

EVALUATING PROGRAMS FOR HIV/AIDS PREVENTION AND CARE IN DEVELOPING COUNTRIES

A HANDBOOK FOR PROGRAM MANAGERS AND DECISION MAKERS

EDITED BY THOMAS REHLE, TOBI SAIDEL, STEPHEN MILLS, AND ROBERT MAGNANI



FAMILY HEALTH INTERNATIONAL



FUNDED BY THE UNITED STATES AGENCY FOR INTERNATIONAL DEVELOPMENT

EVALUATING PROGRAMS FOR HIV/AIDS PREVENTION AND CARE IN DEVELOPING COUNTRIES

A HANDBOOK FOR PROGRAM MANAGERS AND DECISION MAKERS

EDITED BY THOMAS REHLE, TOBI SAIDEL, STEPHEN MILLS, AND ROBERT MAGNANI
WITH THE ASSISTANCE OF ANNE BROWN RODGERS

© 2001 Family Health International (FHI). All rights reserved.

This book may be freely reviewed, quoted, reproduced or translated, in full or in part, provided the source is acknowledged. This book may not be sold or used in conjunction with commercial purposes.

ISBN 1-931547-01-07

The views expressed in this book by named authors are solely the responsibility of those authors and do not necessarily reflect the views of FHI or the United States Agency for International Development (USAID). The mention of specific companies or products implies no endorsement and does not suggest that they are recommended by FHI or USAID over others of a similar nature that are not mentioned.

THIS HANDBOOK HAS BEEN FUNDED BY THE UNITED STATES AGENCY
FOR INTERNATIONAL DEVELOPMENT (USAID) THROUGH FAMILY HEALTH
INTERNATIONAL'S IMPLEMENTING AIDS PREVENTION AND CARE (IMPACT)
PROJECT, COOPERATIVE AGREEMENT HRN-A-00-97-00017-00.

FOREWORD

Evaluation is too often an afterthought in the process of program implementation.

This Handbook is dedicated to the premise that evaluation must be a critical part of the initial phases of planning effective HIV/AIDS prevention and care programs.

Readers of this Handbook will find that the authors have set the stage and provided the tools for a comprehensive and strategic approach to evaluation. The approaches they recommend yield useful and important information on the effectiveness of HIV/AIDS prevention, care, and support programs. Increasingly, the technical “know how” to develop effective interventions is available but the ability to demonstrate effectiveness is lacking. Moreover, the ability to increase and maintain global political will and resources directed to HIV/AIDS programs ultimately rests on our ability to show that our interventions can make a difference in reducing new infections and improving life for people infected with and affected by HIV.

Now more than ever, it is critical that we monitor and evaluate programmatic efforts to justify the expenditure of resources and enable them to grow to the levels that will ultimately have a sustained impact. Toward this end, *Evaluating Programs for HIV/AIDS Prevention and Care in Developing Countries* provides a

comprehensive framework to help country program managers determine the effectiveness of their programs. This Handbook describes the methods needed to answer three simple but important questions: Are we doing the right things? Are we doing them right? Are we doing them on a large enough scale to make a difference? Answering these questions will allow program managers to decide how and when to modify existing programs or design new ones.

As we move into the third decade of this epidemic, it is incumbent upon us to demonstrate the results of our efforts. The third decade of the HIV/AIDS pandemic can be the decade of evidence of program effectiveness. These are times of great challenge and great opportunity. The time is now to show that we can make—and already have made—a difference.

Helene Gayle, MD, MPH

Director (1995-2001)

National Center for HIV, STD, and TB Prevention
U.S. Centers for Disease Control and Prevention

ACKNOWLEDGMENTS

The authors listed in this Handbook are only a handful of the individuals who have directly contributed to the conceptual and methodological advancements presented here from the field of evaluating and monitoring HIV prevention and care programs.

Since HIV was first discovered and interventions established to contain its spread 20 years ago, program managers and researchers alike have grappled with how to best evaluate and monitor interventions to contain the epidemic.

Our collaborators from numerous countries in the IMPACT Project, a U.S. Agency for International Development (USAID)-funded global prevention and care project managed by Family Health International, as well as the United Kingdom’s Department for International Development (DFID)-supported HIV prevention projects in several countries, were generous in sharing examples of their monitoring and evaluation plans. We are also indebted to UNAIDS and the World Health Organization (WHO), which have consolidated the experience of partner agencies worldwide and promoted international best practice standards in monitoring and evaluation. Our evaluation handbook will complement their guidelines on indicators for national HIV/AIDS programs.

Additionally, we are grateful to our colleagues in other USAID-funded projects working in numerous countries—MEASURE, HORIZONS,

AIDSMARK, and SYNERGY—who have tackled evaluation and monitoring issues from various angles and have contributed to many of the ideas expressed here.

Finally, all HIV program monitoring and evaluation efforts depend on the participation of people living with HIV and at risk of HIV infection. Whether through quantitative surveys, in-depth interviews, or focus groups, we ask our evaluation participants to share with us their personal behaviors regarding sex, injecting drug use, and related vulnerability—sometimes not just once, but several times. And they do. If not for the openness and candidness by the participants in these studies, our programs would not improve. We gratefully thank these participants who have represented the universe of individuals at risk of infection and hope that the operational approaches and data collection methods presented here lead to better services for the ultimate beneficiaries of HIV prevention and care programs.

The Editors

CONTRIBUTORS

Joseph Amon, MPH

Washington, DC

Tim Brown, PhD

Population and Health Studies
East-West Center
Bangkok, Thailand

Paul Bouey, PhD

Research and Evaluation
National Native American AIDS
Prevention Center
Oakland, CA

Michel Caraël, PhD

Joint United Nations Programme on
HIV/AIDS (UNAIDS)
Geneva, Switzerland

Thomas Coates, PhD

Center for AIDS Prevention Studies
AIDS Research Institute
University of California at
San Francisco
San Francisco, CA

Gina Dallabetta, MD

HIV/AIDS Prevention and
Care Department
Family Health International
Arlington, VA

Donna Flanagan, MA

Boulder, CO

Barbara Franklin, PhD

Asia Regional Office
Family Health International
Bangkok, Thailand

Steven Forsythe, PhD

The Futures Group International
Glastonbury, CT

Susan Hassig, PhD

Department of Biostatistics
and Epidemiology
Tulane University School of Public
Health and Tropical Medicine
New Orleans, LA

Jan Hogle, PhD

Social Impact
Reston, VA

Claudes Kamenga, MD, MPH

HIV/AIDS Prevention and
Care Department
Family Health International
Arlington, VA

Christine Kolars-Sow, PhD, MPH

Implementing AIDS Prevention
and Care Project (IMPACT)
Cote d'Ivoire Country Office
Family Health International
Arlington, VA

Robert Magnani, PhD

Department of International
Health and Development
Tulane University School of Public
Health and Tropical Medicine
New Orleans, LA

Hally Mahler, MHS

HIV/AIDS Prevention and
Care Department
Family Health International
Arlington, VA

André Meheus, MD, PhD

Department of Epidemiology and
Community Medicine
University of Antwerp
Antwerp, Belgium

Stephen Mills, MPH

HIV/AIDS Prevention and
Care Department
Asia Regional Office
Family Health International
Bangkok, Thailand

Elizabeth Pisani, MSc

HIV/AIDS Prevention and
Care Department
Family Health International
Jakarta, Indonesia

Eric van Praag, MD, MPH

HIV/AIDS Prevention and
Care Department
Family Health International
Arlington, VA

Thomas Rehle, MD, PhD

HIV/AIDS Prevention and
Care Department
Family Health International
Arlington, VA

Deborah Rugg, PhD

U.S. Centers for Disease
Control and Prevention
Atlanta, GA

Tobi Saidel, PhD

HIV/AIDS Prevention and
Care Department
Asia Regional Office
Family Health International
Bangkok, Thailand

Bernhard Schwartländer, MD

Joint United Nations Programme
on HIV/AIDS (UNAIDS)
Geneva, Switzerland

Susan Smith, MBA

The Futures Group International
Glastonbury, CT

David Sokal, MD

Clinical Research
Family Health International
Research Triangle Park, NC

Richard Steen, PA, MPH

Family Health International
Arlington, VA

John Stover, MA

The Futures Group International
Glastonbury, CT

Michael Sweat, PhD

Department of International Health
Johns Hopkins University School of
Hygiene and Public Health
Baltimore, MD

Daniel Tarantola, MD

Office of the Director General
World Health Organization
Geneva, Switzerland

TABLE OF CONTENTS

		List of Tables	
		Comprehensive evaluation framework	1-1 9
		Potential program outcome/impact measures	1-2 11
		Study design: hierarchy of evidence (decreasing strength of evidence)	1-3 13
		Examples of indicators by level of program evaluation	1-4 16
		Program logic model of Project A in Country X	2-1 26
		Illustrative STI control program monitoring and evaluation indicators and sources of data	4-1 52
		Commonly used methods for evaluating quality of STI case management	4-2 54
		The different types of STI surveillance	4-3 58
		STI syndromes used as passive surveillance indicators	4-4 60
		UNAIDS condom programming indicators	5-1 69
		Relative cost-effectiveness of intervention strategies in a typical East African city	5-2 76
		Illustrative voluntary counseling and testing (VCT) program indicators	6-1 83
		Possible sources of data on VCT service aspects	6-2 85
		Setting standards for monitoring the continuum of care	7-1 99
		Recommendations on choosing target groups for repeated cross-sectional behavioral surveys, according to stages of the epidemic	8-1 116
		Examples of respondent groups selected for BSS in various countries	8-2 117
		Key behavioral indicators in HIV/AIDS programs	10-1 153
		Probabilities of HIV-1 transmission per sexual exposure as used in the AVERT model	15-1 225
		Estimated and observed HIV seroconversions in a female sex worker (FSW) cohort, Cameroon	15-2 227
		Modeling the impact of an intervention in South Africa	15-3 228
		Estimated impact of different intervention components	15-4 229
		Guide to estimated life expectancy of assets	16-1 245
		Assessment tool for effectiveness analyses	17-1 258
		List of Figures	
		Monitoring and evaluation pipeline	1-1 18
		Health systems models identifying barriers to effective management in persons with STIs	4-1 50
		Models of STI transmission dynamics	4-2 51
		Interactions between STIs and HIV	4-3 62
		Awareness of condoms as a means of AIDS prevention in Mexico	5-1 72
		Reported condom use and abstinence following voluntary HIV counseling and testing	6-1 89
		The HIV/AIDS continuum of care	7-1 96
		Percent of males reporting sex with a non-regular partner in the past year: APAC Tamil Nadu 1996-1998	8-1 120
1	Introduction		
3	Section I Role of Evaluation in HIV/AIDS Programs		
5	Chapter 1: Conceptual Approach and Framework for Monitoring and Evaluation		
21	Chapter 2: Developing an Integrated and Comprehensive Monitoring and Evaluation Plan		
33	Section II Operational Approaches for Evaluating Intervention Strategies		
35	Chapter 3: Evaluating Behavior Change Communication Interventions		
47	Chapter 4: Evaluating Sexually Transmitted Infection Control Programs		
65	Chapter 5: Evaluating Condom Programming		
79	Chapter 6: Evaluating Voluntary HIV Counseling and Testing Programs		
91	Chapter 7: Evaluating Care Programs for People Living with HIV/AIDS		
105	Section III Methodologies for Measuring Behavioral Trends		
107	Chapter 8: Uses of Behavioral Data for Program Evaluation		
125	Chapter 9: Sampling Strategies for Monitoring HIV Risk Behaviors		
149	Chapter 10: Indicators and Questionnaires for Behavioral Surveys		
165	Chapter 11: Assessing the Validity and Reliability of Self-Reported Behavioral Data		
173	Chapter 12: The Role of Qualitative Data in Evaluating HIV Programs		
195	Chapter 13: Effective Dissemination of Data Collection Results		
203	Section IV Assessing Program Impact		
205	Chapter 14: Understanding HIV Epidemics and the Response to Them: Linking Data on Behavior Change and HIV Decline		
219	Chapter 15: Translating Survey Data into Program Impact: The AVERT Model		
235	Chapter 16: Guidelines for Assessing the Economic and Financial Costs of HIV/AIDS Prevention and Care Programs		
251	Chapter 17: Guidelines for Performing Cost-Effectiveness Analysis of HIV/AIDS Prevention and Care Programs		

120	8-2	Percent of female sex workers reporting always condom use with clients: Cambodia 1996-1999
208	14-1	Relationship between incidence, prevalence, and mortality
212	14-2	HIV prevalence by age group, Nsambya, Uganda
213	14-3	Percentage of population with sexual experience, by age (15-24 years), in 1989 and 1995, Uganda
214	14-4	Trend in HIV prevalence in 21-year-old military conscripts in Thailand
215	14-5	Number of STIs reported from government clinics, by gender, Thailand
216	14-6	Behavioral change and HIV/STI decline among 21-year-old men in northern Thailand
262	17-1	Simple decision tree for two interventions

List of Boxes

25	2-1	Evaluation teams: the value of internal and external perspectives
31	2-2	KABP surveys in every project: useful or not?
42	3-1	Sample question guide
84	6-1	Components and characteristics of voluntary HIV counseling
88	6-2	Illustrative example: The AIDSCAP/UNAIDS/WHO Multicenter Randomized Trial
100	7-1	Suggested standards for essential HIV care in Uganda
101	7-2	Examples of comprehensive care indicators
114	8-1	Identifying and mapping sub-populations
129	9-1	Illustrative primary sampling unit (PSU) or cluster definitions for selected sub-populations
134	9-2	Illustrative example of the sampling frame development process
135	9-3	Illustrative example of a two-stage cluster sampling scheme for establishment-based sex workers
136	9-4	Illustrative application of cluster sampling for men who have sex with men
137	9-5	Illustrative application of targeted snowball sampling for IDUs
138	9-6	Illustrative example of an in-class school survey of youth
139	9-7	Illustrative example of an out-of-class school survey of youth
140	9-8	Illustrative cluster definitions for subgroups of the mobile population target group
142	9-9	Illustrative sample size calculations
143	9-10	Illustrative calculations of sample sizes needed to locate a predetermined number of target group members with specified characteristics
189	12-1	Choosing a software package
192	12-2	Tips on writing qualitative research results
193	12-3	Suggested outline for a qualitative research report (focus group discussions, key informant interviews, rapid ethnographic study)

As the HIV epidemic continues its primary spread throughout developing countries of the world, the quality and effectiveness of interventions designed to reduce transmission are as critical as ever. After more than 15 years of fighting the epidemic, the worldwide track

record is mixed: Some countries have documented success in curtailing new infections, while others continue to witness high incidence rates.

