PART 1: THEORETICAL OVERVIEW OF MEASUREMENT ISSUES AND STRATEGIES

1.0 Introduction

To judge the efficacy of vaccine, it is essential to eliminate competing explanations for changes in disease incidence during the vaccine trial. This requires careful and systematic monitoring of key risk-enhancing and risk-reducing behaviours that have the potential to influence HIV transmission. The following protocol reviews various strategies for monitoring risk behaviour and sets out a prototype design to be used in a Phase III vaccine trial. This document is divided into two parts. The first part gives an overview of risk behaviours for HIV/AIDS and strategies for measuring these behaviours. The second part provides a practical guide for developing a behavioural monitoring system.

2.0 Overview of risk factors for transmission of HIV

The basic mechanisms for transmission of HIV are now well established. Viral transmission requires exchange of blood or blood products containing the HIV virus from an infected source to a susceptible recipient. Infection occurs through viral entry into cells through CD4 and chemokine surface receptors (Royce, Cates et al. 1997). HIV receptive cells are found in the bloodstream (CD4 T lymphocytes, Langerhans’ cells, and other macrophages) and on surface of oral, genital, or rectal epithelia.
The principle modes of viral transmission are also well known. The two most common routes of adult transmission are through penetrative sex and IV drug use. Transmission through sexual contact is estimated to accounts for 75 to 85 percent of cumulative cases of HIV/AIDS to date (HIV/AIDS 1996). The importance of IV drug use varies by setting and accounts for a growing proportion of HIV transmission in industrialized countries. Other modes of transmission include direct blood transfusion, which may account for as much as 10 percent of transmission in sub-Saharan Africa (Cohen and Trussell 1996) where up to 25 percent of donated blood is still not screened for HIV (MAP 1998). In other regions, infection through blood transfusion generally makes a small and diminishing contribution to new caseload. Exposure to blood products of infected individuals through non-sexual contact, such as contact with open wounds or skin sores is assumed to be rare in most settings. Perinatal transmission to children plays a larger role in global transmission, accounting for 15 to 20 percent of new cases in sub-Saharan Africa and 5 to 10 percent worldwide (Quinn 1994), but is not of concern for monitoring adult risk.

Viral transmission is a complex function of host infectivity and susceptibility, environmental factors that affect exposure to infection, and to some extent factors associated with different subtypes of the HIV virus itself. Mode of transmission itself is one of the most important predictors of risk. The risk of transmission per contact event varies with the likelihood of exposure of virus-infected fluids with blood of the recipient. (Copy figure from Royce review article in NEJM?) Vaginal intercourse represents the most common yet the least efficient mode of transmission per contact. Women generally appear to be more susceptible than men due to greater likelihood of trauma to the vaginal epithelium during intercourse, plus the longer exposure to HIV virus in ejaculate after intercourse (Haverkos and Quinn 1995). Traumatic sex including rape, sex during menses or sex using vaginal drying agents common in some African cultures all enhance the likelihood of transmission during vaginal intercourse. The statistical risk of infection through vaginal intercourse is notably low regardless of which partner is infected. In single episodes of heterosexual sex when the male is infected, transmission rates may be as low as 0.1 percent. Female-to-male rates are lower still. Anal intercourse has higher infectivity due to the likelihood of damage to epithelial integrity. Accidental needle stick and needle sharing with infected partners falls within the same range of efficiency as anal intercourse at approximately 1 percent infected per contact.
A number of factors affect host susceptibility to HIV. Once assumed to be universal, susceptibility now appears to be significantly reduced if not eliminated by the presence of a mutation in the CKR5 chemokine-receptor gene. This trait has been found in sex workers and homosexual men who remain uninfected despite having repeated unprotected sex with infected partners. Other factors affecting susceptibility include cervical ectopy (replacement around the cervical os of normal multilayered cervical squamous cells with glandular, single-layered columnar cells that are typically found inside the os) that renders cervical tissue more vulnerable to damage and infection. Women with this condition have 2 to 5 times the risk of infection in most studies. Similarly sex during menses increases susceptibility for both women and men. Male circumcision has a consistent protective effect against infection as well as reducing infectiousness.

Infectiousness is a function of virus concentration in the bodily fluids to which the uninfected person is exposed, and has been shown to vary systematically with selected characteristics. Stage of infection is an important factor. Infectiousness is highest at the earliest acute stage of HIV infection and rises again towards the late clinical stage leading to manifestation of AIDS. Viral load in the bloodstream or bodily fluids which may in turn affect infectiousness is increased by pregnancy and activation of the immune system from vaccination, among other factors. A recent addition to the list of factors affecting infectivity has come through the development of antiretroviral therapies. While the exact physiological mechanism is not clear, drug therapy is associated with a 50 percent reduction in HIV transmission. Consistent condom use is protective against infection. The role of other contraceptives on host susceptibility is uncertain.

One of the most important host factors affecting both susceptibility and infectiousness is the presence of other sexually transmitted disease. Empirically the presence of STDs has always been strongly associated with increased HIV prevalence. Across various sites in sub-Saharan Africa, for example, seroprevalence of HIV among STD patients ranges from 2 to 20 times higher than in the general population (Cohen and Trussell 1996). Even commercial sex workers with STDs suffer higher rates of HIV infection than those without STDs. However, the causal link between prior STDs and HIV has been difficult to establish due to the confounding of STD infection with other risk factors for HIV transmission, primarily sexual promiscuity. An accumulation of recent research has confirmed the causal link (Cohen 1998),
and cast light on what Wasserheit has termed ‘epidemiological synergy’ between STDs and HIV (Wasserheit 1992). Open sores and disturbance to epithelial integrity caused by ulcerative STDs provide a direct route for transmission of the virus. Even non-ulcerative STDs are associated with increased transmission, though different mechanism may be at work (Laga, Manoka et al. 1993). At the same time, infection with STDs appears to increase viral loads in the infected host, possibly through the stimulation of the immune system. Strong evidence of an independent role of prior STD infection comes from the Mwanza, Tanzania trials, where HIV transmission was lowered by 42 percent through syndromic treatment of STDs (Grosskurth, Mosha et al. 1995).

Arguably the least amount of progress has been made in understanding characteristics of the ‘environment’ that facilitates the spread of HIV (Cohen and Trussell 1996). The key environmental characteristic is the behaviour and contextual determinants of behaviour that expose individuals to risk. How behavioural risk is conceptualized is critical to decisions of how best to measure it. Holding physiological susceptibility and infectiousness constant, risk of infection is a function of exposure to the virus, which in turn depends on two sets of factors. The first is individual risk behaviour affecting the number and pattern of sexual partners or needle sharing events. The second is dynamics of sexual networks that affect exposure to infection from outside groups and the spread of the virus within each network once the virus has been introduced. Knowledge of any one dimension is insufficient to explain the risk of infection to the individual or the progression of the epidemic at the community level. An individual without any marked risk characteristics may be at higher risk of infection through a single partner in a high prevalence setting than an individual with multiple risk behaviours in a low prevalence setting.

For a variety of reasons research on behavioural determinants of HIV has focussed primarily on individual risk characteristics in an effort to identify a ‘core group’ of high risk individuals who fuel the epidemic. Morris (1997) and others argue that the central unit of observation in the study of HIV should be shifted from the individual to the partnership to better understand the dynamics of disease spread in the population. This has fundamental implications for data collection strategies, which are discussed in the next section.

3.0. Measurement of Risk Factors for HIV
Neither of the two principle modes of transmission are directly observable due to the socially sensitive nature of sexual behaviour and additional legal prohibitions in the case of IV drug use. The methodological challenge that confronts behavioural research is how to obtain valid and reliable estimates of key indicators of risk on a population level. A variety of strategies have been employed. Generally, they can be grouped into direct methods, all of which rely on some form of self reported accounts of behaviour, and indirect methods that utilize secondary markers of behaviour that do not depend on self reports. The methods are reviewed below.

3.1 Theory and practice of measuring self reported behaviour

A wide variety of data collection strategies for self reported behaviour vary by interview mode, place, format, as well as content, order, and wording of questions. Modes of data collection range from one time questionnaires to daily diaries. Information may be gathered by interviewers in a face-to-face interaction with the respondent or self administered in written questionnaires. Recent innovations have experimented with mixed formats combining face-to-face interviews with self administered modules for the most sensitive questions (Johnson, Wadsworth et al. 1994; Wolff et al. 2000) to audio-administered questionnaires in which subjects respond to tape recorded questions and record their responses on an interview form or directly onto a computer (Turner, Ku et al. 1998). Instruments vary from fixed format questionnaires to open-ended semi-structured interviews. The length of the instrument varies anywhere between 5 minutes for a daily diary to over an hour for a full interview. The wide variety of approaches share a reliance on respondents’ reports of their own sexual behaviour or drug use.

The best configuration of different design features depends ultimately depends on the overall objectives of research (Weinhardt, Forsyth et al. 1998). Methodologically, as will be seen in the discussion to follow, the need to produce valid and reliable measures of sensitive behaviours often comes at odds with the need to produce representative results for large populations. In his pioneering study to establish general patterns of sexual behaviour in the United States in the 1940s, for example, Kinsey (1948) employed detailed diary accounts
from a snowball sample that provided higher validity of reports at the expense of representativeness. Behaviour monitoring for a Phase III trial must meet 3 basic criteria:

- valid and reliable indicators of exposure to risk of HIV transmission among trial participants
- comparable validity and reliability of self-reported measures over time
- applicable without excessive cost in time or money to relatively large sample sizes on the order of thousands to tens of thousands.

