	***0***	***1***	***2***	***3***

Problems specific to Kurdish survivors of torture in Suleimaniyah governate and their families

Symptom	Never or No	Sometimes	Often	Always
B67. Feeling desperate	***0***	***1***	***2***	***3***
B68. Feeling mentally unstable	***0***	***1***	***2***	***3***
B69 and B25. Feeling inferior to others.	***0***	***1***	***2***	***3***
B70. Wishing you were dead	***0***	***1***	***2***	***3***
B71 and B58. Feeling as if you were already dead	***0***	***1***	***2***	***3***
B72. Waiting for your dead relatives to come back	***0***	***1***	***2***	***3***
B73. The brain is tired.	***0***	***1***	***2***	***3***
B74.Able to express your feelings	***3***	***2***	***1***	***0***
B75. Fighting with others	***0***	***1***	***2***	***3***
B76. Poor relationship with family members	***0***	***1***	***2***	***3***
B77. Able to enjoy feasts or other celebrations	***3***	***2***	***1***	***0***
B78. Drinking too much alcohol	***0***	***1***	***2***	***3***
B79. Thinking too much	***0***	***1***	***2***	***3***

Final Instructions

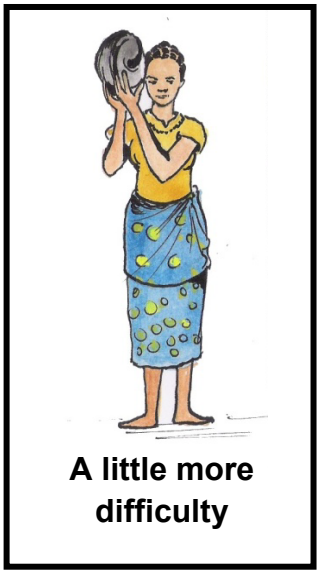
State that this is the end of the questionnaire. Review the questionnaire while still with the interviewee, checking that all questions have been answered and the answers are clear. If not, review the question with the interviewee and insert the missing information. Once you are satisfied that all questions have been answered and the responses are clear, ask the interviewee if it would be OK for you or your supervisor to return if you have further questions. If this is OK, ask if there are any restrictions on when you might return and record these in the space below. Finally, ask the interviewee not to discuss the interview with anyone until after the study is over. This is so that their comments do not affect the answers of others who are yet to be interviewed.

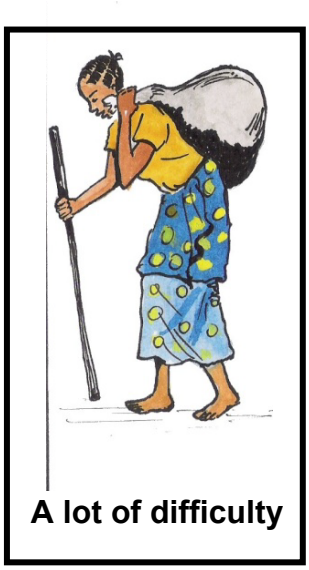
Restrictions on Re-visiting?

End the interview

APPENDIX E: EXAMPLES OF NON VERBAL RESPONSE CARDS

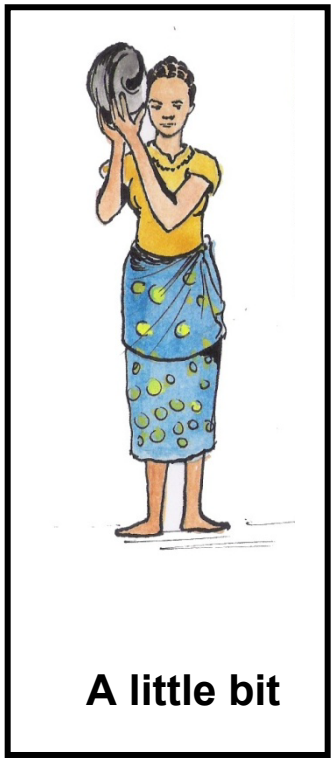
Nonverbal Response Card for Function*





*Adapted by Chishugi Oswald, IRC/Bukavu,

Nonverbal Response Card (symptoms)*



*Adapted by IRC/Bukavu, Democratic Republic of Congo

APPENDIX F: USING A RANDOM NUMBER TABLE

The accompanying random number table is in a typical format, with each line numbered on the left side and the random digits following across the page in groups of ten digits. In drawing random numbers from the table, one can begin anywhere and go in any direction (up, down, forwards, in reverse, diagonally) but, once begun, should continue in that direction. For convenience, most users go forwards, starting at the top left hand number and move to the right, then down to the left side on the next line, as if reading a page of text.