This Handbook is intended to fill an important unmet need in the field. Evaluating HIV prevention and care programs so that they can promote a high level of quality and effectiveness—even in resource-limited settings—is its subject. Just as substantial progress has been made in developing effective interventions in recent years, so have appropriate methodologies, strategies, and indicators for the evaluation of these programs been honed and sharpened. *Evaluating Programs for HIV/AIDS Prevention and Care in Developing Countries* brings together the combined experience of collaborations between scientists and program designers and implementers, universities and local community-based organizations, governments and private agencies—partnerships often hard won, but ones that have produced results.

Nevertheless, compared to other public health areas, the evaluation of HIV/AIDS prevention and care programs is relatively young, and critical questions remain on how best to measure their effectiveness. Vigorous, but healthy and necessary, debates will likely continue over the next years on evaluation and measurement issues. In describing various facets of evaluation issues, the chapters that follow here point out some of these continuing challenges.

The focus of this Handbook is on evaluating programs related to the sexual transmission of HIV. It was developed for a target audience of program managers and decision makers of service delivery programs as opposed to researchers who are evaluating the efficacy of interventions through experimental or quasi-experimental research methods.

Section I lays the foundation for HIV/AIDS program evaluation by describing the current consensus on generic concepts, approaches, and

frameworks (Chapter 1) and by outlining the practical development of an integrated evaluation and monitoring plan for projects (Chapter 2).

Section II describes the operational approaches for evaluating the core program strategies necessary to effectively reduce the sexual transmission of HIV (Chapters 3, 4, 5, 6), illustrating the unique and often very disparate methodologies needed to evaluate different program components. The section concludes with long overdue guiding principles on how to evaluate HIV/AIDS care programs (Chapter 7). This may help to increase the commitment to funding care-related activities by providing donors and decision makers with the necessary feedback to determine whether the invested resources have yielded the expected results.

Section III focuses on measuring behavior change as the key outcome of standard prevention efforts. State-of-the-art methodologies and survey instruments for collecting behavioral data (Chapters 8, 9, 10) and assessing their validity and reliability (Chapter 11) are covered here, complemented by a chapter on tools for collecting qualitative information (Chapter 12). The section concludes with effective strategies for disseminating survey data to key audiences (Chapter 13).

Finally, Section IV tackles evaluation issues related to assessing program impact, emphasizing the need to analyze behavioral and sero-epidemiological data in tandem (Chapter 14). It also describes a newly developed tool for estimating the impact of different prevention strategies on HIV transmission (Chapter 15), and provides practical guidelines for cost and cost-effectiveness analysis (Chapter 16, 17).

The individual chapters in this Handbook attempt to follow the logic of a program evaluation cycle, and though they are written as “stand alone pieces,” they are clearly interconnected. Cross-references to other chapters are found throughout the book to assist readers in understanding how the individual components of evaluation fit into the larger whole.

The authors and editors of *Evaluating Programs for HIV/AIDS Prevention and Care in Developing Countries* hope this book provides program managers and decision makers with clear and practical guidance on a subject that is critical to the continued success of efforts to reduce HIV transmission and contain the epidemic in the developing world.



ROLE OF EVALUATION IN HIV/AIDS PROGRAMS

Conceptual Approach and Framework for Monitoring and Evaluation

5

Developing an Integrated and Comprehensive Monitoring and Evaluation Plan

21

1

CONCEPTUAL APPROACH AND FRAMEWORK FOR MONITORING AND EVALUATION

THOMAS REHLE AND SUSAN HASSIG

Basic Considerations	8
A Comprehensive Evaluation Framework for Country Programs	8
Formative Evaluation	9
Process Evaluation	10
Effectiveness Evaluation: Assessing Outcome and Impact	11
The Attribution Dilemma: Are Observed Changes a Result of Prevention Interventions?	12
The Role of Triangulation	14
Choice of Indicators	15
Differentiating Evaluation Efforts	17
Conclusion	19
References	20

1

CONCEPTUAL APPROACH AND FRAMEWORK FOR MONITORING AND EVALUATION

Evaluating HIV/AIDS prevention and care programs is a never-ending challenge, but recognizing its importance in improving current interventions may help to enhance the success of future initiatives.

There are probably as many definitions of “program evaluation” or “evaluation research” as there are program evaluators. Our approach to evaluating HIV/AIDS prevention and care programs is best captured by a description of evaluation provided by Michael Quinn Patton¹:

“I use the term evaluation quite broadly to include any effort to increase human effectiveness through systematic data-based inquiry. When one examines and judges accomplishments and effectiveness, one is engaged in evaluation. When this examination of effectiveness is conducted systematically and empirically through careful data collection and thoughtful analysis, one is engaged in evaluation research.... Evaluation is applied research, or a type of “action science.” This distinguishes evaluation research from basic academic research....The purpose of applied research and evaluation is to inform action, enhance decision-making, and apply knowledge to solve human and societal problems....Applied evaluative research

is judged by its usefulness in making human actions and interventions more effective and by its practical utility to decision makers, policymakers and others who have a stake in efforts to improve the world.”

Such an evaluation approach is utilization-focused. This approach emphasizes the interests of key stakeholders and primary users of the information at all levels, for example the donor, the host country, and the implementing agencies. It applies socio-epidemiological research to identify ways to improve the design and implementation of HIV/AIDS prevention and care programs.

This chapter first describes several considerations that are fundamental to planning an evaluation effort. It then presents a comprehensive framework for country programs by explaining the major types of evaluation and discussing several important issues related to planning evaluation programs and improving their ability to measure program effects.

BASIC CONSIDERATIONS

Several considerations underlie the decisionmaking process about HIV/AIDS program evaluation. The selection of an appropriate evaluation concept for an AIDS prevention program is crucial because it determines the guiding philosophy behind the actual evaluation process. A number of theorists and evaluation practitioners have proposed various conceptual approaches to evaluation. These approaches differ in their conception as to what evaluation is, what the relationship with the primary client and other stakeholders should be, who should be making the relevant value judgments regarding the program, and the criteria for judging the evaluation process itself.

The conceptual approach debate was, and is for the most part, a debate about the best ways to measure and interpret change. It has highlighted a series of methodological dimensions among which there are variations in emphasis. These dimensions focus attention on some of the options available for making decisions about methods. Today, for example, there is consensus that both quantitative and qualitative data are valued and recognized as legitimate for program evaluation. In fact, these methods are by no means incompatible and should be used in combination².

Deciding what and how much data to gather in an evaluation involves difficult methodological decisions and trade-offs between the quality and utility of information. An evaluation approach that uses multiple data collection methods, both quantitative and qualitative, is more likely to address diverse evaluation needs than is a more limited approach. At the same time, research priorities must be sensitive to competing needs for resources in an environment in which the HIV/AIDS epidemic is growing rapidly and evaluation is sometimes considered a luxury. It is a major task of the evaluator to match

research methods to the reality of particular evaluation questions and to the available resources.

There is also a need for evaluation researchers to play an active role, not merely a consultative one, in making design decisions for program assessments. Although a program evaluator should be a neutral scientific observer, he or she can also mediate between different stakeholder groups, can enable others through a participatory evaluation approach, and can advocate for the dissemination of evaluation results within the larger arena of decisionmaking³. Planning evaluation and data collection activities in a participatory fashion is essential for achieving the delicate balance between practical needs and methodological desirability. Key stakeholders should be included in the planning process and every effort should be made for effective use of limited resources. Ensuring active support and participation of key stakeholders who have an interest in the results obtained by various data collection systems is particularly important for programs funded by external donors that use host country institutions for data collection activities. Data produced by these efforts will have a better chance to be timely and of acceptable quality. Whenever possible, participants, including implementing institutions, host-country collaborators and local representatives of donor agencies, should attempt to reach consensus regarding data needs.

A COMPREHENSIVE EVALUATION FRAMEWORK FOR COUNTRY PROGRAMS

HIV/AIDS prevention and care programs need to be evaluated at different phases of the program cycle. Table 1-1 outlines a framework for comprehensive program evaluation. All stages of evaluation have to be considered together to provide an overall picture of the program because no single data collection approach can supply all the information necessary to improve

program performance or affect policy change. Multiple complementary evaluation approaches and multiple methodologies (qualitative and quantitative) have to be applied to address different evaluation needs.

Formative Evaluation

Formative evaluation should be conducted during the planning (or replanning) stage of a prevention and care program to identify and resolve intervention and evaluation issues before the program is widely implemented. This is the time when flexibility is greatest and program sponsors are freer to make decisions about how to proceed.

Formative evaluation explores the need for interventions, provides the information necessary to define realistic goals and objectives for the program interventions, and helps program planners make tentative decisions about effective, feasible intervention strategies and how to

carry them out. Formative evaluation can also be used as an exploratory tool as the project is being carried out to provide feedback to project managers to help them adjust program objectives to changing situations. Formative evaluation research can identify unacceptable or ineffective intervention approaches, designs, and concepts.

Because of the urgency of the HIV/AIDS problem, many prevention programs have rushed to carry out interventions without preparing first by conducting thoughtful formative evaluation. The lack of this type of evaluation is particularly felt in community-based interventions designed to reduce sexual transmission of HIV. In many cases, interventions have been based on ideas developed outside of the context of the lives of the people to whom the interventions have been delivered. The literature on behavioral change interventions is full of examples of ideas that made perfect sense in the abstract but failed

Comprehensive evaluation framework

TABLE 1-1

Types of Evaluation	Questions Answered by the Different Types of Evaluation
Formative Evaluation Research (Determines Concept and Design)	Is an intervention needed? Who needs the intervention? How should the intervention be carried out?
Process Evaluation (Monitors Inputs and Outputs; Assesses Service Quality)	To what extent are planned activities actually realized? How well are the services provided?
Effectiveness Evaluation (Assesses Outcome and Impact)	What outcomes are observed? What do the outcomes mean? Does the program make a difference?
Cost-Effectiveness Analysis (Including Sustainability Issues)	Should program priorities be changed or expanded? To what extent should resources be reallocated?

completely in the “real world,” mainly because the ideas were unacceptable to the target audience or were not stated in ways that were relevant to the lives of those people. A fuller understanding of the issues might well have led planners to redesign the intervention to make it more appealing to the selected audience. Fortunately, this situation is changing because formative evaluation is now being applied more frequently in designing prevention programs.

Formative evaluations use a mix of research methods that can rapidly provide relevant information to program designers. These methods include:

- reviews of existing information;
- focus group discussions;
- individual in-depth interviews;
- participant observations; and
- short quantitative surveys with structured questionnaires.

The most frequently cited methodological criticism of formative evaluation is its lack of external validity or generalizability. Because the results of the evaluation derive from small-scale rapid assessment procedures and/or pilot studies, one cannot generalize from them to a larger population. Despite this limitation, formative evaluation research can usually identify unacceptable or ineffective intervention approaches, designs, and concepts. However, even with adequate formative evaluation at the program planning stage, there is no guarantee that a prevention program will be effective when finally implemented; it may not be implemented adequately enough to be effective.

Process Evaluation

Once activities are underway, there is a need to examine whether they are being carried out correctly, on time, and within budget. Process evaluation addresses such basic questions as,

“To what extent are planned intervention activities actually realized?” and “What services are provided, to whom, when, how often, for how long, and in what context?” Both input (the basic resources required in terms of manpower, money, material, and time) and output (the immediate service improvement expressed as distributed commodities, trained staff, and service units delivered) are key elements of process evaluation. These questions are often answered in quantitative terms. Qualitative evidence of how and why a prevention program works or fails to work is equally important in answering process evaluation questions. Process evaluation requires getting close to data, becoming intimately acquainted with the details of the program, and observing not only anticipated effects but also unanticipated consequences. An understanding of the processes through which intervention activities achieve effects can help to explain the outcome of the intervention. Process evaluation, however, does not demonstrate whether interventions are effective.

Process evaluation can also play an important role in improving or modifying interventions by providing the information necessary to adjust delivery strategies or program objectives in a changing epidemic. Process-oriented evaluation is carried out throughout the course of the program implementation and should use different methodological approaches to assess service delivery, ranging from reviews of service records and regular reporting systems, key informant interviews, exit interviews of service users, direct observations by ‘mystery clients’ (for example, in sexually transmitted infection [STI] and voluntary counseling and testing [VCT] services) to quantitative population-based surveys to assess program coverage and barriers to service use. Different qualitative and quantitative study designs that are complementary to one another provide together the most comprehensive information. (See Section II—Chapters 3 through 7—for details on these approaches.)

Effectiveness Evaluation: Assessing Outcome and Impact

Evaluating the effectiveness of AIDS prevention programs will almost always require quantitative measurements. These measurements will assess the extent to which the objectives of the program were achieved. Effectiveness evaluation is used to answer the questions, “What outcomes were observed?,” “What do the outcomes mean?,” and “Does the program make a difference?”

Taking into account the various implementation stages of HIV/AIDS prevention programs and the fact that, over time, new age cohorts become sexually active, it is advisable to stratify effectiveness evaluation by **short-term and intermediate program effects (program outcome)** and **long-term program effects (program impact)**. Table 1-2 provides examples of program outcome and impact measures for these different stages. Changes in HIV/AIDS-related attitudes, the reduction of risk behaviors and adoption of protective behaviors, and changes in STI rates are considered to be the most appropriate short-term or intermediate (also called proximate) outcome measures for interventions designed to reduce

sexual transmission of HIV. Long-term effects include impact on HIV/AIDS trends, sustainability issues, and improved societal response.