The size of Phase III trial groups effectively requires some form of population based method, best exemplified by sample survey techniques. For any given individual respondent, sample survey techniques are subject to a number of sources of error that makes them a potentially poor method for assessing risk. For assessing change in patterns of risk behaviour at the population level over time, however, it will be argued here that survey methods can meet the objectives of a vaccine trial. Validity and reliability of survey techniques can be enhanced by employing design strategies developed since Kinsey’s first efforts to minimize the major sources of bias and error in self reports of sensitive behaviour.

### 3.2 Sources of error in survey methods on sensitive behaviours

Sample survey methods often provide the best way to produce quantifiable parameters of behaviour from large populations using limited resources. As Zeller (1993) notes, however, attempts at precise measurement using ill-defined concepts in ill-defined populations can be ultimately counterproductive. The challenges are multiplied by the sensitive nature of information about behaviours at the core of HIV transmission (Lee and Renzetti 1990). The ability of sample survey techniques to measure sexual behaviour and other sensitive topics is the subject of considerable skepticism, particularly among anthropologists and other advocates of close ethnographic research techniques (e.g. Caldwell et al, 1989). The design of population-based measures then must take into account the sources of error and bias that threaten validity and reliability of this approach.

#### 3.21 Threats to validity and reliability: Recall error
There are two major sources of error that threaten validity and reliability of self reported sexual behaviour: recall error and social desirability bias. The first and more easily surmountable is the problem of recall. Self reports on sex typically rely on recall of the number of events and details about each within a given time frame. For example, respondents may be asked ‘How many times did you have sex in the last month?’ or ‘How many sexual partners did you have in the last year’. The salience of an event determines the likelihood of accurate recall, with recurring or everyday type of events being most prone to recall error. ‘Falloff’ of memory has been found to drop rapidly up until 12 months and stabilize afterwards, although it can be improved with extensive probing by the interviewer (Croyle and Loftus 1993). Reliability of couple reports of sexual behaviour has been found to be significantly worse at 12 months than 3 months or two weeks (Kauth 1991).

Recall is also influenced by psychological factors. Pleasant memories are more likely to be recalled than unpleasant memories, and both are more likely to be recalled than neutral events. Since men are more likely to report sex as pleasurable than women, it may contribute to the consistently higher levels of sexual activity reported by men (Catania, Chitwood et al. 1990).

The tendency to simply forget remote or commonplace events is not the only threat to accurate recall. Research on memory suggests that recall depends on two types of memory. ‘Episodic’ memory stores information on specific events, and ‘semantic’ memory that stores general principles and patterns of behaviour (Croyle and Loftus 1993). As common events recede in time, they are likely to be recalled as part of a general pattern which often if not always diverges from actual experience. A clear example is the common tendency of respondents in surveys to report higher coital frequency in an ‘average’ week than in the last week, reflecting the tendency to overlook interruptions of normal routine due to menstruation, illness, travel, or other temporal disruption. Memory has also been observed to be malleable to the influence of more recent events, (for example recall of past diet has been shown to be influenced by current diet), reinterpretation of events by the subject or even influential others. This may help to explain the finding of a study of couple disagreement over sex in Uganda of low concordance of the event and the tendency for both members of the couple to recall ‘winning’ (Wolff, Blanc et al. 2000; Blanc, Wolff et al. 1996).
3.22 Threats to validity and reliability: Social desirability bias

The other major source of error comes from biased reporting of behaviour or non-response due to perceived social desirability of drug use and sexual behaviour. Social desirability reflects social norms concerning the acceptability of engaging in certain types of behaviour, of reporting it to an interviewer or on a questionnaire, or both. It is a particularly difficult problem because the direction of bias depends on the operative social norms in a group and cannot be consistently predicted, and the magnitude is unknowable without validation measures. Sexual behaviour is typically a highly private matter in most settings that evoke feelings of embarrassment, shame or threat among people when asked to report about it to relative strangers (Lee and Renzetti 1990; Johnson, Wadsworth et al. 1994). For drug users the sense of threat is multiplied by the legal implications of reporting prohibited behaviour. It is commonly assumed in the literature on this topic that social desirability bias will lead to underestimates of sensitive behaviours (Bradburn and Sudman 1983; Catania, Chitwood et al. 1990). However the possibility that the opposite tendency to exaggerate reports is encouraged in some groups or settings cannot be ruled out.

3.23 Threats to validity and reliability: Non-response bias

At the extreme, social desirability bias arises from non-response to particular items on a questionnaire or refusal to participate altogether. Item non-response has been shown to increase with sensitive questioning and vary between 6 and 19 percent in published reports (Catania, Chitwood et al. 1990). Non-responders tend to be older, with lower reading ability, and more sexually inhibited than responders, suggesting their exclusion would bias estimates of risk behaviour downwards. The effect of survey non-response is more difficult to estimate, though there is evidence that both highest and lowest risk groups may be underrepresented in typical samples. In Uganda, prostitutes were more likely to refuse to participate in household surveys (Konings, Bantebya et al. 1995) while in the U.S. HIV infected individuals were less likely to volunteer for study participation by a factor of 5 for whites and 9 for black and Hispanic residents (Hull, Bettinger et al. 1988).

3.3 Strategies for improving survey measures of sensitive behaviours
The ability of sample surveys to measure sensitive behaviour ultimately depends on their ability to minimize these foreseeable sources of bias and error. Measurement errors are systematically influenced by characteristics of the respondent, mode of interview, and instrument design. The following section reviews the recent literature on methodological strategies to improve survey reports. While much of the literature applies to sexual behaviour, many of the issues are similar for IV drug users and other sensitive behaviours.

### 3.31 Respondent effects

All reporting error can ultimately be attributed to individual characteristics of respondents that determine how each responds to particular questions, but particular characteristics appear to systematically influence reporting error across settings and measurement strategies. Young age, sexual orientation, risk perception of HIV in sexual behaviour surveys have been cited in the literature (see Catania, Turner et al. 1990, Weinhardt, Forsyth et al. 1998). To the extent that recall problems will be relatively constant across groups, the individual characteristics that are most likely to lead to reporting problems reflect sensitivity to social desirability pressures in the interview situation. Again, it is important to stress that the specific characteristics and the direction and magnitude of potential biases associated with them are likely to be culturally specific.

A universal characteristic to improve reporting quality of reporting over which research design does have some control is the level of psychological motivation. High motivation to participate in a study may overcome normal social desirability biases about reporting sensitive subjects. A subject who is motivated to participate may listen to questions more carefully and attempt to recall more accurately than one who has less investment in the objectives of the study. They are also less likely to be non-responders or take offence at questions about normally private matters (Catania, Turner et al. 1993).

Several strategies of study design have been demonstrated to increase motivation to participate. Recruitment of subjects on a volunteer basis is an obvious strategy that selects for willingness to participate. The advantages of volunteer samples need to be weighed against the likelihood that volunteers are atypical in terms of risk taking behaviour and hence may not provide a representative picture of risk levels in the general population. Investment in building rapport with respondents is another strategy. This can begin at the level of
questionnaire design by such measures as adding more open ended questions or increasing the
number of non-threatening questions before asking about intimate matters. Mode of
interview may contribute. Face-to-face interviews are naturally more conducive to rapport-
building than more impersonal contact through telephone or self administered questionnaires.
The disadvantage is that increased motivation gained by using interviewers may be offset by
the loss of privacy, confidentiality, and convenience. Hence privacy and motivation form two
poles that pose frequent dilemma for research design.

3.32 Mode of interview

Mode of interview has been shown by methodological research to have a strong influence on
quality of reporting. The reasons should not be surprising. A survey interview is a social
interaction as much as a data collection exercise. The motives of a well trained interviewer
are unambiguous; he or she must complete the questionnaire in the expected time and move
to the next interview. To the respondent, whatever desire to comply by reporting the truth has
to be balanced with the natural instinct to present oneself positively in the eyes of a stranger.
Inevitably interpersonal communication during the course of the interview occurs between
interviewer and respondent that is likely to be scanned by the respondent for evidence of
approval or disapproval. To some extent, the same effect may apply for self administered
questionnaires where a respondent is asked to reveal behaviour to an unseen audience. In
either case, the point of interview becomes the focus for social desirability biases.
Table 1: Summary of Practiced Advantages and Disadvantages of Modes of Self-Report Sexual Behaviour Assessment