To use the table, the only information required is how many numbers you want to choose and the range of numbers from which they are to be chosen. For example, if two randomly chosen numbers between one and nine are needed, a user can begin at the upper left hand corner and begin reading, one digit at a time. The first digit is three, which lies between one and nine, and so is chosen. The second digit is also three, but we ignore this since we chose it already. The third digit is six, which also lies between one and nine. Therefore, the numbers chosen at random are three and six. However, if the range was between one and ten, then the numbers should be read off in groups of two because the range includes a number with two digits (i.e., ten). In this case, we would begin by reading off 33, which is discarded because it is beyond the range of one to ten. Likewise we continue discard 60, 52, 44, 95, 50, 14, 37, 61, 98, and all the other two digit numbers on line one until we get to 02. Since this does lie in the range of one to ten, it is chosen. We then continue on to the second line. We pass 02 again since we already selected this number, and continue discarding numbers because they are greater than ten until we get to 04 in the fourth column of the third line. Therefore, the two numbers selected at random between one and ten are 02 and 04.

Similarly if the highest number in the range has three digits (in the hundreds) or four digits (in the thousands) we read off groups of three or four digits respectively, disregarding any that we already have chosen or that are outside the selected range.

In the assessment survey we assign each supervisor and interviewer a different line at which to begin and instruct them to read forwards. Each time they use the table they make a mark at the last number they read. The next time they use the table they begin at that number. It does not matter if the last time they used the table they were reading off groups of two digits and this time they need only one digit at a time, or vice versa. In the unlikely event that any user would reach the end of the table, they continue at the top left hand corner. When numbering houses from a sector, the range will be the number of houses and the supervisor will select 20 numbers from within that range. At houses, the range will be the number of adults in the house and the interviewer will chose the first number within that range.

RANDOM NUMBER TABLE

Line

01	3360524495	5014376198	4092518211	4834425845	8785455990	2432950253
02	7594681597	1718506620	1668398885	1987315094	9969967480	7140176214
03	3155820299	2182292666	2573204381	1300170469	4172713081	3349300541
04	1550513541	6185291791	9996957059	3595322298	0606144798	2994766062
05	9156090237	4381909493	8633029343	2082377861	4085955645	6363082323
06	2152245223	4393750031	9500104699	1563610483	5523502529	3643165151
07	4139890686	2583731685	4307502757	0352006951	9676405681	3868005188
08	2116667187	8487166164	5187482854	2508357958	3393529942	7653671331
09	8869344418	6774685416	7954539048	0286266566	6487481182	8184060977
10	8984219998	7469760039	3937014341	6658918808	7179887038	1645014950
11	2091067709	8064775779	7573857383	3080074211	3934169074	3002612812
12	1569587231	7402674576	3925184851	9289399083	1330147508	7429396765
13	9640040001	9849986321	1710238888	9906328413	7438692269	1887606767
14	5597455639	5462039332	7021330318	7211738082	9984216642	0268033790
15	7570965109	7391244310	0231014110	8699569727	7244279578	8269205853
16	6892148494	4862175480	6668397770	8478285224	1567226853	3814410193
17	9624292909	1795674132	8573122253	6623346104	7452470732	9756733917
18	3085162618	7780200167	8317057616	2002706493	7905456761	6774539671
19	9767176267	4255115627	5090643200	3382944533	9058557780	9767768386
20	3193736696	3781784220	6044624555	8756912407	0457483205	1283856235
21	1252582078	5226931051	5229352210	9161182880	7269090892	3403605307
22	6427125360	3312711833	0814458908	2917631318	7156242681	2532462507
23	1118217502	9402145998	2292608345	5538921257	9370896154	8904268280
24	9502367931	8345932400	2544505153	4293858115	6114125591	0665575637
25	7687472946	8648068897	5888683181	3874706245	4661618897	3481488494
26	0058210328	7991410647	9914032815	9395746474	0241742410	5923470018
27	4557755740	8653196793	2273047172	9907646107	7447832929	0829954813
28	8376241758	7749349845	0904352341	2961099543	3357908853	3030197563
29	4631541792	6120295024	7671656566	8251720550	5851673235	2290520132
30	3618775144	6267156764	4513210649	7900421800	0568170673	1463546044