Outcome and impact evaluation is intimately connected with process evaluation. Process information can help the evaluator to understand how and why interventions have achieved their effects and, perhaps, what is actually making the difference. Examining outcome/impact indicators without assessing the process of program implementation could lead to erroneous conclusions regarding the effectiveness of the intervention.

Program goals and objectives have to be carefully defined to allow the selection of appropriate outcome and impact measures to assess the effectiveness of an AIDS prevention program. Effectiveness evaluation is generally based on indicators that provide quantitative value from which the outcome and impact of interventions can be measured. Because multiple interventions working synergistically together are most effective in producing behavior change, surveys should not be typically designed to capture the effects of one single intervention (see Figure 1-1 later in this chapter and Box 2-2 in Chapter 2).

Potential program outcome/impact measures

TABLE 1-2

Program Outcome (short-term and intermediate effects)	Program Impact (long-term effects)
<ul style="list-style-type: none"> • changes in HIV/AIDS-related attitudes • HIV/STI-related risk behaviors • trends in STI rates (e.g., gonorrhea) • increase in social support/community response 	<ul style="list-style-type: none"> • sustained changes in HIV/STI-related risk behaviors • trends in HIV/AIDS rates • AIDS-related mortality rates • reduced individual and societal vulnerability to HIV/AIDS • sustained changes in societal norms

Rather, they should be designed to measure behavioral trends in population groups who are exposed to combined interventions. The evaluation of one intervention is usually conducted through rigorous and expensive controlled trials.

Cost-effectiveness Analysis

Cost-effectiveness analysis also measures program effectiveness, but expands the analysis by adding a measure of program cost per unit of effect (for example, per number of HIV infections averted). By comparing the costs and consequences of various interventions, cost analyses and cost effectiveness estimates can assist in priority setting, resource allocation decisions, and program design. (Chapter 17, “Guidelines for Performing Cost-effectiveness Analysis of HIV/AIDS Prevention and Care Programs” provides more detail on this type of evaluation.)

THE ATTRIBUTION DILEMMA: ARE OBSERVED CHANGES A RESULT OF PREVENTION INTERVENTIONS?

The ultimate goal of any HIV prevention program is to reduce the number of new infections. Program evaluation is intrinsically complex, however, due to the temporal evolution of epidemics and our poor understanding of how different behaviors and epidemiologic factors influence epidemic patterns as they move from an epidemic phase to an endemic state. Several factors unrelated to intervention effects can contribute to the observed stabilization or decreases in the prevalence or incidence of HIV in a given setting. They include:

- mortality, especially in mature epidemics;
- saturation effects in populations at high risk;
- behavioral change in response to the experience of HIV/AIDS among friends and relatives;

- differential migration patterns related to the epidemic; and
- sampling bias and/or errors in data collection and analysis.

Determining whether observed changes in HIV incidence and prevalence are a reflection of the natural history of the epidemic or due to intervention effects is a critical evaluation issue. This is particularly true when evaluating behavior changes in the face of growing numbers of people with AIDS-related illnesses because there is evidence that secular trends toward risk reduction will occur. For example, having a friend or relative with HIV/AIDS may influence adolescents to delay the onset of sexual relations or motivate those with non-regular sex partners to use condoms. Human sexual behavior is influenced and shaped by many factors and exposure to an HIV prevention program is only one of them.

The question of whether behavior changes can be attributed to prevention programs, especially in countries with advanced HIV epidemics, has created some friction between stakeholders and program implementers at the field level. Their different perspectives on this issue also reflect fundamental differences regarding the criteria for judging the process of program evaluation itself. From a public health perspective, it may not matter whether the observed changes are due to a particular intervention. What is most important is that sexual practices have become safer and HIV infection should subsequently decrease.

From the cost-effectiveness or policy perspective, however, it is important to determine what caused the observed changes in sexual behavior. If the changes would have occurred without a particular intervention that was designed to contribute to the observed changes, the costs of the intervention could be considered as resources better spent on something more useful⁴. Prevention programs are under growing pressure

to estimate which approaches work best for specific target populations in different epidemiologic settings with a given level of inputs in order to allocate resources in a cost-effective manner⁵. Effectiveness evaluation, therefore, is critical because it can answer a basic question, “Does the program make a difference?”

A vexing task of assessing program effectiveness is to disentangle the attributable affects of a prevention program from the gross outcome and impact observed. Such estimates can be made with varying degrees of plausibility, but not with certainty. A general principle applies here: The more rigorous the research design, the more convincing the resulting estimate⁶. A hierarchy of evidence based on the study design can be established that reflects the degree of certainty in concluding that a given proportion of the observed changes in behavior is attributable to the intervention program and is not the result of other factors⁷.

Table 1-3 presents a ranking of different study designs according to their decreasing strength of evidence. Non-experimental observational methods with no control groups have been routinely used in behavioral outcome evaluations. It is important to recognize, however, that a before-and-after evaluation design with no comparison groups may be useful for assessing a prevention program’s proficiency in delivering services, but it is not a very convincing design to measure program effectiveness. The inference of cause and effect from such a design is problematic because competing explanations for across-time behavioral changes cannot be ruled out⁸.

In some situations, the evaluation could assess “exposure” to an intervention program or a specific element, and determine the extent of the association of that exposure with the desired outcome. This method, however, can be limited in its utility by factors such as lack of association of services or products with the intervention and inaccurate reporting by respondents of their participation in the intervention.

The interpretation of program evaluation data should always be approached with caution. In most situations, the program and evaluation process as a whole is not a rigorously controlled experimental trial. The ability of an evaluation to precisely determine the true extent of a program’s effectiveness is often limited by time, resources, and the lack of a rigorous design. Many factors can confuse or confound the results measured, and biases can be introduced by a range of factors inherent to the problem of HIV/AIDS, the available measurement options, and those conducting the evaluation. One of the most difficult questions to answer in any evaluation is that of attributing any measured effect to the program being evaluated. Defining the web of interacting and overlapping influences is extremely difficult, and is one of the reasons why so many programs have difficulty attributing results to their actions. At some point, we need to stop worrying about attribution in such settings and focus on monitoring the changes as they occur.

Study design: hierarchy of evidence (decreasing strength of evidence)

TABLE 1-3

I Experimental studies	a) randomized controlled trials
II Quasi-experimental studies	a) controlled trials with no randomization
III Observational studies	a) cohort studies with concurrent controls b) cohort studies with historical controls c) case-control studies
IV Cross-sectional surveys with no control groups	a) repeated surveys in target populations b) pre-post intervention surveys

THE ROLE OF TRIANGULATION

Triangulation can be achieved through using multiple data sources, different researchers, multiple perspectives to interpret a single set of data, or multiple methods applied to a single program, problem, or issue.

In the absence of rigorous controlled trials, data triangulation procedures must be applied to substantiate a link between interventions and observed behavior changes. For example, process evaluation data on condom sales, the intensity of peer education, or the quality and coverage of media campaigns can be combined with an analysis of behavioral outcome data to provide an understanding of the process through which an intervention has achieved its effects. Results from behavioral surveys should be analyzed together with findings from qualitative evaluation research that is carried out in sub-samples of surveyed target populations. Such research can include focus group discussions, key informant interviews, and

In the absence of rigorous controlled trials, data triangulation procedures must be applied to substantiate a link between interventions and observed behavior changes.

rapid ethnographic studies. This type of analysis will allow a more appropriate interpretation of observed outcome data because they are the likely results of the aggregate effects of multiple interventions as well as environmental and personal factors.

Many of the areas that need to be measured to evaluate HIV/AIDS programs are sensitive and very personal in nature, such as sexual behaviors or personal attitudes toward persons with HIV/AIDS. Validity and reliability are critical issues for sexual behavior research because the behaviors cannot be directly observed. Self-reports of sexual behaviors in the absence of additional evidence are often considered invalid and unreliable by stakeholders for whom such data are sensitive and run against firmly held cultural norms.

One of the best methods for promoting reliability and validity, therefore, is to triangulate behavioral data with all other available and relevant biological, behavioral, and process data to explain more comprehensively the context in which risk behaviors take place. (Chapter 11, “Assessing the Validity and Reliability of Self-reported Behavioral Data,” provides more detail on this issue.)

Given the abundance of AIDS-related research conducted in many countries, secondary data are a source of material for triangulation. Multiple-method triangulation is probably the most common triangulation technique. Rapid ethnographic research, combining semistructured information gathering with mapping, participant observation, and in-depth interviews, is another possibility. Focus group discussions have been widely used, as have individual in-depth interviews or key informant interviews, to obtain stakeholders’ (or other key individuals’) opinions about target groups’ behavior. (Chapter 12, “The Role of Qualitative Data in Evaluating HIV Programs,” provides more information on qualitative evaluation tools.)

Finally, it is important to realize that behavior change interventions have to be in place for sufficient amounts of time and on a large enough scale to have an impact on personal behavior, social norms in communities, and ultimately on the epidemic. The example of Thailand shows that a focused intervention strategy implemented

on a national scale can result in substantial declines in HIV incidence and prevalence in targeted populations. It is also an example of applied triangulation of data: STI/HIV trends were systematically collected by sentinel surveillance systems, and behavioral surveillance data provided the necessary supplementary information to interpret the observed epidemiological trends. There is now growing consensus that country programs need to monitor risk behavior trends together with trends in HIV infection. (See Chapter 8, “Uses of Behavioral Data for Program Evaluation” for more on this issue.)

CHOICE OF INDICATORS

One of the critical steps in designing and carrying out an evaluation of an HIV/AIDS program, or any program for that matter, is selecting appropriate indicators. This can be a fairly straightforward process if the objectives of the program have been clearly stated and presented in terms that define quantity, quality, and time frame of a particular aspect of the program. Even with well-defined objectives, however, the choice of indicators for the evaluation of many programs requires careful thought and consideration of both theoretical and practical elements.

The following questions can be helpful in selecting indicators:

- Is the focus of the objective a parameter that can be measured accurately and reliably?
- Are there alternative measures that need to be considered?
- What resources (human and financial) does the indicator require?
- Are there areas for congruency, either in the content of the indicator or the means of gathering the data?
- Are there any additional measures that

would help in interpreting the results of the primary objective?

Selecting indicators and setting targets is usually done during the process of program planning and replanning, preferably in a participatory way with the implementing agency and key stakeholders. Setting targets and benchmarks should also include information from similar types of interventions, so that the targets set are realistic from the perspective of the target population, resource allocation, and intervention type.

While the level of attainment to be measured by the indicator is not actually part of the indicator itself, it is a critical factor. The magnitude of the level to be measured affects the size of the sample of the population needed to estimate that level accurately. It may also help evaluators select additional or supplemental indicators that might assist in later interpretations of the results. Ideally, indicators should be:

- **Valid**—They should measure the condition or event they are intended to measure.
- **Reliable**—They should produce the same results when used more than once to measure the same condition or event.
- **Specific**—They should measure only the condition or event they are intended to measure.
- **Sensitive**—They should reflect changes in the state of the condition or event under observation.
- **Operational**—It should be possible to measure or quantify them with developed and tested definitions and reference standards.
- **Affordable**—The costs of measuring the indicators also should be reasonable.
- **Feasible**—It should be possible to carry out the proposed data collection.

Validity is inherent in the actual content of the indicator and also depends on its potential for

being measured. Reliability is inherent in the methodology used to measure the indicator and in the person using the methodology. Many familiar

outcome indicators in HIV/AIDS prevention, such as measures of condom use, provide challenges to the evaluator with respect to validity and reliability.

TABLE 1-4 Examples of indicators by level of program evaluation

Levels of Evaluation		Indicator
Process Evaluation		
INPUTS		<ul style="list-style-type: none"> Resources allocated (e.g., percent of national budget) Condom availability at central level (Prevention Indicator¹⁰ [PI] 2)
OUTPUTS		<ul style="list-style-type: none"> Knowledge of HIV transmission (PI 1) Condom availability in periphery (PI 3) Proportion of those 12-17 years of age receiving AIDS/sexual health education Percentage of services with improved quality, e.g., STI case management (PI 6 and PI 7), voluntary counseling and testing (VCT), care for people living with HIV/AIDS Percentage of blood transfusion facilities with uninterrupted supply of appropriate HIV screening tests
Effectiveness Evaluation		
OUTCOMES	Proximate determinants addressing the risk per sexual contact	<ul style="list-style-type: none"> Condom use during last act with non-regular partner (PI 5) Prevalence of urethritis among men aged 15-49 in last year (PI 9) Prevalence of positive syphilis (RPR/VDRL) serology among antenatal women aged 15-24 (PI 8)
	Proximate determinants addressing the frequency of sexual risk contacts	<ul style="list-style-type: none"> Proportion of men with non-regular partner during last year (PI 4) Prevalence of concurrent partnerships during last month Median age at first sex of boys and girls Proportion of young women less than 18 years of age having had sex with men more than 30 years of age during last year
IMPACT	HIV incidence/prevalence	<ul style="list-style-type: none"> HIV prevalence among women less than 25 years of age in antenatal clinics (to approximate incidence) (PI 10); HIV prevalence among high risk groups, e.g., STI patients/sex workers/injecting drug users HIV prevalence among adult men aged 15-49
	AIDS mitigation	<ul style="list-style-type: none"> HIV-associated mortality rates among adults aged 15-59 Number of communities with increased coping capacity
OVERALL PROGRAM EFFORT		<ul style="list-style-type: none"> AIDS Program Effort Index (API)

Source: Modified from Reference 9
PI = WHO/GPA prevention indicator

Interpreting outcome indicators for behavioral interventions that promote safer sex is further complicated by the fact that risk behaviors are measured in relative terms. For example, percentage figures of condom use measure the proportion of sexual exposures that are considered to be safe. These may or may not reflect the absolute number of sex acts that place individuals at risk for exposure to sexual transmission. Ten percent condom use in 10 HIV-associated sexual episodes is still “safer” than 75 percent condom use in 100 HIV-associated sex episodes (9 versus 25 unprotected HIV-associated sex acts, respectively). Therefore, in this example it also would be important to determine the frequency of condom use in absolute terms in a given risk situation. Behavioral surveys have begun to address this dilemma by collecting additional data on “always or consistent” condom use in the context of sexual episodes with non-regular partners.