<table>
<thead>
<tr>
<th>Assessment mode</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-administered questionnaire (SAQ)</td>
<td>• Privacy</td>
<td>• Literacy dependent</td>
</tr>
<tr>
<td></td>
<td>• Efficiency</td>
<td>• No potential to probe responses</td>
</tr>
<tr>
<td>Face-to-face interview (FTFI)</td>
<td>• Ability to build rapport</td>
<td>• Less privacy (especially with household interviews)</td>
</tr>
<tr>
<td></td>
<td>• Ability to probe ambiguous, internally inconsistent, or nonresponse</td>
<td>• Reaction to interviewer</td>
</tr>
<tr>
<td></td>
<td>• Ability to explain unfamiliar terms</td>
<td>• Inefficiency</td>
</tr>
<tr>
<td></td>
<td>• Enhanced credibility</td>
<td>• Possibility of inaccurate interpretation of responses by interviewer</td>
</tr>
<tr>
<td></td>
<td>• Minimizes nonresponding</td>
<td></td>
</tr>
<tr>
<td>Telephone interview</td>
<td>• Ability to probe ambiguous, internally inconsistent, or missing responses</td>
<td>Obscene phone-caller phenomenon (less credibility)</td>
</tr>
<tr>
<td></td>
<td>• Ability to explain unfamiliar terms</td>
<td>• Many high-risk individuals are homeless or transient and have no phone</td>
</tr>
<tr>
<td></td>
<td>• More privacy than FTFI (no visual contact)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Less interpersonal reaction to interviewer</td>
<td></td>
</tr>
<tr>
<td>Postal questionnaire</td>
<td>• Privacy</td>
<td>Literacy dependent</td>
</tr>
<tr>
<td></td>
<td>• Efficiency</td>
<td>• Less credibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Low response rate</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>• Minimizes effects of memory error: Less time between behaviour and recall</td>
<td>Reactive-responding may impact sexual behaviour and reporting of sexual behaviour</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requires more effort and commitment from participant</td>
</tr>
<tr>
<td>Audiotape-administered interview/questionnaire</td>
<td>• Privacy</td>
<td>Literacy dependent</td>
</tr>
<tr>
<td></td>
<td>• no interpersonal reaction to interviewer</td>
<td>• No potential to probe responses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Relatively difficult to repeat questions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• reaction to voice on tape: who is this person, and how would they judge me?</td>
</tr>
<tr>
<td>Computer-administered interview/questionnaire</td>
<td>• Ability to prove ambiguous, internally inconsistent, or missing responses</td>
<td>Literacy dependent</td>
</tr>
<tr>
<td></td>
<td>• Ability to explain unfamiliar terms</td>
<td>Expense and inefficiency: One participant per computer</td>
</tr>
<tr>
<td></td>
<td>• Privacy</td>
<td>Level of comfort/familiarity with computers may affect responses.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The demand for improved population estimates of risk behaviours had led to a number of methodological innovations in modes of data collection in large scale studies. A list of principle modes and the advantages and disadvantages of each is shown in Figure 1. Each mode can be evaluated on several key dimensions that affect the quality of reporting. These include degree of privacy, motivation and rapport building, and quality control.

**Face to face interviews (FTFI)**

Face to face interviews represent the common standard of interviewing against which others are judged. The advantage of FTFI design is its flexibility and robustness across any configuration of instrument design, setting, and respondent characteristics. FTFI interviews can be administered virtually anywhere to anybody. The presence of a live interviewer allows greater complexity of questionnaire design (e.g. use of skips and filters, multiple loops), question wording and question emphasis. Although interviewers contribute an independent source of error to survey data, they can significantly reduce response error from respondent miscomprehension and detect and probe for internal inconsistency or incomplete responses. FTFI hold another advantage in quality control over self administered formats when thorough training and consistent monitoring are practiced. Finally, live interviews may build motivation to participate by building a sense of social connection to the study and conveying greater effort on the part of the research project than more impersonal appeals for information over the phone or through the mail system.

The main disadvantage of FTFIs are their costs and tendency to magnify social desirability biases, particularly in the context of sensitive behaviours. Evidence for this can be found in a large literature on methodological research on costs and interviewer effects (see Groves 1989). For example, female respondents reported significantly fewer sexual partnerships to male interviewers than to female interviewers (Schopper, Doussantousse et al. 1993). It is also evident from comparison with other modes of interview discussed below.

**Telephone interviews**

The main advantages of telephone interviews lie in low cost, rapid data collection, and wide geographic coverage. Quality control can be assured through standardized training and monitoring of interviewers. Phone interviews improve privacy through their relative
anonymity and by removing the possibility of non-verbal interaction between respondent and interviewer. On the other hand the impersonality of this form of interview may reduce the credibility of the interviewer in the eyes of the respondent and reduce motivation to participate or respond truthfully. The difficulty of tracing very high risk populations such as IV drug users through phone surveys may disqualify this from consideration for regular risk monitoring, though they may be considered for validation exercises among subsamples of respondents who agree to participate during face-to-face contact.

**Self administered questionnaires (SAQ)**

Self administered questionnaires represent a significant improvement in privacy and should improve validity of responses by eliminating interpersonal dynamics between interviewer and respondent as a source of error. They may be administered under a wide variety of conditions ranging from organized classroom settings to mail-back samples. SAQ suffer from several drawbacks. First, depending on the mode of administration they may lack the credibility necessary to motivate sincere participation. Particularly in the case of mail surveys, non-response is significantly higher than in face-to-face or phone interviews. Second, they assume high levels of literacy that limits their applicability in many settings. Finally, interviewer error or ‘noise’ with the lack of formal training or any checks against incomplete, missing or erroneous responses becomes a serious consideration with SAQ. The degree of complexity of the questionnaire must be limited. One strategy adopted in a national study of sexuality in the UK is to administer rapport-building questions in a face-to-face interview and leave the most sensitive questions for a self-administered questionnaire filled in the interviewers presence but without their knowledge (Johnson, Wadsworth et al. 1994). Responses were returned in sealed envelopes only identified by a code number.

**Audio-administered or computer-administered interviews**

One promising new strategy for obtaining the quality control of face-to-face interviews with the privacy of self administered questionnaires is through audio or computer-administered techniques. In audio-administered designs, the respondent listens to an audio taped interview and records responses on a self administered questionnaire. Subjects must be literate to fill the questionnaire. For computer assisted techniques, the interview is recorded on disk or CD, and responses entered onto computer screen using a keyboard or touchscreen device. Order, phrasing and emphasis of the questionnaire is standardized, and clarification or additional
information on response options is available with special screens. Proponents of this approach claim it can be used in low literacy settings, though basic ability to learn keyboard or screen commands as well as the infrastructure necessary to operate computers must be assumed. As part of the 1995 National Survey of Adolescent Males in the U.S., 1690 respondents were randomly assigned to answer questions using audio-computer assisted survey instrument (audio-CASI) or the more traditional self administered questionnaires SAQ(Turner, Ku et al. 1998). Interestingly, for more conventional risk behaviours such as male-on-female sex or number of lifetime partners, computer techniques yielded similar results to SAQs, suggesting that improvements to reporting validity through such extraordinary measures to protect privacy are limited to the highest risk behaviours. Reporting of behaviours considered less shameful is less sensitive to mode of interview.

Summary
The quality of reporting on sensitive issues can be improved by altering the mode of interview. Self-administering strategies appear to reduce social desirability effects and improve reporting of highly sensitive behaviours. This is particularly relevant in a study of a high risk population where many of the critical risk behaviours are socially stigmatised, illegal, or both. The feasibility of self-administered designs ultimately depends on a minimal level of literacy in the study population. Design strategies also need to balance the need to maintain privacy against the added advantages of building rapport and motivation to participate among respondents that are best achieved with personal interaction with interviewers. A hybrid approach where the most sensitive questions are self administered in the presence of interviewers seems to address both concerns.

3.33 Instrument effects.

Question wording, question order, interview length, and open versus closed format questions have all been shown to affect reporting error in sample surveys. As in mode of interview, their influence stems from how questionnaire construction amplifies or diminishes recall error social desirability biases, or motivation to comply among respondents.
**Question wording**

Principles of question wording for sensitive behaviours have changed little since the original Kinsey studies. The first is to phrase questions on sensitive topics in such a way to place the burden of denial on the respondent. Yes-no questions on sensitive behaviours should be avoided in favour of questions that ask for more detail on any dimension and deny any easy retreat to a socially acceptable answer. For example, the prevalence of premarital sexual activity reported by adolescents in a Mexican study was 15 percent when respondents were asked, ‘Have you ever had sex?’ compared to over 40 percent when the question was phrased, ‘How old were you when you first had sex?’. (Morris, personal communication).

This strategy assumes a general tendency to underreport, and may yield overestimates if that assumption is incorrect (Catania, Chitwood et al. 1990).

A second principle is to ask questions in a direct fashion, without apology or hesitancy. Embarrassment on the part of the interviewer can only intensify temptations to provide the social desirable response. Another strategy to improve response validity is to preface sensitive sections of the questionnaire with statements designed to counteract known social desirability biases. For example, the following statement is taken from an actual prospective study of women working in bars where exchange of sex for money or goods is common. The statement is read before embarking on questions about multiple sexual partnerships.