Table 1-4 lists possible indicators related to different levels of program evaluation. The advantage of relating indicators to specific evaluation levels is that it also helps to identify opportunities for triangulating data. For example, survey data on condom use can be compared with information on condom distribution and availability in a defined intervention area. Or, available data on incident STIs, such as gonorrhea, in the surveyed population could be correlated with the condom data.

In collaboration with national and international partners, the United Nations AIDS Programme and the World Health Organization (UNAIDS/WHO)¹¹ have developed a standard set of indicators for country programs that will refine and expand the prevention indicators (PI) developed by WHO’s former Global Program on AIDS (GPA). Moreover, because HIV/AIDS/STI prevention and care programs are affected by many factors, including political commitment, available resources, and the socio-cultural and economic context, a new approach

is currently being developed to capture the overall effort of national HIV/AIDS programs. *The AIDS Program Effort Index (API)* is a composite score comprised of the main components of an effective national response.

The potential advantage of this instrument is that it may yield useful information on the above issues even in the absence of more rigorous monitoring and evaluation systems. Using the key informant assessment approach, it also allows an assessment of areas that are difficult to capture with more objectively measurable indicators (e.g., political support and commitment). However, major concerns have been expressed with regard to the subjectivity and reliability of the API approach. The score depends entirely on the choice of informants, and the informants are likely to change from year to year. Questions have also been raised about the utility of a single composite score in which improvements in some areas may be masked by deterioration in other areas¹¹.

DIFFERENTIATING EVALUATION EFFORTS

Because of the various constraints on time, available funds, and trained staff, program managers and evaluation planners must balance what is ideal or preferred against what is feasible, useful, relevant, and essential when choosing how to evaluate a particular intervention or program. A useful approach for differentiating evaluation efforts is to define them in three different dimensions: the individual project dimension, the country program dimension, and the international dimension. Using this multi-dimensional approach (individual, country, and international project) to set priorities for the degree of rigor needed to evaluate programs and projects may alleviate some of the tension that arises when universal, standardized evaluation practices conflict with the objectives of individual projects.

One can think of the individual project dimension as an area of service delivery that, in most cases, does not require a rigorous research design to judge its proficiency, unless it is piloting a new intervention or responding to an unanswered research question, such as would occur with a demonstration project. Individual projects carrying out standard intervention strategies that have already been shown to be

effective in other similar settings should focus their evaluation activities on formative evaluation (when needed for project planning), process evaluation, and capacity building assessment. Figure 1-1 shows the number of projects in relation to the different levels of evaluation efforts and reflects the current situation in program evaluation. The monitoring and evaluation “pipeline” illustrates that there is usually a

reduced number of projects that actually warrant the evaluation of the effectiveness of their implemented prevention activities.

Within the dimension of a country program, several categories of evaluation should be emphasized—intervention outcomes, socioeconomic impact, and changes in societal norms. The guiding principle here is that in a situation in which multiple donors are conducting multiple interventions with overlapping target groups, certain types of evaluation are not appropriate for the scope of an individual project, but rather, should be coordinated and conducted by country or regional programs. Use of such an evaluation approach, especially in the area of behavioral surveys, not only saves money, but also makes sense in environments where the effects of individual projects from different donors cannot be sorted out anyway.

Country program evaluation includes (but is not limited to) the analysis of behavioral trends in different population groups in conjunction with an analysis of HIV/STI surveillance data; the evaluation of social marketing activities related to condoms, drugs, and services; STI case management; scoring the overall effort of the national program (for example, through the AIDS Program Effort Index); socioeconomic impact assessments; and epidemiological modeling of the country’s epidemic. Countries will have different programs of evaluation activities, reflecting their different information needs, which are determined by the stage of their epidemic, as well as their political and social environment, existing capacity for research, and available financial resources.

Evaluation efforts on the international dimension may address still existing uncertainties about which set of prevention interventions works best, in which setting, for whom, and under what circumstances (emphasis on cost-effectiveness analysis). This type of evaluation, however, requires large-scale community-based controlled

trials that are certainly beyond the scope of individual projects or even national programs.

Given the difficulties and high costs associated with directly measuring the impact of HIV prevention programs through large-scale incidence studies, more emphasis has now been placed on developing other methods of assessing impacts, for example through modeling. (Chapter 15, “Translating Survey Data into Program Impact: the AVERT Model,” provides more detail on one such effort.)

CONCLUSION

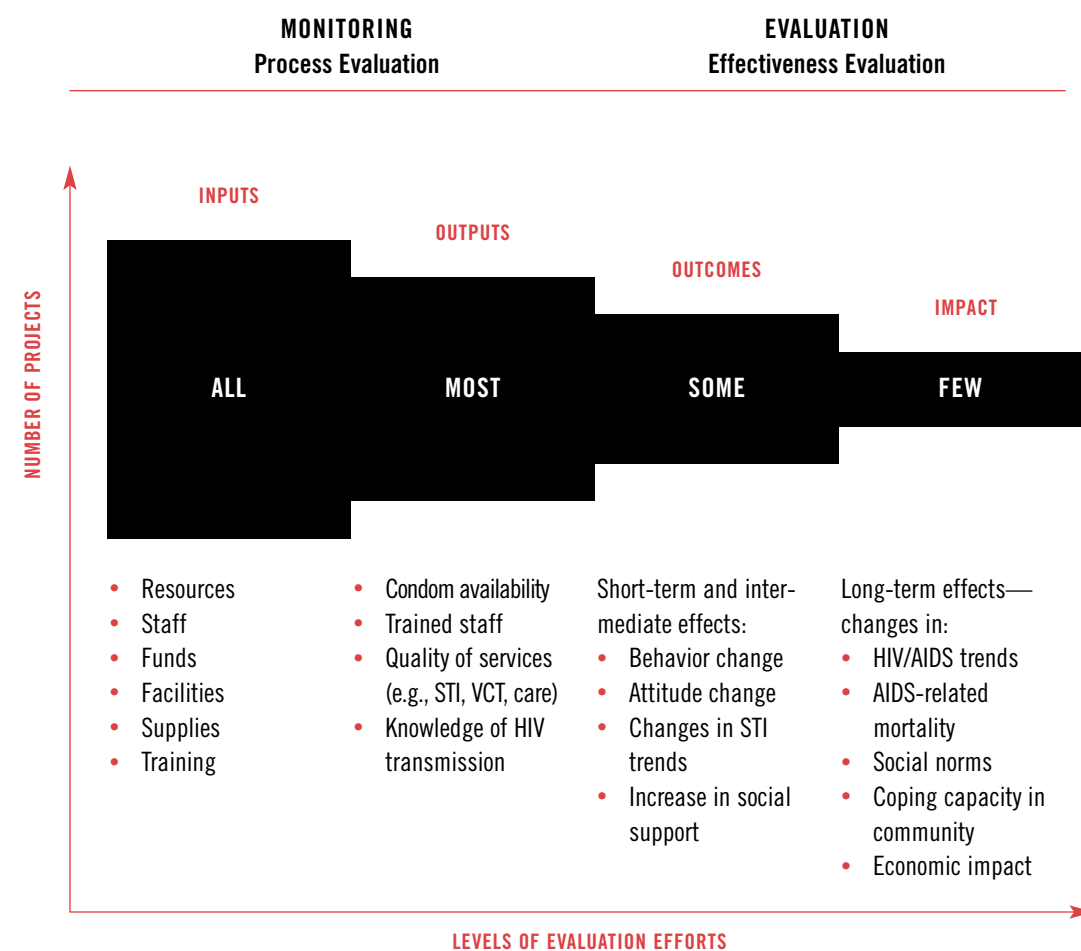
HIV/AIDS prevention and care program evaluation is applied socio-epidemiological research whose main purpose is to identify and solve practical problems and guide program managers and planners in improving the design and implementation of prevention and care activities. This perspective not only determines the role of program evaluation but also how an evaluation should be conducted, including the choice of indicators and levels of efforts in a given setting.

By applying different methods from several disciplines to many types of problems, program evaluation is a comprehensive research approach committed to meeting the needs of stakeholder groups as well as the requirements of the scientific community. This Handbook is designed to assist program evaluators in this critically important task.

Although program evaluation is context specific, a comprehensive framework as outlined in this chapter is helpful in defining the questions that are to be answered by the different types of evaluation during the program cycle. We advocate for a utilization-focused evaluation approach that emphasizes the interests of stakeholders as the primary intended information users. To achieve the delicate balance between practical needs and methodological desirability, it is therefore essential that program evaluations are planned in a participatory fashion with key stakeholders.

FIGURE 1-1

Monitoring and evaluation pipeline



Decisionmaking is a political process and program evaluators can play a major role in this process when evaluation efforts are expected to provide information of policymaking significance and relevance. Although evaluation researchers should be neutral scientific observers, there is also a need for them to assume a more active role and, if necessary, mediate between stakeholders with different and sometimes conflicting interests, perspectives, and information needs.

Given the limited resources in most developing countries, assessing the effectiveness of HIV prevention programs will often depart from

scientifically ideal designs. In the absence of more rigorous evaluation designs, we urge program managers and evaluators to apply triangulation procedures using multiple complementary methods as well as different data sources. Such a triangulated analysis will provide information comprehensive enough to allow a plausible and valid interpretation of observed outcome data, such as changes in risk behaviors, because they are the likely results of the aggregate effects of multiple interventions as well as environmental and personal factors.

REFERENCES

- 1 Patton MQ, editor. Qualitative evaluation and research methods. 2nd edition. Newbury Park (CA): Sage Publications; 1990. p. 11-12.
- 2 Patton MQ, editor. Utilization-focused evaluation. 2nd edition. Beverly Hills (CA): Sage Publications; 1986. p. 177-217.
- 3 Thompson JC. Program evaluation within a health promotion framework. *Can J Public Health* 1992;83(Suppl 1):S67-S71.
- 4 Coyle S, Boruch R, Turner C, editors. Evaluating AIDS prevention programs. Washington (DC): National Academy Press; 1991. p. 50-82.
- 5 Mertens TE, Caraël M. Evaluation of HIV/STD prevention, care and support: an update on WHO's approaches. *AIDS Educ Prev* 1997;9(2):133-145.
- 6 Rossi PH, Freeman HE, editors. Evaluation: a systematic approach. 5th edition. Newbury Park (CA): Sage Publications; 1993. p. 215-259.
- 7 Susser M. Some principles in study design for preventing HIV transmission: rigor or reality. *Am J Public Health* 1996;86(12):1713-1716.
- 8 Cook TD, Campbell DT, editors. Quasi-experimentation: design and analysis issues for field settings. Chicago: Rand McNally; 1979. p. 95-146.
- 9 Boerma T, Schwartländer B, Mertens T. A proximate determinants framework for monitoring and evaluation of HIV/AIDS/STD programs. From: Towards Improved Monitoring and Evaluation of HIV Prevention, AIDS Care and STD Control Programs. A joint UNAIDS, WHO, USAID/MEASURE evaluation workshop, Nairobi (Kenya); November 17-20, 1998.
- 10 WHO/GPA/TCO/SEF/94.1. Evaluation of a national AIDS programme: a methods package. Geneva: World Health Organization; 1994.
- 11 UNAIDS/00.17E. National AIDS programmes: a guide to monitoring and evaluation. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2000.

2

DEVELOPING AN INTEGRATED AND COMPREHENSIVE MONITORING AND EVALUATION PLAN

DEBORAH RUGG AND STEPHEN MILLS

Rationale for Developing an Evaluation Plan	24
Key Elements of an Evaluation Plan	24
Developing an Integrated and Comprehensive Evaluation Plan	24
Step 1: Identify Program Goals and Objectives	25
Step 2: Examine Existing Data and Past Evaluation Studies	28
Step 3: Identify Internal and External Evaluation Resources and Capacity	28
Step 4: Determine Evaluation Questions, Their Feasibility, and Appropriate Designs and Indicators	28
Step 5: Plan for Disseminating and Using Evaluation Findings	29
Using Technical Support	29
Implementing the Plan, Checking Progress, and Deciding When to Make Mid-course Corrections	30
Conclusion	32
Recommended Reading	32

2

DEVELOPING AN INTEGRATED AND COMPREHENSIVE MONITORING AND EVALUATION PLAN

Incorporating evaluation at the program design stage is an essential element of ensuring that evaluation activities will produce useful results. Planning an intervention and designing an evaluation strategy should be inseparable activities. To ensure the relevance

and sustainability of evaluation activities, project designers, in collaboration with national and local stakeholders and collaborating donors, must work in a participatory manner to develop an integrated and comprehensive evaluation plan.

The purpose of this chapter is to provide practical guidance to HIV prevention program managers at various levels in developing realistic integrated and comprehensive monitoring and evaluation plans. Depending on whether such plans are designed for the national or project level, they will likely consist of widely different objectives, indicators, and methodologies to obtain those indicators. National and large geographic areas, such as provinces, may focus more on impact-related measurements, such as prevalence and behavioral estimates for HIV and sexually transmitted infections (STIs). Project-level evaluation may more appropriately concentrate on process and immediate outcome indicators, such as number of people reached and STI

clients treated, as well as qualitative research methods to determine whether intervention strategies are appropriate for the target audiences.

Projects at all levels, whether they consist of multiple integrated projects or single interventions, should have an evaluation plan for assessing the progress of the program in achieving program goals and objectives and informing key stakeholders and program designers about the results of the evaluations. Such plans will guide the design of evaluations, highlight what information or data need to be collected, and how best to collect it.

Comprehensive evaluation plans should describe the overall purpose(s) of the evaluation, the specific evaluation questions to be addressed, the evaluation designs and methods to be used, what data are to be collected and how, the resources that will be necessary, who will implement the evaluation, and the basic evaluation plan timeline. They are often written to cover a 4- to 5-year period because they may involve

multiple evaluative efforts on multiple interventions for multiple target populations, some of which require time to observe intervention or program outcomes (immediate or short-term effects), as well as overall program impact (long-term effects).

This chapter applies the conceptual evaluation framework offered in Chapter 1 and describes the steps involved in developing an integrated and comprehensive evaluation plan, the factors that may influence this process, and key questions that should be asked by every program.