*It is common for many women in busy towns like this to have more than one partner, even to earn extra money by having sex with men. To understand about health, we need to ask about some questions about different kinds of partners. Please do not be offended if some questions do not apply to you. Everything you say will be strictly confidential.*

The third principle on adopting language familiar to the respondent over more formal scientific terminology remains a central point of debate. Advocates of using informal language claim that formal terms may lead to more measurement error through lack of recognition or through intimidation; Informal language will be clearer to the respondent and help to build rapport and comfort with the interview process. Critics point out that use of informal terms often lack exact definition of behaviour and may counteract rapport building if they appear inappropriate or condescending to the respondent. For example, the term ‘sex’ may be readily recognized by respondents, but may or may not correspond to the intended
meaning of ‘penetrative vaginal intercourse’. Moreover the sheer diversity of alternative terms for sex or drug use in socially diverse samples presents the difficulty in arriving at a common language for the same behaviours. With trained interviewers it is possible to tailor the instrument to each respondent by establishing common terms and definitions to be used in questions on behaviour to follow (Johnson, Wadsworth et al. 1994). For self-administered questionnaires this is more difficult to accommodate individual or cultural variations within a single sample. The best solution depends on the ultimate objectives of the study. In a risk monitoring exercise for a vaccine trial, the threat of alienating or unfamiliar language may be outweighed by the need for precise scientific definitions of behaviour where no suitably clear, commonly shared vernacular terms exists. In case local language is used, it is advisable to include the scientific terms in parentheses after the question for the purpose of probing (Weinhardt, 1998, Johnson et al., 1994).

Question ordering, instrument length, open vs. closed ended questions

A variety of techniques can be employed in the construction of questionnaires to help improve quality of reporting on sensitive issues. Ordering question in increasing order of sensitivity, starting with the least threatening and moving progressively towards more intimate topic areas, is thought to allow more time for rapport to develop between interviewer and respondent. In one explicit methodological study testing this hypothesis, the amount of time and number of questions devoted to rapport building did not significantly affect the level of sexual behaviours reported (Padian, Aral et al. 1995). However both respondents and interviewers expressed marked discomfort with the more abrupt approach through short form questionnaires without rapport building questions in advance. Longer questionnaires and addition of open-ended questions may both contribute to the ‘conversational’ ambience of the interview that is conducive to rapport building. Question ordering may bias responses through the well-established tendency of interview process to ‘sensitize’ respondents and influence their responses on later questions (Bradburn and Sudman, 1983). To avoid additional bias on key behavioural indicators, some researchers insist on postponing attitude questions that ask for moral judgements about the rightness or wrongness of certain behaviours until after actual behaviour questions (see Johnson et al., 1994). Others have used sensitizing nature of moral attitude questions to their advantage. Reported levels of male-on-male sex increased when behaviour questions were prefaced by a series of attitude questions on homosexuality (Catania, Chitwood et al. 1990). behaviours in the wider population, which
helped to give respondents some forewarning of the direction of the interview, and place behaviours in a wider social context that helps remove the fear of moral condemnation.

*Improving recall*

Several instrument design techniques have been suggested to improve the quality of retrospective reports of behaviour that compensate for the elusive and malleable nature of human memory. Most effective of all is to shorten the period of recall. The less memorable or more frequent the event, the more rapidly the quality of recall deteriorates. Recall periods should be clearly demarcated. Initiating a question, “In the last 12 months’ rather than ‘In the last year’ will avoid potential confusion between calendar and chronological years. For certain types of event recall, establishing a time line with the respondent that helps to sequence events and place them in historical context has been shown to be an effective memory aid, although this is expensive in terms of numbers of questions and time. A less expensive option is ‘seed questions’ to sharpen the boundaries of observation periods in the respondent’s mind. For example, reliability of events in the last month were improved by prefacing the question with another about the same event in the last 6 months (Croyle and Loftus, 1994). Information on specific events (i.e. When was the last time you had sex with your regular partner?) is typically more reliable than information on frequency or general patterns (i.e. how many times a week do you normally have sex?) that rely on sex.

3.4 *Specific application to repeat monitoring exercises*

Reviewing different methodological techniques to improve validity and reliability of survey measures of sensitive behaviours, there is no single best strategy, especially when budget and time constraints are taken into consideration. Each must be evaluated according to the overall objectives of the study. The objectives of monitoring risk behaviour for a vaccine trial place extra demands on valid measurement on one hand, but remove some of the normal sources of error that affect other population-based measures. For example, the usual threats to validity and reliability may be eased by high motivation of trial participants. All are volunteers. All are selected based on identification with a high risk population and possibly through willingness to report high risk behaviours in screening interviews. All are subjected to repeated visits as part of the vaccine trial regime that should reduce the need for rapport building measures for each interview.
On the other hand, the difficulties of monitoring risk behaviour are increased by the need to obtain precise estimates of HIV risk behaviours at repeated points in time and by the unique biases introduced by the vaccine intervention. Behavioural measures must not only identify high risk characteristics for individuals but shifts in the pattern of sexual or drug networking that independently affect transmission. A screening instrument to distinguish high risk from low risk individuals is not sufficient. The presence of treatment intervention adds a new set of potential biases to reporting behaviour. Receiving a vaccine may influence subsequent risk behaviour, reporting of that behaviour, or both. Changing behaviour as a result of participation in a vaccine trial should not pose a threat as long as an effect monitoring system is in place. Changes in reporting patterns may pose more of a threat, but one that is not necessarily fatal to the study evaluation if it occurs evenly over intervention and control populations. For example, repeated monitoring may improve attention to and recall of high risk behaviours over time, thus leading to an artifactual increase over time in the prevalence of related risk factors. The real danger for a monitoring system is if changes in reporting occur differentially across treatment and control groups. Double blinding of intervention is essential to prevent differential reporting through conscious or unconscious shifts in reporting by respondents, interviewers, or both. In the absence of a control vaccine, the need for validation measures that do not rely on self report becomes essential. External validation is advisable in any case given the multiple threats to validity of self reported behaviour on surveys discussed above. Different strategies for validation are discussed in the following section.

4.0 Validation strategies

4.1 Internal validation

Approaches to validation of survey data can be grouped into internal and external strategies. Internal validation measures generally rely on asking the same question in more than one way at different points in the interview to catch internal inconsistencies. For example, a respondent might be asked to give the number of sexual partners they have had in the last year, and then asked to recall the most recent partners in reverse chronological order. The number of partners recalled in the last year can be compared to the reported total. Internal measures require no additional resources aside from increased numbers of questions, and
increase confidence in important parameters in the absence of external validation. However they are generally of limited value beyond alerting the presence of certain types of reporting bias. Arguably they are most useful if interviewers can be alerted to inconsistencies at the time of interview and encouraged to probe for more accurate information.

4.2 **External validation: population based measures**

External validation requires comparison with data collected from independent sources, different modes of interview or different instrument design. It can take one of three forms; comparative population-based measures, in-depth ethnographic techniques on smaller subsamples, or collection of biomarkers of risk behaviour. Comparison of external sources of population based data probably represents the least desirable method of validation. For sexual behaviour, a common validation technique is comparison of partner reports as a measure of reliability. Validity of any particular measure cannot be judged using partner reports, since neither partner can be expected to provide the ‘true’ benchmark of behaviour. However comparative levels of partner agreement are useful to judge the *relative* reliability of different measures or methods. Lack of partner agreement can also be taken as a clear sign that any particular measure or method lack validity as well as reliability. Partner studies have been used to demonstrate declining reliability of self reports with length of recall (Ellish, Weisman et al. 1996; de Boer, Celentano et al. 1998), to evaluate different modes of data collection on sexual behaviour (Padian, Aral et al. 1995), or to judge the comparative reliability of survey measures of different types of sexual behaviour (Kinsey, Pomeroy et al. 1948; Coates, Calzavara et al. 1988; Schopper, Doussantousse et al. 1993). This method has a number of disadvantages for the purposes of a vaccine trial. Besides inability to provide a true measure of validation, its use is limited to sexual behaviours within stable unions which account for the lesser part of HIV risk behaviour.

One imaginative strategy to judge the accuracy of survey measures is to deliberately include a group with known risk characteristics into the standard sample. Konings et al. (1995) tested this approach by incorporating a subsample of commercial sex workers in a double blinded household survey of sexual behaviour. The proportion of the subsample reporting this to an interviewer provided a simple indicator of validity for comparative interview techniques. An
important assumption of the indicator is that CSWs are equally likely to underreport high risk behaviours than others in the sample.

4.3 External validation: Ethnographic techniques

A second strategy for validation is to conduct in-depth investigation of behaviour on focused subsamples of larger survey studies. In-depth methods rely on intensive investment in rapport building to overcome the normal problems with motivation and social desirability biases in reporting sensitive behaviours. Trust and respect that is developed through longer, less structured, sometimes multiple visit interviews is viewed as the key difference in gaining access to such socially privileged information. The two main approaches are semi-structured interviews (SSI) and diary methods. Semi-structured interviews cover topics in a systematic fashion but allow the interviewer greater flexibility in determining how questions are asked and wider range to explore narrative prose data in addition to more quantifiable outcomes (Kikwawila study group 1994). The common assertion that less structured interview techniques yield improved validity has some empirical support. In the Konings study cited above, the validity of three measurement techniques for estimating the prevalence of high risk sexual behaviours was compared - short form questionnaires, long form questionnaires, and semi-structured interviews. Between 50 and 59 percent of the subsample of CSWs reported multiple partners in the last 4 weeks in long and short form questionnaire studies, compared to 87 percent for the in-depth interviews. Assuming all CSWs had multiple partners in the last 4 weeks, SSIs still underestimate true risk behaviour, but are a significant improvement on either survey method.

Another ethnographic approach involves collecting diary accounts of sexual or drug use behaviours. Information regarding risk events is recorded on a daily or weekly basis over a limited period of observation. Optimally information collected on diary forms is kept to a minimum, so a daily record requires only a few minutes to complete. Diaries potentially address the two main sources of error in report of sensitive behaviours. Recording events on a daily or even weekly basis diminishes potential for recall error. Regular contact with research staff through diary training or periodic data collection allows rapport building and reinforcement of motivation to participate that in turn reduce temptations to bias responses towards socially acceptable standards. Depending on the mode of data collection, diaries can
privacy, confidentiality, and convenience for respondents. The disadvantage of diaries is that they require high levels of motivation on the part of respondents and a potentially high level of investment of research staff time to monitor or collect diary accounts. Dropout rates are a problem with this approach, and times of observation need to be limited (Catania, Chitwood et al. 1990). High levels of literacy are required for self administered diary designs, although diaries using face-to-face interview methods have been successfully employed in low literacy settings through regular visits by local interviewers (Pickering, Okongo et al. 1997).

Methodological studies have shown that diaries improve reliability of reporting on sexual behaviour. In a study of 62 married couples in Senegal that involved 5 weekly diaries followed by a retrospective survey, Lagarde et al. (1995) found that couple agreement on sexual frequency was significantly higher in the weekly diaries than the retrospective survey. Weekly surveys effectively eliminated the commonly observed gap in reported coital frequency between men and women from the retrospective survey. Number of extramarital partnerships was increased in weekly reports, but the exact results are not reported in the article. Similarly, Hornsby (1989) found that diary methods corrected coital frequency estimates reported by 91 women from an average of 2.5 times per week in retrospective reports to 1.8 times per week reported in diaries. Leigh (1998) found diaries improved reporting of drinking behaviour and coital behaviour in adolescents, but made no significant difference for adults.

4.4 External validation: Biomarkers

The third category of external validation involves collection of indirect evidence of behaviour that explicitly do not rely on self report. An imaginative array of measures have been attempted in different locations. One example is monitoring condom sales and supplies in the observed population as an indirect measure of changes in protective behaviour which was proposed as one of 10 prevention indicators in a WHO protocol to measure HIV/AIDS prevention related indicators (Global Programme on AIDS, 1994). Sewer traps have been proposed to obtain an indicator of condom use(Catania, Chitwood et al. 1990).

STD biomarkers offer an increasingly attractive validation method for sexual risk behaviour with the development of new techniques that are relatively inexpensive and practical in field
settings. On a population level, changes in STD rates have been used to document shifts in underlying sexual behaviour. The rise of HIV rates in early stages of the epidemic were marked by similar increases in anorectal gonorrhoea, syphilis, and enteric infections among homosexual men in the US (Quinn 1996). More recently sharp declines in a variety of STDs have been attributed to the aggressive implementation of a national condom campaign in Thailand (Hanenberg, Rojanapithayakorn et al. 1994). In very high prevalence settings such as those observed in the Mwanza trial in Tanzania, HSV-2 infection is found in the majority of sexually active adults and is being used to monitor adolescent transition to sexual activity (Obasi, Mosha et al. 1999). On a smaller scale, STDs have been used to validate self reported behaviour. HSV-2 (genital herpes) infection has been found to be significantly correlated with years of sexual activity, lifetime sexual partners, history of past STDs, and has been proposed as good serological marker of high risk sexual lifestyle, particularly among women (Cowan, Johnson et al. 1994). Reductions of anal gonorrhoea rates correlated well with accounts of sex behaviour among gay men in San Francisco (Coates, Stall et al. 1989). Among a sample of high risk urban adolescents in the U.S., the likelihood of new infection with gonorrhoea, chlamydia, and trichomonis vaginalis correlated perfectly with reports of complete sexual abstinence, but poorly with level of sexual activity among those who reported any sex (Orr, Fortenberry et al. 1997). In a study among young STD clinic attenders, level of reported condom use had virtually no correlation with incident infection in the last month (Zenilman, Weisman et al. 1995). While the possibility of incorrect condom use or misclassification of STD reports is recognized, the more likely explanation is recall error or social desirability pressures to overreport condom use.

Using STDs as a biomarker for sexual behaviour has notable limitations. Presence or absence of infection is a poor proxy for risk behaviour at an individual level. As with any disease, likelihood of STD infection is a complex function of host, disease agent, and environmental factors. As Pickering et al. found in their Ugandan studies, communities with high rates of partnership change and other high risk behaviours for transmission of HIV have remained insulated from the epidemic by lack of exposure to outside sexual networks (Pickering, Okongo et al. 1997). At the opposite extreme, biomarkers potentially lack sensitivity to exposure levels in the highest risk groups. Those infected multiple times within an observation period will be indistinguishable from those infected only once. Empirical research has demonstrated a curvilinear relationship between STDs and multiple sex partners.
(Catania, Chitwood et al. 1990). Practically, the utility of a biomarker component is limited by the need for either high transmission rates or very large numbers (Quinn 1996). For treatable infections, monitoring requires additional collection of detailed information on treatment behaviour in the reporting period, introducing yet another layer of potential reporting error and bias.

Given these shortcoming, the applicability of biomarkers as external validation for HIV risk behaviours needs to be evaluated in each setting.

**How essential is a biomarker in the presence of ethnographic validation measures?** The answer to this question depends on the level of faith placed in self reporting methods and the intensity of bias, which include ethnographic approaches themselves. In a controlled trial with any possibility of differential reporting patterns developing from the intervention between vaccine and placebo arms, then the justification for biomarkers becomes compelling. In the absence of a placebo vaccine, they become essential. The lack of a simple one-to-one relationship between risk behaviour and STD incidence means that biomarkers can only be interpreted on the group rather than individual level. Consequently, biomarkers may augment but should not replace ethnographic validation measures.

**Are transmission rates sufficiently high to justify a biomarker study?** Transmission of the candidate STD must be high enough to detect behaviour change at statistically significant levels within the anticipated sample size. Orr et al. (1997) based their validation study on as few as 255 women, but in a setting where nearly one third of sexually active women had acquired a new STD within 3 months of enrolment. The chances of sufficiently high transmission rates should be improved by the high risk selection criteria for Phase III vaccine trials.

**Which STD should be adopted as a marker?** The general prevalence patterns in the study population needs to be established in advance as an indicator of levels of exposure to candidate STDs. As seen in Figure 2 below the burden of different STDs varies strongly by region. Vaccine trials in eastern Europe, for example, would require much larger samples for a biomarker study than trials in most sub-Saharan African settings, and may not be feasible. For a prospective risk monitoring study in a high risk population as envisioned for a Phase III
vaccine trial, the possibility of reinfection with the candidate STD is important. HSV-2 would not be a suitable for monitoring in a population such as Mwanza, Tanzania where over 75 percent of women over age 25 are already infected (Obasi, Mosha et al. 1999). As discussed in the practical protocol, three of the most common curable STDs on a global scale – Neisseria gonorrhoeae, Chlamydia trachomatis, and Trichomonas vaginalis - can now be tested with high sensitivity and specificity from the same first void urine sample (Davies, Low et al. 1998).

*Can the expense and logistical requirements of STD testing be accommodated in the research design?* Infrastructure to allow testing is a significant consideration. Even new testing techniques for STDs based on first void urine tests require processing, storage, transport, careful documentation, and quality assurance. Ethical obligations to confirm diagnoses, inform and provide treatment for positive tests must be considered.

Figure 2: Estimated incidence of STDs (millions) in people aged 15-49 years for 1995

<table>
<thead>
<tr>
<th></th>
<th>Syphilis</th>
<th>Gonorrhoea</th>
<th>Chlamydia</th>
<th>Trichomoniasis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>North America</td>
<td>0.072</td>
<td>0.072</td>
<td>0.83</td>
<td>0.92</td>
</tr>
<tr>
<td>Western Europe</td>
<td>0.10</td>
<td>0.10</td>
<td>0.60</td>
<td>0.63</td>
</tr>
<tr>
<td>Australasia</td>
<td>0.005</td>
<td>0.005</td>
<td>0.063</td>
<td>0.069</td>
</tr>
<tr>
<td>Lat. America &amp; the</td>
<td>0.56</td>
<td>0.70</td>
<td>3.45</td>
<td>3.67</td>
</tr>
<tr>
<td>Caribbean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>1.56</td>
<td>1.97</td>
<td>7.30</td>
<td>8.38</td>
</tr>
<tr>
<td>N. Africa &amp; Middle East</td>
<td>0.28</td>
<td>0.33</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>E. Europe &amp; Central Asia</td>
<td>0.050</td>
<td>0.050</td>
<td>1.17</td>
<td>1.16</td>
</tr>
<tr>
<td>East Asia &amp; Pacific</td>
<td>0.26</td>
<td>0.30</td>
<td>1.80</td>
<td>1.47</td>
</tr>
<tr>
<td>South &amp; SE Asia</td>
<td>2.66</td>
<td>3.13</td>
<td>14.56</td>
<td>14.55</td>
</tr>
<tr>
<td>Total</td>
<td>5.55</td>
<td>6.66</td>
<td>30.54</td>
<td>31.62</td>
</tr>
</tbody>
</table>
Part II. PRACTICAL PROTOCOL

5.0 Introduction

Carballo et al. advocated the use of face to face interviews using fixed questionnaires as the best method of estimating sexual behaviour on a population level (Carballo, Cleland et al. 1989). The subsequent decade of experience and applied methodological research has expanded the variety of ways in which questionnaires can be administered, and refined way to construct questionnaires, but it has not reduced reliance on sample survey techniques to estimate sexual behaviour for large populations. For all the error and biases known to occur in the collection of sensitive subjects through this deliberately impersonal approach, surveys still offer the only practical approach to monitoring behaviour in large populations. At the same time the need for parallel validation measures remains equally strong. A combined approach offers the best way to obtain valid and reliable estimation of levels and trends of HIV risk behaviours.