RATIONALE FOR DEVELOPING AN EVALUATION PLAN

It is often helpful in the beginning stages to review with all stakeholders the reasons for developing a comprehensive evaluation plan. Some of the benefits that can be derived from the evaluation planning process are:

- Evaluation planning will provide program managers and stakeholders alike with the opportunity to assess the evaluation needs, resources, capabilities, and priorities in their area.
- Having an evaluation plan will show stakeholders how the program plans to be accountable for the resources they have received.
- In the process of developing the evaluation plan, existing data sources and past or concurrent evaluation activities are often identified. Capitalizing on such existing data sources and past evaluative efforts can lead to a more efficient, less redundant plan for new evaluation monies.
- Having a long-term evaluation plan can clarify future decision-making regarding evaluation priorities.
- Finally, having a comprehensive evaluation plan in place may also favorably influence donor decision-making.

KEY ELEMENTS OF AN EVALUATION PLAN

In developing an evaluation plan, a planning group should focus on the following key elements:

- **Scope of the evaluation**—Specifying the goals and objectives of the program and developing a conceptual framework (or logic model) that integrates and correlates the inputs, activities, outputs, outcomes, and impact, and establishing realistic expectations for what the evaluation will provide or show.
- **Methodological approach**—Developing an evaluation design, including specifying the outcome indicators or measures and the data source and plans for data analysis.
- **Implementation plan**—Delineating activities, roles, and responsibilities, and a timetable for identified activities with realistic expectations of when data will be analyzed and results will be available.
- **Dissemination plan for the results**—Determining who will translate the results into language that is useful to program designers, managers, and decision makers, how findings will be disseminated (for example, written papers, oral presentations, program materials), and implications for setting priorities for future evaluation activities.

DEVELOPING AN INTEGRATED AND COMPREHENSIVE EVALUATION PLAN

In some cases, these elements are already in place and the task of the evaluation planning group is to pull them together into an integrated logical whole. If these elements have not been previously developed, then the following steps may be helpful in beginning an evaluation plan.

Evaluation planning generally involves several common steps. These steps are described here in a general way so as to allow for differences

among program types, target populations, stakeholders' needs for information, and the various levels such as a national program, a district or provincial response, and individual projects reaching specific target groups. Throughout the process, it is important to involve the various stakeholders: program planners, evaluators, Ministry of Health and National AIDS Control Program staff, as well as donors. Involving members of the target community also helps inform the evaluation planning process. This stakeholder involvement in the early phases helps to ensure that the evaluation results will be used in the end. As Box 2-1 shows, both internal and external perspectives are critical.

Step 1: Identify Program Goals and Objectives

The first step involves identifying the program goals and objectives and establishing a program logic model. This is done through writing a clear statement that identifies the program's goals and objectives (and sometimes sub-objectives) and describes how the program sees itself operating to achieve these objectives. When this is done, a program logic model can be easily diagrammed and used to establish an evaluation plan.

Each component of the program should have objectives and sub-objectives specified, and each objective and sub-objective should designate measures of success indicators that should be collected to determine progress. Table 2-1 illustrates a program logic model showing sample goals, objectives, activities, indicators, and methods to obtain indicators, at both the country and project levels.

This model illustrates the way that the role of national governments in monitoring and planning HIV prevention is distinct from, yet complements, the implementing strengths of individual projects at the local level. For example, individual projects do not typically conduct impact evaluation because their impacts are typically seen in tandem with those of other

projects that work synergistically at multiple levels toward the same goals. Impact evaluation is most appropriately measured in large geographic areas, such as a province or country, and examines whether the collective efforts of numerous projects are reaching their desired impacts. These impacts can be measured through the serosurveillance systems, which monitor trends in HIV and STI prevalence as

Evaluation teams: the value of internal and external perspectives

BOX 2-1

Individuals from inside as well as outside projects and their respective organizations are necessary for quality evaluations. These internal and external perspectives provide a number of benefits:

Internal perspectives

- bring a full understanding of the project to the evaluation
- are typically more familiar with the target groups' characteristics, needs, barriers to behavior change
- know the strengths and weaknesses of the organization
- can suggest how and whether the organization can implement recommendations suggested by the evaluation team

External perspectives

- bring technical expertise perhaps not available to the organization
- offer other models and strategies to the organization
- offer a fresh view of the project, unencumbered by organizational dynamics

TABLE 2-1 Program logic model of Project A in Country X

Narrative Summary	Measurable Indicators	Means of Verification	Important Assumptions
<p>GOAL To reduce sexual transmission of HIV infection and other STIs in Country X (time frame: several years)</p>	<ol style="list-style-type: none"> 1 Stabilization in gender specific HIV prevalence 2 Reduction in prevalence of other STIs in female sex workers (FSW) and male truckers 	<ul style="list-style-type: none"> • National HIV sentinel surveillance survey conducted by Ministry of Health (Measurable Indicator 1) • HIV/STI prevalence surveys in Year 1 and 5 conducted by project A (Measurable Indicator 2) 	<ul style="list-style-type: none"> • Stable HIV prevalence reflects low transmission • STI decreases among FSW and truckers signals declines in other population groups
<p>OBJECTIVES Reduce high-risk behaviors among the primary target populations of FSW and their clients (time frame: several years)</p>	<ol style="list-style-type: none"> 1 Increase of X percentage points in proportion of FSW who report using condom with last client and who report consistent condom use with all clients in past month 2 Increase of X percentage points in proportion of selected male sub-population groups who report condom use with last FSW and who report consistent condom use with FSW in the past 6 months 3 Decrease of X percentage points in proportion of male sub-population group who report having sex with a FSW in the past year 4 Other behavioral and related indicators as appropriate 	<ul style="list-style-type: none"> • Behavioral surveillance survey (BSS) in project area (Measurable Indicators 1-4) 	<ul style="list-style-type: none"> • Reported behavior is valid and reliable indicator of actual behavior • Outreach promotes risk reduction • Condoms continue to be available from government facilities and through social marketing
<p>ACTIVITY 1 Implement targeted behavior change intervention (time frame: several months to several years)</p>	<ol style="list-style-type: none"> 1 X number of target population selected and trained as peer educators in NGO 1,2,3,4 in project area by month X 2 X number of behavior change communication (BCC) materials distributed by month X 3 X number of target group (FSW, clients) reached by peer educators by month X 4 X number of target group (FSW, clients) referred to quality STI service 5 X spots aired (TV, radio) by month X 6 X events held (concerts, sports, theater) by month X 7 Improved quality of interventions 	<ul style="list-style-type: none"> • Process data from implementing organizations (Measurable Indicators 1-6) • Quarterly narrative report (Measurable Indicators 1-6) • Special qualitative research conducted with target groups (FSW, clients) every 6 months to evaluate reach and acceptability of interventions with target group (Measurable Indicator 7) 	<ul style="list-style-type: none"> • Research indicates intervention quality
<p>ACTIVITY 2 Implement improved STI control program, including increased use of quality services (time frame: several months to several years)</p>	<ol style="list-style-type: none"> 1 Increase in proportion of selected targeted health providers who provide quality STI care, including adherence to national standards and promotion of condoms and partner notification 2 Increase in use of improved services by STI clients 3 X number of health providers trained in appropriate STI services 4 X number of clients referred to facilities by outreach programs from NGO 1,2,3,4 	<ul style="list-style-type: none"> • Health provider survey through interview, observation, or mystery client and accompanying qualitative research (Measurable Indicator 1) • Clinic reports (Measurable Indicators 2-4) 	<ul style="list-style-type: none"> • STI prevention reduces risk of HIV transmission • Providers who have substantive numbers of STI clients can be reached for training
<p>ACTIVITY 3 Establish and/or expand condom promotion and distribution (social marketed and free) (time frame: several months to several years)</p>	<ol style="list-style-type: none"> 1 Increase in number of outlets that sell and distribute condoms in areas convenient for target groups 2 X number of condoms sold 3 X number of condoms distributed free 4 X number of promotional events conducted to market condoms 	<ul style="list-style-type: none"> • Process data from implementing organization (Measurable Indicators 1-4) • Quarterly narrative reports (Measurable Indicators 1-4) 	<ul style="list-style-type: none"> • Condoms sold and distributed are used
<p>ACTIVITY 4 Develop and increase capacity to design, manage, and implement quality HIV prevention programs (time frame: several months to several years)</p>	<ol style="list-style-type: none"> 1 Yearly strategic and work plan formulated 2 Increase in self-reported organizational capacity 3 X training sessions held in areas in need of improvement by month X 4 X people trained by month X 	<ul style="list-style-type: none"> • Organizational plans on file (Measurable Indicator 1) • Use of standardized organizational capacity assessments (Measurable Indicator 2) • Organizational reports (Measurable Indicators 3-4) 	<ul style="list-style-type: none"> • Increased capacity remains with organization

well as through repeated behavioral risk surveys. Local organizations, which have direct contact with target groups through outreach, STI treatment, and care, for example, should focus their evaluation efforts on assuring quality program implementation as opposed to outcome or impact evaluation. This means more of a concentration on quality inputs, such as training and pretesting of communication messages, but also includes the coverage and density of interventions.

This model also illustrates the length of time needed to show progress at various levels—ranging from several months for process-level accomplishments, such as staff trained, to several years for outcome- and impact-level goals.

If the information or data necessary to complete Step 1 are not readily available, then plans

the evaluation questions may need revision later in the evaluation plan development process.

Step 2: Examine Existing Data and Past Evaluation Studies

The second step consists of identifying existing data sources as well as other evaluative activities that may have been done in the past, are ongoing, and/or may have been sponsored by other donors. At this step, evaluation planning teams should assess whether other groups are planning similar evaluations and invite them to collaborate in the planning and coordination.

Step 3: Identify Internal and External Evaluation Resources and Capacity

Identifying evaluation resources means not only identifying the funds for the evaluation, but also

Local organizations, which have direct contact with target groups through outreach, STI treatment, and care, for example, should focus their evaluation efforts on assuring quality program implementation as opposed to outcome or impact evaluation.

should be developed to collect them. This step concludes with a written specification of the program logic model, using the framework provided in Table 1-1 in Chapter 1 and the format of Table 2-1, and a general statement about priority evaluation questions to be addressed in the evaluation plan.

This last aspect—framing the evaluation questions and then setting priorities among them—is sometimes quite difficult when there are multiple stakeholders involved and resources, time, and evaluation expertise are limited. Thus,

other types of resources, such as personnel experienced in evaluation to assist in planning and conducting the evaluation activities. It also means determining the program's capacity to manage and link various databases and computer systems.

Step 4: Determine Evaluation Questions, Their Feasibility, and Appropriate Designs and Indicators

In the fourth step, evaluation experts and program managers clarify the priority evaluation questions, appropriate evaluation designs, outcome measures or indicators, data needs, and

the methods by which this information will be collected and analyzed. Practical ways for obtaining data and maintaining a data system that is sustainable and easily accessed should be discussed. An operational plan for the comprehensive evaluation plan should be developed at this step as well. Thus, this step should conclude with a revised written plan briefly outlining the evaluation questions and evaluation design, data collection methods and analysis plan, and overall timeline for the comprehensive plan.

Step 5: Plan for Disseminating and Using Evaluation Findings

The fifth step is not always performed, but should be because it is extremely useful in ensuring that evaluation findings are used to inform program improvement and decision-making. This step involves planning how evaluation results will be used, translated into program policy language, and disseminated to all relevant stakeholders and decision makers. It should also involve a feedback loop to the planners of the next evaluation and a feedback mechanism should be built in so that past lessons learned can effectively inform new efforts.

It is important that evaluation findings be translated into language that is useful for program designers in their efforts to improve programs and that they contain implications and actions needed, if appropriate, at national and/or local levels. This fifth step is often best conducted in collaboration with individuals or groups familiar with the intervention. Evaluation results, like all messages, must be marketed and “packaged” for their target audiences. (Chapter 13, “Effective Dissemination of Data Collection Results,” provides useful guidance on ways in which programs can better carry out this step of the evaluation planning process.)

This step is often overlooked in the planning and implementation phase of the evaluation, and often only becomes an issue when there is a

problem at the end of the evaluation and someone asks the question, “How has this evaluation plan been implemented and how have the results been used to improve HIV prevention programs and policy?” If there has been no plan for disseminating results and using the findings, this question often cannot be answered because the people involved in the evaluation have forgotten the details or have moved on. The lack of such a plan can undermine the usefulness of the evaluation and future activities. Inadequate dissemination also commonly leads to redundancy in evaluation efforts because others are not aware of the findings of previous evaluation efforts. It also reinforces a negative stereotype about evaluations, which is that they are not really intended to help improve programs. For all these reasons, programs should include a plan for disseminating evaluation findings in their overall evaluation plan.

USING TECHNICAL SUPPORT

After an evaluation plan has been developed following the above steps and including the essential elements described, program managers are faced with the question, “How are we really going to implement this plan?” Clearly, the success of the plan depends on the technical capacity of the program and its associated staff to carry out the evaluation activities. This invariably requires evaluation technical assistance. Evaluation expertise typically can be provided by individuals internal or external to the Ministry of Health or National AIDS Control Program. Such individuals usually have knowledge and training in the theories and methods of applied program evaluation. Potential evaluation experts might be found in planning and evaluation units in the Ministry of Health, local academic institutions, non-governmental organizations, and private consulting firms.

Technical capacity at the field level to design, implement, and maintain data collection systems is important to ensure the uninterrupted flow of consistent data. Participation of local people with a vested interest in the results should occur throughout the process of planning, data collection, analysis, and feedback.

Some traditional methodologies in evaluation, such as surveys on knowledge, attitudes, behaviors, and practices, may not always be appropriate for all evaluations. Technical guidance is especially important in complex evaluation designs, such as quantitative surveys where sampling designs, sample sizes, and questionnaire design must be carefully determined (see Box 2-2).

IMPLEMENTING THE PLAN, CHECKING PROGRESS, AND DECIDING WHEN TO MAKE MID-COURSE CORRECTIONS

A comprehensive evaluation plan as described above has multiple components, some of which need to occur in a sequential fashion. Additionally, sufficient evaluation resources (both human and fiscal) for the entire plan are often not available at the launching of the plan. Thus, implementing components of the evaluation plan in phases is not only logical and practical, it is also essential.