5.1 Overview of monitoring strategy

Four phases of data collection are proposed to provide the most complete and accurate picture of risk behaviour in the trial population. Each phase is associated with the different data collection method. A prototype instrument for each phase that involves interviews is presented here. The four phases, associated data collection method, and target population is listed below.

<table>
<thead>
<tr>
<th>Data collection phase</th>
<th>Data collection method</th>
<th>Target group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formative phase</td>
<td>1) Focus group interviews</td>
<td>Selected informants from high</td>
</tr>
<tr>
<td>Risk groups</td>
<td>2) STD prevalence/incidence estimation</td>
<td>Secondary data if possible</td>
</tr>
<tr>
<td>Behaviour monitoring phase</td>
<td>Repeat survey questionnaire</td>
<td>All trial participants</td>
</tr>
<tr>
<td>Validation phase</td>
<td>1) Sexual and IVDU diaries</td>
<td>Random subset of survey pop.</td>
</tr>
<tr>
<td></td>
<td>2) STD biomarkers</td>
<td>Random subset of survey pop.</td>
</tr>
<tr>
<td>In-depth phase</td>
<td>In-depth interview</td>
<td>Subset of diary sample</td>
</tr>
</tbody>
</table>
6.0 Formative phase

A formative phase is optional but advisable to improve the quality of risk monitoring and validation instruments. Two methods of data collection are proposed. Focus group interviews among the high risk groups can assist the construction of the risk monitoring instrument. For the design of the validation phase, it is important to establish the prevalence of STDs among high risk groups in order to select the appropriate biomarker.

6.1 Focus group discussions

6.11 Objectives

A small number of focus group interviews prior to the onset of a Phase III trial can serve several useful purposes. These include:

- Identify high risk groups in the community
- Assist in location and recruitment strategies for high risk groups
- Document normative patterns of high risk behaviour in different groups
- Document language variation and local terminology for risk behaviours among the study population

6.22 Rationale

Focus group discussions (FGD) are one of a number of rapid assessment techniques to serve as a form of proxy informant on community behaviours, norms, and attitudes. Focus groups typically involve 6 to 12 participants in a general discussion of broad guideline points led by a trained moderator. With the atmosphere of trust and shared experience in a well conducted group, FGDs are particularly suited to eliciting information on sensitive subjects such as sexuality and drug use, provided that participants are not obliged to report on their own
behaviour but rather discuss general patterns of behaviour in the surrounding community (Krueger 1994).

FGDs serve two overall objectives in the risk behaviour monitoring study. The first, summarised in the first two items in the summary list above, is essentially a marketing strategy to improve coverage of high risk populations enrolled in a Phase III trial. Initial groups recruited from obvious high risk populations such as STD clinics or drug rehabilitation centres can identify other high risk groups in the population and develop strategies to locate and recruit high risk subjects into the study. The second main objective is to use groups to confirm or improve the content and wording of the risk monitoring instrument. Participants can provide an overview map of risk behaviours in the community which may suggest alteration or addition to the monitoring instrument. Of central importance to the validity and reliability of self report is the clarity of the terminology used in the questionnaire to the respondents. Focus groups can contribute importantly to building a lexicon of local or ‘poetic’ terms for high risk behaviours that can be substituted or preferably added to the formal scientific terms in the instrument. For example, the HRBS instrument adopts the term ‘hit up’ instead of inject drugs, although the latter term is retained parenthetically to ensure standard definition. Use of the poetic terms has been shown to improve reporting of sensitive behaviours (Cantania et al. 1990, 1993; Weinhardt, Forsyth et al. 1998).

6.23 Study design

A small number of focus groups should be conducted to represent the key high risk groups for HIV in the proposed trial setting. The two essential criteria for participation in FGDs are experience of the specific high risk behaviour and ability to contribute freely to a group discussion. Both criteria can be defined broadly and somewhat subjectively, taking a number of factors such as convenience and personality traits into consideration. Random selection or strict representativeness should not be a consideration. For example, discussions with men and women should be conducted separately where cultural precedent makes it difficult for women to contribute equally to discussion of a particular subject in the presence of men. Other criteria may be added to enhance the homogeneity of the group and its ability to provide targeted information on specific groups. Initially four groups should be tentatively
planned - two for high risk sexual behaviour and two for high risk drug behaviour. Additional groups can be organized as necessary.

6.24 Recruitment

Once criteria for FGD participation are established, recruitment can take a variety of strategies. If eligibility is defined by presence in an STD or drug rehabilitation clinic, for example, simple face-to-face recruitment is possible without preliminary screening. If eligibility involves more criteria, participants may be selected through brief screening forms administered in advance.

6.25 Requirements

Focus group discussions can be conducted in a wide variety of settings with few personnel or material requirements. These include:

- trained focus group moderator
- notetaker to tape record the session and keep a record of individual contributions to the discussion
- transcriber to transcribe the tape after the discussion for analysis - the notetaker is the ideal candidate to transcribe.
- focus group guideline - a set of 6-10 general questions to stimulate discussion
- tape recorder and supplies
- meeting space - preferably someplace free from outside interruption and quiet enough to make tape recorded discussion audible.

6.3 Formative research on STD prevalence

The selection of an appropriate biomarker to validate shifts in high risk sexual behaviour will depend in part on the incidence of specific STDs in different Phase III study populations. This information may be obtained through secondary data such as hospital or clinic records in
the study site or previous STD studies in comparable populations. In the absence of secondary sources, a small scale study to indicate the prevalence of candidate STDs among high risk groups for a biomarker may be warranted.

7.0 Monitoring Survey

7.1 Objectives

- provide the valid and reliable indicators of HIV risk from sexual behaviour and IV drug use in a high risk population.
- provide a readily interpretable, easy-to-calculate summary indicator of HIV risk behaviour to monitor trends over time.
- allow summary score to be disaggregated into component indicators to monitor separate categories of risk behaviour over time.
- collect information in the fewest number of questions possible to reduce the burden of repeat surveys for large sample sizes.

7.2 Sampling strategy

The risk behaviour monitoring survey should be administered to all participants of Phase III trial. Saturation sampling should remove any concern about reaching sufficient statistical power to monitor risk behaviour. For example, to detect a 5 percent change in the prevalence of a specific risk behaviour from a baseline level of 20 percent, assuming 95 percent confidence intervals and 90 percent power, would require a cohort size of approximately 1000. It can be safely assumed that the sample sizes to detect the minimum change in HIV transmission required of a Phase III trial would be much greater. Since the risk monitoring instrument is brief and may be administered in conjunction with vaccination visits, there is no strong reason to restrict monitoring to a subsample. (NB. This assumes that control groups will be following a similar visit schedule as the vaccine group.)

7.3 Monitoring instrument

The objectives of monitoring risk behaviour for vaccine trials pose contradictory demands for construction of the instrument; It must capture enough detail to provide valid and reliable
estimates of risk behaviour in a relatively large population required for a Phase III trial. At
the same time it must be brief and easy to administer to accommodate repeated measures of
large samples. The prototype presented here attempts to meet both demands. In addition,
coding categories are designed to be additive to allow calculation of simple summary
indicators of HIV risk at the end of each interview. Averaged over samples, this should allow
simple analysis of time trends to meet the most important objective of the survey. More
detailed analysis strategies are presented in a following section.

The strategy of the questionnaire design is to construct summary indicators of HIV risk from
a fixed number of recent sexual partnerships or drug events. The questionnaire is divided into
two sections, one for sexual risk behaviour, the other for recent drug events. For sexual
behaviour, a series of 6 questions are repeated for the five most recent partners in the last
month besides a formal spouse. (Depending on the level of risk in the population and the
pressure to keep the questionnaire short, the instrument may be easily adapted to
accommodate partnerships in the last 6 months.) Summary measures of total number of
partners in the last 6 months, last one month, and number of formal spouses Condor use
with a normal spouse or partner is included to track
A. Sexual risk behaviour questionnaire

To be read out loud or included in a taped prologue at the beginning of the interview:
It is very important for the success of the vaccine trial to monitor behaviours that affect the spread of AIDS. The following questions ask about sexual partnerships and IV drug use since the last interview. Please try to answer them as honestly as possible. Please do not be offended by questions that do not apply to you.

<table>
<thead>
<tr>
<th>Category and question</th>
<th>Coding categories and scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION: Tell me about the most recent sexual partner you had besides your husband\wife?</td>
<td></td>
</tr>
<tr>
<td>FOR PREVIOUS PARTNERS: Tell me about the last partner you had before that?</td>
<td></td>
</tr>
</tbody>
</table>

FILTER: Recency of partnership
How long ago did you last have sex more than 6 months ago A-SKIP S8
(PROBE FOR A OR B: Please think carefully. within last 6 months B-SKIP S8
Have you ever had sex with any other partner last month C-NEXT
in the last month besides your husband\wife?

S1. Duration of partnership
How long ago did you start sexual relations more than 6 months ago 0
with this person? Within last 6 months 1
Within last month 2

S2. Sex of partner (MEN ONLY)
Was this a man or a woman? Woman 0
Man 1

S3. Age range
Is this partner closer to age 20, age 35, or age 50? 20 0
35 1

S4. Condom use
In the last month, how often did you use a condom
with this person? (PROBE FOR ALWAYS: Was Always 0
there ever a time in the last month when you did not use a condom?)

Sometimes 1

Never 2

S5. Anal sex

In the last month have you ever had anal sex with this partner with or without a condom?

No 0

With a condom 1

Without a condom 2

S6. Commercial sex

Have you ever exchanged money or goods for sex with this person, or has this person paid you for sex?

No 0

Yes 1

RETURN TO TOP (UNTIL 5 PARTNERS OR PREVIOUS PARTNER MORE THAN 1 MONTH AGO)

S7. Altogether how many different sexual partners have you had in the last 6 months, including clients or commercial sex workers?

IF UNCERTAIN: Between _____ and _______

S8. Altogether how many different sexual partners have you had in the last month?

IF UNCERTAIN: Between _____ and _______

S9. How many of these are either a husband\wife or someone you live with on a regular basis?

S10. In the last month, how often have you used a condom with your spouse(s) or partner(s) you normally live with?

Always 0

Sometimes 1

Never 2
B. Intravenous drug use

INTRODUCTION: When was the last time you ‘hit up’ (injected a drug) in the last month?

FOR PREVIOUS DAY OF DRUG USE: When was the most recent day you injected drugs before that?

D1. What were you injecting that day?
(Scores for different drugs?)

D2. How many times did you inject that day?

D3. How many people injected before you that day?
(record exact number, or record range)

D4. Was the needle cleaned before you used it?
(new needle - 0, always cleaned before - 1, sometimes cleaned before - 2, never cleaned before - 3)

D5. How was the needle cleaned before you used it?
(never shared = 0, bleach\boiled = 1, other cleaning = 3, not always cleaned = 4)

RETURN TO TOP (X 5)

D6. In the last month, how often did you inject?
(never, less than once a week, at least once a week, once a day - 3, two to three times a day - 4, more than 3 times a day - 5).

D7. In the last month, how many times have you injected with other people?
(never-0, rarely\sometimes-1, about half of the time-2, frequently\most of the time -3, all the time -4).
The extent to which the summary indicators are valid and reliable indicators of true risk of HIV depend on a set of underlying assumptions that guided construction of the questionnaire. The first key assumption is that Phase III trial subjects are more highly motivated and more willing to report high risk behaviours a typical respondent randomly selected into a household survey. This assumption is based on selection criteria of Phase III trials, specifically 1) all participants are required to be volunteers, 2) eligibility depends on willingness to report high risk behaviours, and 3) after the first round all participants will have answered the questions before and will know what to expect. In the interest of brevity, then, the questionnaire dispenses with ‘rapport-building’ questions and asks a range of highly intimate questions about recent partnerships in a direct manner. An introductory statement is designed to enhance motivation to respond accurately and minimize potential embarrassment by reassuring respondents that many questions will only apply to the highest risk categories of behaviour.

The second assumption is that a truncated history of partnerships and days of drug use will not affect the sensitivity of summary indicators to real changes in HIV risk at the population level. In the case of sexual behaviour, limiting partnership history at five partners in the last month should affect only a small proportion of the highest risk groups. In the case of IV drug use where the retrospective approach captures a maximum of 5 days out of the last month, the summary measure should act more like a random sample of drug use that would not threaten validity. For both sexual behaviour and drug use, total counts of high risk events in the last month are included into the summary indicator score to reflect high risk counts.

### 7.4 Interview schedule and mode of administration

The risk behaviour monitoring survey will be administered at the vaccination visit that occurs on the 28th day after the 6 month interval visit. The purpose of this timing is to allow validation with one month diary forms that cover the same month observation period.

Mode of interview will depend on the location and facilities available at each Phase III test site. It is well established that the validity of reporting sexual behaviour is improved by increasing confidentiality of information. The short, factual nature of the questionnaire
would benefit from a self administered form completed at the time and location where vaccination takes place. In low literacy situations where self administered questionnaires are not feasible, face-to-face interviews will be required. Advocates of computer-assisted interview techniques claim that this approach can be implemented in low literacy situations (Turner et al, 1998) but this still assumes a minimum level of infrastructure and conceptual familiarity if not facility at reading computer screens and operating a computer keyboard.

Personnel requirements depend on the mode of interview selected. For low literacy populations, face-to-face interviews would be required. At a fixed site with a steady flow of respondents, a trained interviewer should be able to complete 15 to 25 face-to-face interviews a day, depending on the level of risk behaviour in the groups. The total number of interviewers required for each site can be calculated accordingly. For high literacy populations, self administered questionnaires would be preferable on validity and reliability grounds as well as eliminating the need for full time interviewers. Personnel time would be required to train participants at enrolment on correct completion of questionnaires, and to provide quality checks.

7.5 Analysis

The monitoring instrument is constructed to allow calculation of simple summary scores for HIV risk behaviour in the last month. Each item response is given a numerical score proportional to the level of HIV risk. Scores can be calculated by means of simple addition of item responses for each column. Summing across each item total yields summary scores for sex and drug use respectively. A total risk score for each individual is calculated from the sum of sex and drug use scores. Analysis of levels and trends at the individual level will require data entry and analysis in a statistical package, but single individual scores and group averages over time may be tracked with no more than a hand calculator if necessary or desired.

A. Summary measures
Total risk score - sum of sex risk score and IDVU risk score
Sex risk score - sum of sexual behaviour items
IVDU risk score - sum or intravenous drug use score

B. Sexual behaviour item scores

The sex risk score is designed to provide a single summary indicator for risk of HIV from partnership behaviour in the last month. The following component indicators from the sex and IVDU risk scores may be independently monitored. To test for interaction of risk factors, a list of partner-specific or event-specific outcomes are also included in each section.

Total sex partners in the last 6 months
Total sex partners in the last month
Number of regular partners in the last month

The following item scores are constructed from information about the last 5 partnerships in the last month.

Partnership duration score
Age mixing score
Condom score
Anal sex score
CSW score

C. IDVU item scores

The construction of an overall intravenous drug use score and component risk scores follows a similar procedure.

Overall IVDU score
Needle score
Sharing before score
Sharing after score
Cleaning score
Post sex score
8.0 Validation

Two validation measures are proposed to check for the anticipated biases in survey reports of sexual behaviour and intravenous drug use. Diaries for a small subset of respondents are proposed as the anthropological measure best suited to address the dual sources of error in self reports from recall error and social desirability bias. STD biomarkers are proposed to provide an external measure of validation for self reports. Justification for two validation measures is reviewed in Part 1 of this protocol. Field procedures are discussed below.

8.1 Monthly diary

While it is intensive and time consuming, the diary method provides the ‘gold standard’ for self reports by addressing the two main threats to validity and reliability of self reported behaviour. Closer contact to the study, especially for oral diary methods involving weekly visits, favours increased motivation to participate and stronger rapport between researchers and subjects. Both factors should reduce social desirability bias for self reported behaviour. The other critical source of error is poor recall, which is addressed through regular recording of events in diaries. The disadvantage of diaries is that they are labour intensive, particularly for oral diaries collected by mobile interviewers, necessarily restricted to small samples, and prone to high dropout rates. Designed correctly, they offer an excellent measure for validation of larger scale survey reports.

Sample size of diary reports depends on the number of subpopulations for which validation is desired and whether diaries are orally collected by interviewers or maintained by respondents themselves. A rough guide of 30 to 50 diaries per subgroup would yield stable estimates upon which to qualitatively judge the validity of comparative questionnaire reports. Greater numbers are possible with self administered mode of sexual and drug diaries. A recent study of sexual networking in Uganda provides a useful model of diary methods in a low-literacy setting (Pickering, Okongo et al. 1997). Trained interviewer visited between 20 to 30 informants, making with 3 to 4 visits per week on an informal, unscheduled basis. Interview times were notably brief, often under 5 minutes (Pickering, personal communication). For the self administered diary approach, subjects received standard debriefing to insure quality of
data, make necessary corrections, and generally reinforce motivation. Prototype diary forms with coding categories are presented below.