Carrying out and completing components of the evaluation in phases provides the opportunity for the evaluation planning group to assess how well the evaluation plan is working. Sometime after the group completes the first or second evaluation activity, the plan may need some revisions or corrections. It is helpful for the group to spend time at this point ensuring that the plan is still useful. This process need not be very involved. The group can begin by discussing the following questions:

- Are the evaluation activities going as planned?
- Are the evaluation questions that were initially posed being sufficiently answered? Are other data needed to answer these questions? How can such data be obtained?
- Do the evaluation questions themselves need re-framing? Have other evaluation questions arisen since the initial planning that need to be incorporated into the plan at this point?
- Are there any methodological issues that need addressing or changes that need to be made to the evaluation designs? Are there any other factors, practical or political, that need to be considered in the evaluation activities yet to be implemented?
- Are any changes in the plan needed at this time? How will these changes be made? Who will implement them?
- Is the right mix of personnel and fiscal resources still available to carry out the rest of the evaluation plan?
- How are findings from the evaluation activities so far being used and disseminated? Does anything need to be done to enhance their application to programs?

This assessment is most helpful if it occurs annually throughout the plan. If the plan is kept current and relevant, it will help ensure the usefulness and quality of the remaining components. It will also foster the overall success and usefulness of all of the evaluation activities by the end of the project period. Reviewing the evaluation plan and doing mid-course corrections as needed also facilitates the connection between the evaluation activities and the programs as well as the design of subsequent plans.

KABP surveys in every project: useful or not?

BOX 2-2

Knowledge, attitude, beliefs, and practices (KABP) surveys were a staple of public health research long before the advent of HIV. However, the advent of the epidemic brought a flood of surveys that researched sexual knowledge and behavior. While many of these surveys have provided rich information guiding intervention design and policy efforts, the results of some surveys have been difficult, if not impossible, to interpret and potentially harmful to prevention efforts because of careless sampling, quickly constructed questionnaires, untrained interviewers, and overall poor survey implementation.

These quality control issues are found in large-scale national surveys as well as project-level surveys. For example, a recent review of 33 national-level behavioral surveys noted substantial methodological shortcomings in design and, of the surveys reviewed, only two had multiple data points from which population-level trends could be deduced.

Donors and project managers have historically promoted pre- and post-intervention KABP surveys as an appropriate way to measure the outcome of interventions. But these project-level surveys, as a whole, have yielded questionably useful data because of the quality control issues mentioned above. There are several challenges in the use of quantitative surveys to evaluate project outcomes:

- **Methodological limitations to measuring program outcomes**—The effects of a specific project can only be truly measured through the use of control and/or comparison groups used in experimental or quasi-experimental designs. These designs are inappropriate for most service delivery projects because they require extensive funding and research expertise and they necessitate withholding interventions to certain groups

for research purposes. Without control groups, behavioral change observed in pre- and post-surveys cannot solely be attributed to a specific project because other influences may be promoting change as well.

- **Limited research capacity in intervention organizations**—Even if the interpretation limitations mentioned above are accepted, behavioral surveys to measure program outcomes are extremely complex, necessitating carefully designed sampling plans, measurement of intervention exposure, and multivariate statistical analyses of results. Typically, these skills are not found in organizations that specialize in interventions. Because of external and internal pressures to show results, surveys that compromise quality have often been carried out, however, leading to erroneous results and frustrated relationships between donors and implementing organizations.

- **Disregard of synergistic, mutually-reinforcing interventions**—Donors and program managers too often concentrate on measuring the effects of one intervention, disregarding the preponderance of research evidence that suggests that the most effective HIV prevention is a mixture of different types of interventions working together synergistically. Individuals and communities change behavior because of a variety of influences and interventions, such as mass media communications, individualized outreach, counseling in STI treatment settings, the participation of people with HIV/AIDS in policy and public discussions, and condom social marketing, to name but a few. While certain interventions still require specialized and controlled research to investigate their effects, most service delivery projects should use tested interventions, monitor the quality of implementation, and work harmoniously with other project interventions toward collective results.

CONCLUSION

Although projects may differ greatly in their scope, target audience, objectives, and activities, they all need to include realistic and comprehensive monitoring and evaluation plans. Incorporating evaluation in the program planning and design phase is essential to ensuring that evaluation activities will produce useful results. Involving national and local stakeholders and collaborating donors is an essential element of successful evaluation activities, as is broad dissemination of results.

This chapter has described a five-step process to guide program managers in developing evaluation plans that describe the purpose of the evaluation, the specific questions to be addressed, the evaluation designs and methods, the resources that will be needed, the people involved, and the timeline. It also has presented a program logic model that evaluation planners can use as a framework for laying out the evaluation plan. By following these steps, program planners can greatly enhance their chances of conducting evaluations that are useful, informative, and timely for HIV prevention efforts.

RECOMMENDED READING

- Academy for Educational Development. Critical issues in HIV prevention evaluation: a monograph. Washington (DC): AED; 1998. Funding provided by the Centers for Disease Control and Prevention.
- Catania JA, Moskowitz JT, Ruiz M, Cleland J. A review of national AIDS-related behavioral surveys. *AIDS* 1996; 10 (Suppl A): S183-S190.
- Centers for Disease Control and Prevention (CDC). HIV prevention evaluation guidance. Atlanta (GA): CDC; 1998.
- Chelimsky E, Shadish WR, editors. *Evaluation for the 21st century: a handbook*. Thousand Oaks (CA): Sage Publications; 1997. p. 145-201.
- Coyle S, Boruch R, Turner C, editors. *Evaluating AIDS prevention programs*. Washington (DC): National Academy Press; 1991. p. 102-123.
- Cronbach LJ. *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass; 1982.
- Fetterman DM, Kaftarian SJ, Wandersman A, editors. *Empowerment evaluation: knowledge and tools for self-assessment and accountability*. Thousand Oaks (CA): Sage Publications; 1996. p. 100-207.
- Patton MQ, editor. *Utilization-focused evaluation*. 3rd edition. Thousand Oaks (CA): Sage Publications; 1997.
- Rossi PH, Freeman HE, editors. *Evaluation: a systematic approach*. 5th edition. Newbury Park (CA): Sage Publications, Inc.; 1993.
- Rugg DL, Buehler J, Renaud M, et al. Evaluating HIV prevention: a framework for national, state, and local levels. *Am J Eval* 1999;20(1):35-56.

II

OPERATIONAL APPROACHES FOR EVALUATING INTERVENTION STRATEGIES

Evaluating Behavior Change Communication Interventions	35
Evaluating Sexually Transmitted Infection Control Programs	47
Evaluating Condom Programming	65
Evaluating Voluntary HIV Counseling and Testing Programs	79
Evaluating Care Programs for People Living with HIV/AIDS	91

3

EVALUATING BEHAVIOR CHANGE COMMUNICATION INTERVENTIONS

BARBARA FRANKLIN, DONNA FLANAGAN, AND HALLY MAHLER

Audience-centered Communication	37
Planning BCC	38
Conduct Situation Analysis	38
Conduct Audience Analysis	39
Develop BCC Objectives	40
Developing BCC	41
Pretest Communication Materials	41
Implementing BCC	42
Conduct Process Evaluation	42
Assess Communication Effects	44
Conclusion	45
Recommended Reading	46

3

EVALUATING BEHAVIOR CHANGE COMMUNICATION INTERVENTIONS

This chapter will look at the essential evaluation activities performed at different stages of implementing behavior change communication (BCC) projects and discuss critical issues and inherent challenges. Behavior change communication is one component of behavior

change interventions (BCI), a broader designation that includes many other types of interventions, such as services, commodity distribution, and community mobilization, in addition to communication interventions.

The evaluation of a BCC project is an integral part of its design and must be considered at different stages throughout the project's development: during the stages of formative research (situation analysis and audience analysis), during pretesting, while monitoring implementation, as well as during the evaluation of communication effects. In other words, evaluation of BCC projects must be considered an integral part of the audience-centered communication cycle, with implications for each of these stages.

AUDIENCE-CENTERED COMMUNICATION

Audience-centered communication is an approach to BCC based on a model of dialogue between

those who wish to promote behavior change and a participating audience. It is essentially a consultative partnership, focusing on interaction at every stage between the communicating institution and the public. While a number of methodological tools are available (focus groups, key informant interviews, and various participatory learning and action [PLA] techniques, among others), no technique will be truly effective unless it is imbued with an attitude of respect for the target audience and a determination to understand the audience's point of view and address the audience's central concerns throughout the BCC campaign, culminating in its evaluation.

Audience-centered communication has three stages. The first stage determines the parameters for effective BCC through a series of research and planning activities. During this stage, communicators look at research already done, and, if necessary, carry out two types of formative research: a situation analysis and an audience analysis.

The second stage involves developing a communication concept and messages based on the formative research findings, drafting materials, and pretesting them with the target audience. The third stage involves implementing BCC activities, and includes monitoring communication activities, checking comprehension, and assessing communication effects. While evaluation is often considered a separate phase that follows implementation, experience shows that it is more useful to think of it as an integral part of a communication intervention. This emphasizes the need to generate immediate feedback, verify comprehension, and develop an awareness of effects. The steps of the audience-centered communication cycle are listed below. Activities with a particular relevance to evaluation are in bold face.

I PLANNING BCC

- Review baseline data
- Conduct situation analysis**
- Segment the audience
- Designate target audience(s)
- Establish behavior change goal
- Conduct audience analysis**
- Develop BCC objectives**
- Make action plan

II DEVELOPING BCC

- Develop BCC concept
- Develop messages
- Choose channels
- Develop BCC materials
- Pretest BCC materials**
- Produce BCC materials

III IMPLEMENTING BCC

- Carry out BCC activities
- Conduct process evaluation**
- Assess communication effects**

The evaluation activities conducted at each of these stages are described in greater detail below.

PLANNING BCC

Two types of formative research should be conducted at the planning stage of a BCC project to gather information to guide the project strategies. Planners need to investigate both the individual and community environment before formulating a strategy for behavior change communication. This research is the basis for developing communication objectives—the criteria on which the BCC project ultimately will be evaluated.

Conduct Situation Analysis

The situation analysis (also called community investigation) reveals the parameters within which BCC will take place. It includes an analysis of the:

- **Demographic environment**—The population in terms of size, density, location, age, gender, race, occupation, education, income, family composition, and other statistics.
- **Epidemiological environment**—The incidence and prevalence of sexually transmitted infections (STIs), including HIV, among various target audiences.
- **Economic environment**—The factors that affect people’s purchasing power and spending patterns, which will permit a better understanding of how, for example, prostitution, condoms, and health care are purchased. An understanding of the macroeconomics of a country helps explain the role that the infrastructure, the media, and the market will play in distributing health care messages and facilitating programs.
- **Political environment**—The laws, government policies, agencies, and advocacy groups that influence and limit various organizations and activities. The political environment can affect condom advertising, public health, access to health care,

the economic power of women, the sex trade, sex education, the operation of STI clinics, and the illicit drug trade.

- **Cultural environment**—The institutions and other forces that shape society’s basic values, perceptions, preferences, and behaviors. Religion, language, educational institutions, literature, popular music, the press, and theater all play a role in determining the status of women and in influencing sexual behavior, attitudes toward AIDS, and family values.
- **Organizational/development environment**—The programs, projects, and interventions that are already in place so that complementary and/or shared efforts can be encouraged and duplication avoided.

The situation analysis tells how the community is organized and how it works—essential parameters within which any BCC activity must function. A carefully conducted situation analysis can save time and money by helping to ensure that BCC activities are chosen well and that they correspond to the realities of the situation, avoid traps, and make best use of their opportunities. It also takes into consideration the need to create a supportive environment, which may lead to a focus on advocacy or community development issues needed to support the individual changes BCC is promoting.

On the basis of the situation analysis, the BCC planning team can segment the audience, and choose target audiences from among the segments for specific BCC activities. A target audience is defined as the group that needs to change its knowledge, attitudes, or behaviors to achieve the program goal. Target audiences for effective BCC include not only groups at risk because of their behavior, but also others who have an impact on the risk situation, including authorities (such as police), gatekeepers (such as brothel owners), and service providers (such as STI clinic staff).

Conduct Audience Analysis

The situation analysis describes the community and larger environment within which change takes place, but ultimately much of the behavior change with respect to HIV/AIDS and STIs takes place on an individual level. Behavior change is usually not a one-time, definitive event, but rather is preceded by a series of steps, including changes in knowledge, shifts in attitude, and tentative preliminary steps to behavior change, such as trying a behavior once. For this reason, developing an effective BCC program requires an in-depth “insider” understanding of the members of each of the target audience segments and where the segment as a whole is on the steps of change at any specific time—their level of knowledge, their relevant attitudes (beliefs, values, preferences, expectations), the barriers to change they face and express, and the factors that might motivate them toward behavior change (as well as their communication preferences, essential information for a strategic choice of channels.)

The audience analysis should seek to provide a clear picture of the interior reality of members of each segment and should provide answers to the following questions:

- How do members of the target audience segments conceive of HIV/AIDS and STIs?
- Where are they in the process of behavior change?
- What is their level of knowledge of the specific facts of HIV/AIDS and STIs, their transmission and prevention?
- What is their estimation of their personal risk of becoming infected with HIV and STIs? If they feel they are not at risk, why not?
- What are their perceptions of peer and social norms governing sexual behaviors and HIV prevention?
- What peer groups and significant others are most important to them?

- What are the behaviors that put them at risk for HIV and STIs?
- How do they currently understand and practice preventive behaviors?
- What do they see as the benefits of changing their behaviors?
- What do they see as the disadvantages of changing behavior? What pressures make it difficult for them to change their behaviors?
- What power do they believe they have to change their behavior, and if it is limited, why, and by whom or what?

Develop BCC Objectives

On the basis of the situation analysis and in-depth audience analysis, the BCC planning team then develops communication objectives. These objectives should specify the nature of the change the BCC campaign will engender. They should not be limited to behavior changes but should also include changes in knowledge, attitudes or the decision or intention to try behaviors that precede definitive behavior change. Experience shows, for example, that it is often unrealistic to expect that a BCC campaign will bring about an

The process of behavior change is complex and incremental, and BCC planners can best approach it on those terms—like chipping away at a big stone, rather than trying to move it with a single push.