<table>
<thead>
<tr>
<th>Sex Diary</th>
<th>Start date:</th>
<th>Stop date:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Partner Name/ID</strong></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Column #1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Monday</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Date:</strong></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Recorded:</strong></td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Extra</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tuesday</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Date:</strong></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Recorded:</strong></td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Extra</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wednesday</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Date:</strong></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Recorded:</strong></td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Extra</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Etc.…

**Coding categories**

**Name / partner ID:** Unique identifier of partner each partner

**M/F:** - Sex of partner. (M = Male, F = Female)

**#** - Total number of episodes with each partner

**Action codes** -

- V - Vaginal sex (y/n)
- A-G - Anal give (y/n)
- A-R - Anal receive (y/n)
- O-G - Oral give (y/n)
- O-R - Oral receive (y/n)

**Condom:** Condom use with partner (a=all episodes, s = some episodes, n = not used)

**Payment:** Was any payment given or received for sex? (g = give, r = receive, n = no)
<table>
<thead>
<tr>
<th>Column #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date:</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded:</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuesday</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date:</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded:</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wednesday</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date:</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded:</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thursday</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date:</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded:</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Etc….

Coding categories

**Injected:** Substance injected for specific episode

**Needle type:** Needle source (n = new, u = used, pp = prepackaged with substance at purchase)

**Sharing:**

Y/N – Inject with others (y) or alone (n).

# before: number who used the same needle before subject injected

**Cleaning method:** How was needle cleaned before subject used? (nn = new needle, bl = bleach, bb= boiled, ww = water + non-bleach cleaner, oc = other method, dk = don’t know, not = not cleaned)

**Sex after:** Engage in sex with injection partners?

**# of partners:** number of different partners after injection episode.
STD biomarker

STD biomarkers provide a community-level validation measure of changing risk behaviour reflected in changing patterns of STD transmission. It is important to establish in the formative phase of research whether base levels of STD transmission are high enough to make this practical within anticipated sample sizes. Other considerations include choice of biomarker, sample size, and treatment monitoring.

8.21 Choice of biomarker

The suitability of biomarkers and choice of STD to monitor should be determined during the formative phase of research based on several main considerations. The first is the prevalence and transmission rate of candidate STDs in the study population or comparable group for which information exists (see Figure 2 in Part 1 of protocol). Higher exposure to the candidate STD reflected in prevalence and incidence rates will increase the sensitivity of biomarker to changes in sexual risk behaviour. Potential for reinfection is desirable for monitoring recent behaviour rather than accumulation of lifetime risk, which should be uniformly high in a Phase III population.

Globally, the candidates for common, curable STDs in ascending order of incidence include syphilis, gonorrhoea, chlamydia, and trichomoniasis (Gerbase, Rowley et al. 1998). Incidence levels and priority ordering vary across regions. Even though chancroid is thought to be the most important cause of genital ulceration in tropical regions (Mindel 1998), it is excluded because transmission can occur through non-sexual means (ref?).

Choice of biomarkers also depends on testing considerations. The range of options available for testing STDs has significantly expanded through the development of molecular diagnostic techniques such as polymerase chain reaction (PCR) and ligase chain reaction (LCR) (Mayaud and Mabey 1998). Gonorrhoea, chlamydia, and trichomoniasis can now be tested from urine samples with equal if not superior specificity and sensitivity to the former ‘gold standard’ methods based on blood tests (Davies, Low et al. 1998). Urine specimens are awkward to handle in the field and require centrifuging to reduce the sample. Once processed, however, samples are relatively insensitive to handling, remaining stable at room
temperature up to 24 weeks. The same technology allows combined (multiplex) assays for multiple target STDs from the same clinic sample. Given the high prevalence of all three conditions in high risk populations of both men and women, a urine sample collected every 6 months for three biomarkers is an attractive option a validation strategy.

The gold standard test for syphilis still requires blood specimens, although this may represent less of an impediment if STD testing can be incorporated into normal HIV testing routines in the field trial. Syphilis testing has the added capability of detecting reinfection with new strains (Van Dyck, Behets et al. 1996).
8.23 Sample size

The sample size required will depend on the incident rate of the candidate biomarker in a normal 6 month observation period between test visits. Given the high risk nature of a Phase III trial and the direct role of STDs in transmission of HIV, incorporating STD monitoring into HIV screening routines for the full sample may be warranted. Otherwise additional specimens should be collected from a random subsample selected at the beginning of the trial and followed throughout. Exact sample size calculations will depend on incidence calculations.

All subjects involved in the diary validation should be included in the biomarker sample.

8.24 Other considerations

To interpret STD trend from a sample or the full population, it will be important to know the history of any treatment since the last test. Information on treatment in the last 6 months should be gathered at the time specimens are collected. Questions may be limited to a few:

1) Medical treatment for STD in the last 6 months?
2) Treatment received
3) Treatment completed?
4) Antibiotic treatment for any other condition in the last 6 months?

9.0 In-Depth Interview

The united aim of the methods and instruments recommended above - repeat surveys, daily behavioural diaries, and collection of blood and urine specimens to test for STD infection - is to describe actual risk behaviours as accurately as possible. The more precise picture of shifting patterns of risk behaviour in the population can be captured, the more precisely change in transmission rates of HIV can be attributed to the effect of the vaccine. However none of the behavioural measures can help to explain the motivations behind behaviour
change. A critical question for the vaccine trial is how potential changes in risk perception from participation in a vaccine trial will impact on actual behaviour. The complexity of risk perception and multilayered motivations for behaviour require more in-depth qualitative approaches. Individual in-depth interviews provide an appropriate tool for this task.

9.1 Objectives

- To gain understanding of the nature of risk perception among subjects in high risk populations.
- To explain motivations for changes in risk behaviour as a result of participation in a vaccine trial.
- To validate sexual behaviour reports among a small subset of the trial population.

9.2 Interview technique

In-depth interview is a standard tool in ethnographic research. The logic and structure of IDIs reflect the fundamental difference between ethnographic approaches to gaining knowledge and more objectively oriented approaches implicit in much sample survey research. The objective of a sample survey is to understand behaviour by gathering selective, factual, often quantifiable information that are interpretable within a logical framework imposed by the researcher. The objective of in-depth qualitative techniques is to gather a more detailed, less selective account of individual behaviour from the respondent’s own perspective. The two approaches overlap in their interest in describing behaviour, and in-depth approaches have been used for this purpose to validate survey estimates (refs!). However the emphasis on subjective meaning of behaviour ultimately differentiates ethnographic techniques from sample survey techniques and determines qualitative study design.

If fixed questionnaires simulate controlled laboratory measurement with human subjects, in-depth interviews aspire to be guided conversations where the respondent has more control over the process(Kvale 1996). Interviews are semi-structured. The general objectives of discussion are established by the interviewer at the beginning of each session. Within these
loose constraints, the structure of conversation, the level of detail, the duration of the interview and the course of the discussion may vary from interview to interview. An in-depth questionnaire normally consists of general themes, questions and probes intended to initiate or guide the flow of discussion without constraining it to specific answers. In short, much depends on the reaction of the respondent which cannot be foreseen.

The quality of in-depth interviews depends on a basic rapport and trust between interviewer and interviewee. The design of in-depth interviews should actively build such rapport. The first essential ingredient is time. Some manuals advocate dedicating the first interview to general undirected discussion to establish a foundation of trust (Spradley 1979). Another technique is to start an interview by asking the interviewee to recount his life story or a particular passage in life in order to build a bond of common interest before proceeding to areas of substantive interest (Kvale 1996). During the interview, rapport is enhanced by maintaining a reasonable balance of power between interviewer and respondent over the direction and content of discussion within the framework set out at the beginning. The more trust is established, the more comfortable the respondent becomes, and the more truthful and complete the data are presumed to be. The resulting advantage of in-depth approaches is to insulate data from the normal social desirability biases that plague more impersonal interview styles.
9.3 Semi-structured questionnaire

Themes
1) Evolution of knowledge about HIV (first interview – rapport building and establish theme)
   - How did you first learn about HIV?
   - Where does HIV come from?
2) General risk perception (also first interview)
   - Do you know anybody who has HIV or who has died from AIDS?
   - How did that affect you?
   - Have you had any health problems in your life?
   - Did any of those affect your behaviour? Why/why not?
2) Detailed account of partnership behaviour in last 6 months (later interviews)
   - behaviour
   - perception of risk
3) Detailed account of drug use in last 6 months
   - behaviour
   - perception of risk
4) Perceived effect of participation in last 6 months on behaviour

9.4 Sample

The time invested in rapport building, semi-structured interviewing, and analysis of textual data effectively limit the size of in-depth studies to small numbers of subjects. The numbers required depends on the objective of this component of research. To document change in risk perception in a qualitative method, no statistical criteria for sample size need be applied. Rather, ‘saturation’ methods may be adopted where the number of respondents and the number of interviews should be limited to the level where additional numbers do not contribute any new information (Bernard, 1994). Upper limits are set by the time for interviewing and textual analysis that can be devoted to this task. A minimum of 10 interviews per subgroup should be considered. If the purpose of in-depth interviewing is to validate sexual behaviour reports, a statistical minimum of 30 observations per subgroup should be applied.
Selection of subjects for in-depth interviews should be made purposively to sample key risk categories or individuals of particular interest. At the beginning of the vaccine trial, an initial set of informants should be selected to represent the spectrum of high risk behaviours. Rolling sample design may be achieved by interviewing informants every 6 months for 18 months. New recruits can be selected among those who demonstrate significant reductions or increases in risk behaviours within broader risk categories.

**References**