- What would it take for them to change to a safer behavior or continue with safe behavior?
- When and where do they usually get information about sexual and health topics?
- How do they communicate with others? Where do they meet? When? Who is in their immediate social network?
- What sources of information on sex and health do they find most credible?
- When and where would be the best times to talk with them about HIV prevention, distribute condoms, make STI treatment services available, and who would do this most effectively?

immediate increase in condom use, but a lack of change in condom use does not necessarily mean that nothing has changed. Important changes may have occurred and may be preparing the way for later behavior change. The process of behavior change is complex and incremental, and BCC planners can best approach it on those terms—like chipping away at a big stone, rather than trying to move it with a single push. Well-formulated BCC objectives recognize these steps of change.

Knowledge changes are the easiest to identify, and as a result, many BCC campaigns focus on increasing knowledge long past the time when lack of knowledge is really the problem for a target audience. Attitude changes are often even more important but neglected by communicators. For example, audience members may

change their idea of condoms from a pregnancy prevention device, to something that can keep them safe from STIs and HIV, or they may increase their personal perceived risk of HIV/AIDS. Norms may also change; it may become more usual for women to negotiate the terms of sexual behavior as the result of a BCC campaign, or people may develop new concepts of masculinity, or a sense that condoms can be sexy. There may also be changes in law or policy, or changes in the public discourse, as reflected in the media.

Communication objectives should be stated in the future tense, in terms of changes in knowledge, attitudes, and skills or in the policy environment that will be evident after the BCC activity, and they should directly reflect needs identified through the in-depth audience analysis. Here, for example, are some communication objectives aiming at changes that precede behavior change. Note that they are stated in behavioral terms so that their success is easy to measure:

- **Change in knowledge**—After the mass media campaign, audience members will say they cannot get HIV through haircuts, manicures, or mosquito bites, or by sharing clothes or dishes.
- **Change in attitude (beliefs)**—After the campaign, when asked what they think of when they hear the word “condom,” audience members will more often use words reflecting a positive image of strength and protection.
- **Change in attitude (perceived risk)**—After the workshop, audience members will say it is quite possible that they or their friends could get HIV if they have sex outside of a monogamous relationship and do not use condoms.
- **Change in skills**—After seeing the video and practicing with the peer educator, audience members will be able to show how they would negotiate condom use with a partner.

- **Change in the discourse**—After the campaign, a survey of the popular press will reveal less judgmental language applied to sex workers.

DEVELOPING BCC

Pretest Communication Materials

Once the BCC planning team has written communication objectives, developed concepts, chosen the media channel(s), designed messages (that is, the specific words and images) to express the concept, and developed prototype materials to match the objectives, the materials must be pretested with the target audience. Pretesting is an evaluation activity that gives immediate feedback on the effectiveness of BCC messages and materials. It is an early evaluation of the BCC planning process that can and should lead to immediate adjustments in the materials or methods to be used in the BCC project.

Thorough and systematic pretesting is an essential part of marketing and plays no small role in the success of commercial advertising—a success social marketers have emulated. For example, in Nepal’s AIDSCAP II project, implemented by FHI, alternative images of a condom were thoroughly pretested to find one that would appeal to a wide spectrum of audience members and would carry a strong association of condoms with AIDS prevention. A number of options were rejected based on the pretests, and the winning image was further refined. The resulting image of “Dhaaley Dai” (Big Brother Condom) shows a smiling, muscular condom character kicking out a small figure representing the virus. It is accompanied by a slogan (rhyming in Nepali): “Wear a condom and drive away AIDS.” This message has been widely disseminated in Nepal and has achieved a very high level of recognition.

There are a number of good pretesting techniques, including focus group pretesting, and

pretesting through individual or group interviews. Alternative versions of materials should always be pretested, so that the pretest has meaning. While the pretest may look like an informal discussion, it should be structured around a definite set of questions that reveal the

different qualities for which the materials are being tested. For example, interviewers could use the question guide in Box 3-1 to pretest a pamphlet with groups or individuals.

IMPLEMENTING BCC

Conduct Process Evaluation

Process evaluation is conducted throughout the implementation of BCC activities and actively involves key stakeholders, such as project managers, beneficiaries, organization staff, and donors. This participatory approach to process evaluation allows the stakeholders themselves to identify the essential indicators they want to measure and report on and helps to insure that the evaluation will be relevant and useful for designing future activities. Its purpose is to determine whether activities are proceeding according to the plan and if not, to indicate where changes need to be made. Questions asked during process evaluation obviously reflect the activities of the program. They might include such questions as the following (for a peer education activity):

- Were peer educators selected, trained, and supervised?
- Are the peer educators performing the duties that were expected of them?
- Is supervision being conducted as planned?
- Are the communication channels being used as planned?
- Have the radio or TV messages been broadcast?
- Was the target audience involved with message development?
- How many target group members have been reached?

Process evaluation can also examine strengths and weaknesses of an ongoing intervention. What follows is an instrument for assessing and monitoring behavior change communication interventions that sets out standards for effective projects. It then asks specific questions that lead the user to recognize whether the intervention is adhering to the established standards for high-quality BCC interventions. These questions are meant to be asked throughout the implementation of BCC interventions.

Standard 1: Interventions should focus on well-characterized, specific target audiences.

- 1 Who is the primary target audience for this BCC intervention?
- 2 Has this primary target group been appropriately divided by segmenting variables? If not, which variables have not been considered that now appear important to the segmentation?
- 3 Are there other people who influence the primary target group who are not yet being addressed? If yes, who are they?
- 4 How can the project address these other people?
- 5 What is the risk behavior(s) that the primary target audience is practicing? What is the desired behavior?

Standard 2: HIV/AIDS prevention interventions and messages must be crafted to motivate and appeal to the specific target audience's perceived needs, beliefs, concerns, attitudes, present practices, and readiness to change.

- 1 What additional knowledge is needed, what attitudes need to change, and which skills need to be mastered before the target audience will be able to adopt the desired behavior?
- 2 What are the main messages used in this intervention? Do the main messages address the needed knowledge, attitudes,

and skills? If not, what is missing, or what does not match these needs? (To answer this question more specifically, the following additional questions are useful.)

- 3 What gaps in knowledge appear to influence audience members behavior with respect to HIV/AIDS/STIs? Do the messages match the audience's gaps in knowledge?
- 4 What does the audience perceive as its most important needs? Do the messages match the audience's perceived needs?
- 5 What are the audience's main beliefs related to sexuality and HIV/AIDS/STIs? Do the messages make appropriate use of (appeal or respond to) these beliefs?
- 6 What are the main concerns of the audience members? Do the messages refer to these concerns?
- 7 What attitudes presently inhibit change in the audience? Do the messages respond to the attitudes and encourage/model different attitudes?
- 8 What undesirable behaviors do audience members currently have? Do the messages specifically address the disadvantages of these undesirable behaviors?

Standard 3: At-risk individuals must be provided with both skills and supplies to prevent HIV.

- 1 Are any new skills needed for the audience members to change? If so, what skills? Do the messages model the needed skills?
- 2 Are supplies that are needed for safe behavior available to all the audience members?
- 3 Are supplies that are needed for safe behavior affordable for all the audience members?
- 4 Is STI treatment easily available and affordable for all the audience members?

BOX 3-1

Sample question guide

Ask these questions about EACH PAGE of the pamphlet

- 1 Are there any words on this page that you do not understand? What are they? Can you explain why they are unclear?
- 2 Is there anything on the page that you do not believe to be true? What? What would you find more believable?
- 3 What about the picture on this page? Tell me what you see. (If appropriate ask, "Where is this place? What kind of person is this?")
- 4 Is there anything about the pictures or the writing on this page that might offend or embarrass some people? What? What would be better?
- 5 Is there anything on this page that is confusing? What? How could it be clearer?
- 6 Is there anything on this page that you really like? What?
- 7 Is there anything on this page that you don't like? What? How would you like to change it?

Ask these questions about the ENTIRE PAMPHLET

- 8 Do you think the pamphlet is asking you to do anything in particular? What? Do you think you would do it, and if not, why not?
- 9 Do you think the pamphlet is meant for people like yourself, or is it for other people? Why?
- 10 What do you think can be done to make this a better pamphlet?

Ask the above questions for EACH VERSION of the pamphlet, then ask this question

- 11 Which version of this pamphlet do you prefer? Why?

Standard 4: A supportive environment needs to be created for HIV prevention and for the protection of those infected with HIV.

- 1 What are the social, cultural, environmental, political, and organizational conditions that may influence the target audience's HIV/AIDS risk behaviors?
- 2 Does this intervention try to influence these social, cultural, environmental, political and/or organizational factors? For example, does it:
 - support traditional and cultural values that encourage low-risk behaviors?
 - persuade government officials to change public health policies?
 - influence organization/corporate officials to discontinue discriminatory practices or policies?
 - mobilize support among the general public to work for changes in public policy?
 - promote the social acceptability of alternatives to risk behaviors?
 - protect human rights of all people affected by HIV/AIDS?
 - actively fight discrimination?
 - educate the whole community for care, compassion, and prevention?

Standard 5: Mechanisms need to be created to maintain and sustain HIV prevention behaviors and activities over time.

- 1 Does this BCC intervention include follow-up mechanisms to reinforce and encourage the maintenance of newly acquired attitudes and behaviors? For example:
 - periodic follow-ups and re-certification of peer educators;
 - HIV prevention messages mainstreamed into school curriculum at all grade levels;
 - campaigns to reinforce messages focused on maintaining new behaviors;

- annual meetings for organizations working in the HIV prevention area;
- meetings organized to discuss “lessons learned.”

Standard 6: BCC planners should identify and use opportunities to work collaboratively and in different sectors of the community/country.

- 1 Does this intervention actively collaborate with other partners and implementing agencies?
- 2 Does this intervention take into consideration activities and materials aimed at this target audience by other organizations?
- 3 Is this intervention designed to involve the resources and expertise of other individuals and organizations in the public and private sectors?

Standard 7: A monitoring plan is essential to guide the adequate implementation of behavior change communication projects.

- 1 Does this intervention have a monitoring budget?
- 2 Does this intervention have staff available for monitoring and supervision?
- 3 Have new directions been identified as a result of monitoring? If so, what are they?

Assess Communication Effects

Monitoring an ongoing BCC program according to the above standards will go far toward developing an understanding of the effectiveness of that program. However, it is also useful to pause, after the conclusion of a BCC program or campaign, to take a close look at the net effect of that program on the target audience. This can be done using quantitative or qualitative methods.

It is possible to conduct a systematic, quantitative evaluation of any change in knowledge, attitude, intention, skill, or reported behaviors (actually a measure of norms) if these are defined specifically enough in the objectives,

and if an appropriate methodology is chosen (Section III of this book provides more information on the methodologies for measuring behavioral outcomes).

However, this type of outcome evaluation is often not necessary at the individual project level (see also the section on “Differentiating Evaluation Efforts” in Chapter 1). A simpler method, and one that usually yields richer insights, is to repeat some of the in-depth formative research with different members of the same target audience segments, to identify any evidence of change. This method, called assessment of communication effects, does not attempt to prove that a certain communication campaign caused a certain change (such causality is misplaced, in any case, when dealing with such a profoundly social realm as communication). Rather, it is an attempt to take a second look at the audience, to see how things have changed since the BCC program began.

For example, suppose that a negative image of condoms has been identified as an attitudinal barrier as a result of focus groups or in-depth interviews in which participants referred to condoms in a variety of negative ways, and a campaign has been conducted to change the image of condoms. After the campaign, follow-up focus groups or interviews with the same segments may reveal some change in the discourse on condoms. A striking example from Vietnam came when follow-up focus groups were held after a BCC campaign in which one of the objectives was to give condoms a more friendly, protective, and less medical image. When the groups were asked, “What do you first think of when you think of condoms?” (a question also asked in the original research) many participants responded, “condoms are like a loyal body-guard.” This phrase, echoing a message from the BCC campaign, provided evidence that the campaign had had its intended effect.

Assessing communication effects can also produce data about exposure to media messages, about recall of specific messages, and about the relative impact of different communication channels. An assessment of communication effects not only reflects the success or failure of a previous BCC campaign, but also provides feedback on audience concerns that can be addressed in the next campaign. In this way, the BCC process can become more like a dialogue between the BCC planning team and the audience. Thus, the assessment of communication effects guides future BCC campaigns, closing the planning-implementation-evaluation circle.

CONCLUSION

This chapter has shown how evaluation is an integral part of the design of audience-centered communication projects, not only at the final stage but also at other key points during the project cycle. This ongoing concern with evaluation from the outset helps to ensure that the communication activities are structured around a dialogue between the members of the target audience and the BCC planning team. Elements of evaluation are included in the situation analysis and audience analysis (formative research), in setting BCC objectives, in pretesting of materials, as well as in the process evaluation (monitoring of implementation) and the final assessment of communication effects. Assessing communication effects allows the audience to give feedback on the campaign as well as on changes within the audience itself relating to the campaign objectives. This then allows the BCC team to plan the next campaign. Thus, the final assessment closes the audience-centered communication cycle and assures that the process maintains its essential character of a dialogue between BCC planners and the audience.

RECOMMENDED READING

- Bertrand JT. Pretesting communication materials. Chicago: Communication Laboratory, Community Family Study Center, University of Chicago; 1978.
- Franklin B. Targeting young men: audience-centered communication for AIDS prevention in Vietnam. Monograph Series No. 4. Hanoi: CARE International in Vietnam; July, 1994.
- McKenzie J, Smeltzer JL. Planning, implementing, and evaluating health promotion programs. 3rd ed. Needham Heights (MA): Allyn & Bacon; 2000.
- Piotrow PT, Kincaid DL, Rimon JG, Rinehart W, editors. Health communications: lessons from family planning and reproductive health. Westport (CT): Greenwood Publishing Group; 1997.
- Rogers EM. Diffusion of innovations. 4th ed. New York: Free Press; 1995.

4

EVALUATING SEXUALLY TRANSMITTED INFECTION CONTROL PROGRAMS

GINA DALLABETTA, RICHARD STEEN, TOBI SAIDEL, ANDRÉ MEHEUS

STI Control Considerations	50
Monitoring and Evaluating STI Control Programs	51
Formative Evaluation	51
Process Evaluation: Inputs and Outputs	53
Effectiveness Evaluation: Outcomes and Impact	56
Other Considerations	61
Laboratory Issues	61
STIs as Biologic Proxies for HIV Transmission	61
Conclusion	63
References	64

4

EVALUATING SEXUALLY TRANSMITTED INFECTION CONTROL PROGRAMS

Sexually transmitted infections (STIs) are major causes of reproductive morbidity and mortality. Because many of the widely prevalent STIs also facilitate transmission of human immunodeficiency virus (HIV), STI control has been recognized to be an

effective means of reducing HIV transmission. The proportion of new HIV infections attributable to STI cofactor effects—the population-attributable fraction—is larger in populations or subpopulations with HIV infection in a growth phase than in populations where the HIV epidemic is mature. In keeping with the rest of this Handbook, this chapter considers evaluation of STI control programs in the context of HIV prevention, with reduced prevalence and incidence of STIs as a desired goal contributing to reduced HIV transmission at the impact level. In this context then, monitoring and evaluating STI programs has two goals. The first is to ensure that STI services are proficient (both clinically and with respect to coverage) in managing and preventing STIs. The second is to obtain proxy biologic measures for both the potential risk of HIV infection spread in a population and the impact of HIV prevention interventions. The relationship between HIV and STIs is complex,

however, and the use of STIs as a proxy for HIV is discussed in greater detail later in this chapter.

The term “sexually transmitted infection” (STI) refers to infections that may or may not cause symptoms and recognizes the asymptomatic and sub-clinical nature of many of these infections. The term “sexually transmitted disease” (STD) refers to infections with sexually transmitted pathogens that cause recognized symptoms or clinical signs in individuals. Throughout this chapter and the remainder of the Handbook, the term STI will be used except in specific instances when referring to symptomatic individuals.

Much more experience exists in monitoring and evaluating STI control programs than with most other areas of HIV prevention and care. This reflects the fact that STI control has been a disease control strategy in its own right and the fact that STIs were identified almost a decade ago as a key component of HIV prevention interventions. In this chapter, we provide general

guidelines for monitoring and evaluating STI control programs and for STI surveillance. While the specific target populations, target STI pathogens, and the mix of approaches and strategies for STI control will vary from one country to another and from one program to another, this chapter is intended to serve as a practical reference for service providers, STI and HIV program managers, and others called upon to evaluate STI control programs.

STI CONTROL CONSIDERATIONS

Effective STI control requires a coordinated series of interventions at both community and clinic levels to facilitate access to services, to ensure that services are appropriate to the needs of those at greatest risk, and to ensure that the services are of high quality. An explanatory model of STI transmission dynamics and barriers to prevention and

control should thus guide both the design and evaluation of interventions. Two commonly cited models are helpful for illustrating elements that should be addressed in evaluation. Health systems models (Figure 4-1) take as a starting point the subset of the population with symptomatic STIs and analyze barriers to identification and effective treatment. They are most useful for pointing out areas where services could be improved to maximize the proportion of STI cases that are effectively managed and rendered non-infectious. Limitations of such models are their focus on formal health care services and their treatment of populations as homogeneous with respect to STI transmission dynamics within a population.

Transmission dynamics models (Figure 4-2) address questions about which subgroups of the population are most at risk of acquiring and transmitting STIs. Such models better explain what occurs at the community level with STI transmission than do health system models and

can be useful for orienting interventions to those who can most benefit from them and have the largest impact at the population level. While less informative about needs at the service delivery level, analysis and evaluation within this framework is better suited to pointing out significant gaps (such as poor access to services for those at greatest STI risk) that could have important implications for STI control.

Using both of these models together, an explanatory model for designing and evaluating STI control efforts can be constructed that considers aspects of coverage as well as proficiency of individual service elements. With information from formative research, such models can be adapted to fit local conditions and contribute to effective intervention design. Once an intervention is designed and program goals and objectives are clearly defined, appropriate indicators to monitor and evaluate the STI control effort can be selected. Table 4-1 lists illustrative indicators and data sources that may be useful in addressing the different levels of monitoring and evaluation for STI control programs:

- process evaluation, including inputs, routine data collection on activities, and special study output information; and
- effectiveness evaluation, including intermediate program outcomes and expected program impact.

MONITORING AND EVALUATING STI CONTROL PROGRAMS

Formative Evaluation

As a first step in the program development cycle, information should be collected to describe the current situation, identify needs, and orient an effective response. A situation analysis should consider available data on STI prevalence and incidence, capacity and quality of existing services, current use of those services, as well as access and acceptability issues. The following are some components of a situation analysis for a program seeking to institute or improve STI interventions:

- review of existing STI data;
- community mapping to identify priority areas for interventions and availability of formal and informal sources of care;
- health facility audit to assess availability of prevention and treatment supplies;
- assessment of the quality of STI case management to define training and support needs and establish a baseline;
- key informant interviews in the community to better understand issues related to commercial sex or other high-frequency STI transmission dynamics; and
- focus group discussions with sex workers, clients, and “gatekeepers” to identify

FIGURE 4-1 Health systems models identifying barriers to effective management in persons with STIs

	Barrier to Control	Program Objective
MORE PEOPLE ↓ FEWER PEOPLE	Many people with STIs do not recognize symptoms or are unaware of risk.	Increased awareness of symptoms/risk
	Many people with symptoms delay seeking care or consult unqualified sources.	Improved health care seeking behavior
	Many STIs are not detected at health care facilities.	Effective STI detection
	Detected STIs may be treated incorrectly.	Effective STI case management
	Reinfection is likely if partner treatment and prevention advice is neglected.	Promotion of partner treatment and prevention of reinfection

Models of STI transmission dynamics



FIGURE 4-2

TABLE 4-1

Illustrative STI control program monitoring and evaluation indicators and sources of data

Monitoring and Evaluation Indicators		Sources of Data
Process Evaluation		Process Evaluation
INPUTS	Community outreach, targeted peer education <ul style="list-style-type: none"> Number of active peer educators by target group Number of trainings, refresher trainings, and supervisory visits 	Program records Program records
	Provision, use, and proficiency of curative services for STIs <ul style="list-style-type: none"> Number/proportion of health workers trained in STI case management Number of supervisory visits Number and distribution of clinics with qualified staff Adequate supplies of STI drugs and condoms 	Program records Program records Program records Clinic reports/facility surveys
OUTPUTS	Community outreach, targeted peer education <ul style="list-style-type: none"> Number of peer educator contacts Proportion of target population reached Number of peer educator referrals to clinical services Number of condoms distributed 	Routine reports Reports/census data/special studies Routine reports Routine reports
	Provision, use, and proficiency of curative services for STIs <ul style="list-style-type: none"> Number of persons seeking care for STI complaints Proportion of pregnant women in antenatal clinics screened for syphilis Proportion of population with knowledge of STI symptoms and STI transmission Partner referral rate Adequacy of national STI treatment guidelines with respect to diagnostic accuracy and treatment effectiveness Adequacy of laboratory testing (where laboratory support is part of national STI treatment protocols) Adequacy of supply of STI drugs [UNAIDS STI Service Indicator 3]¹ Quality of service provision: Proportion of diagnosed STI patients who are treated according to national STI treatment guidelines (composite indicator) [UNAIDS STI Service Indicator 1]¹ Proportion of STI patients given prevention advice (condom use, HIV testing) and advice on partner referral (composite indicator) [UNAIDS STI Service Indicator 2]¹ 	Clinic records Clinic/laboratory records Population based surveys Clinic records STI studies STI studies/quality assurance system reports Facility surveys Facility surveys Facility surveys
Effectiveness Evaluation		Effectiveness Evaluation
OUTCOMES (SHORT-TERM AND INTERMEDIATE EFFECTS) <ul style="list-style-type: none"> Decrease in time between onset of symptoms and seeking appropriate treatment Proportion of persons with STI symptoms seeking care for STI complaints [UNAIDS STI Service Indicator 4]¹ Decrease in specific STI prevalence and incidence 		Behavioral surveys Behavioral surveys Routine STI surveillance/STI studies
IMPACT (LONG-TERM EFFECTS) <ul style="list-style-type: none"> Decrease in STI related morbidity: congenital syphilis, infertility, ectopic pregnancies (STI Program goal) Decrease in HIV prevalence and incidence 		Routine STI surveillance/STI studies HIV studies

structural obstacles impeding safer sex and access to services, as well as acceptability of current services.

Information from the situation analysis should be synthesized and fit into the explanatory model that describes the dynamics of STI transmission in the community and identifies limitations of existing services. Potential biases in the data should be recognized, however. The most easily collected data are often obtained from formal public-sector health care facilities. Results may be subject to bias when services do not reach the people at greatest risk in the community. Formative research should take a wider community perspective to avoid missing population groups important to STI transmission. The meetings and interviews that comprise formative research activities should also be seen as important opportunities to promote trust and participation of key target populations.

Process Evaluation: Inputs and Outputs

Attention to service delivery through monitoring program inputs and outputs, including service quality and coverage, is crucial for program management. These indicators describe how efficiently program activities are carried out and many, although not all, are collected as part of implementing program activities. All important program activities should be monitored and should be linked to indicators. One of the most important purposes of monitoring process indicators is keeping services on track. Indicators should be reviewed with project staff on a regular basis to improve management, identify problems, and motivate performance.

Examples of process evaluation indicators and possible data sources for different types of STI program activities are presented in Table 4-1. Sources of data for input-level process indicators are generally obtained from routine monitoring records of service implementation as well as program records on training and supervisory

activities. Similarly, data for many output-level process indicators include simple routine reporting instruments such as clinic registers, log books, referral slips, and program records of activities.

At regular intervals, however, it is important to take a closer look at aspects of quality and coverage of services than is possible through the process monitoring described above. Periodic evaluations of these aspects of service delivery

can provide useful information for reorienting and strengthening services. These periodic output-level evaluation exercises require special surveys and studies to be undertaken to obtain the necessary data.

Quality of STI Services

Methods for evaluating the quality of STI services have been tested and refined, and indicators

form part of the recommended set of evaluation criteria for HIV prevention programs^{1,2}. STI service indicators 1 and 2 (previously called Prevention Indicators [PI] 6 and 7) look specifically at the curative and preventive aspects provided at health care facilities by STI case management providers. Key questions addressed in monitoring service quality include:

- Are services providing quality care for men and women infected with STIs?
- Are necessary drugs, diagnostics, and other commodities for treatment and prevention available?
- Are services used effectively as an entry point for HIV and STI prevention?

Protocols for measuring these indicators generally include enumerating health care facilities where STI case management is provided and evaluating the adequacy of health facility infrastructure for STI case management¹. Other areas of quality that are assessed include provider attitudes toward STI clients, partner notification, and referral for voluntary HIV counseling and testing.

Several methodologies for evaluating the quality of STI services from providers have been compared³. The approach recommended by UNAIDS is a health facility survey with both provider interview and provider observation components. While the standardized methodology permits comparison across time and place, disadvantages include expense (observers may have to wait several days for a patient with STI symptoms to present) and the need to select sites with sufficient volume to permit observation. Additionally, biases related to the method of data collection itself can occur (for example, a provider may alter performance during observation and the knowledge elicited in interviews may not be reflected in actual patient management).

Table 4-2 presents alternative strategies for assessing the quality of STI care, including use

of mystery patients, exit interviews, and record reviews. Adaptations of these methodologies have also been used at other sources of care, including private practitioners, pharmacies, and “non-qualified” practitioners. Mystery patients (for health care facilities) or mystery shoppers (for pharmacies) are observers who are trained to pose as patients with an STI complaint. While this approach reduces bias that occurs when someone is aware of being observed, the expense may limit the number of observations or types of STI syndromes evaluated.

Reliable data on the quality of STI case management is important for effective management of STI control programs. The challenges of accurately assessing this indicator through health facility surveys should not be underestimated, however. In many cases, the choice of method is determined by what the health care providers will allow, particularly in the case of the private sector. For examples, private providers may participate in a short interview but will not allow their records to be reviewed or allow observations of their patient management sessions.

Unless the physician or pharmacist community has consented, mystery client surveys may be viewed as intrusive and be ultimately counterproductive in standardizing high quality STI services in a community. It may be more practical to include a limited set of quality indicators into ongoing monitoring systems. For example, data routinely collected in patient registers, charts, or specially designed patient encounter forms can be analyzed to determine the proportion of patients managed according to protocol. A comprehensive approach would be to combine routine monitoring with periodic in-depth health facility surveys. Further discussion of these issues can be found in the review by Saidel and colleagues³. When this information is collected, data collection protocols and staff should strictly adhere to the principles of confidentiality and respect for persons.

TABLE 4-2

Commonly used methods for evaluating quality of STI case management

Method	Advantages	Disadvantages
Provider interviews	<ul style="list-style-type: none"> • commonly used method • relatively easy • permits evaluation of a range of knowledge • standardized protocols exist 	<ul style="list-style-type: none"> • reporting bias (knowledge does not necessarily reflect actual practice)
Direct observation of provider	<ul style="list-style-type: none"> • commonly used method • transparent • standardized protocols exist 	<ul style="list-style-type: none"> • logistically difficult and expensive • biases: <ul style="list-style-type: none"> – need to select health care facilities with sufficient volume to permit observation – providers’ awareness of being evaluated
Mystery patient/shopper	<ul style="list-style-type: none"> • may be less expensive than direct observation • less likelihood of observation bias 	<ul style="list-style-type: none"> • possible negative reactions of health care workers to being observed without their knowledge • mystery patient/shopper does not actually have clinical signs that may affect management
Record review	<ul style="list-style-type: none"> • avoids observation bias • less costly 	<ul style="list-style-type: none"> • information in records can be limited and often inconsistent
Patient exit interviews	<ul style="list-style-type: none"> • may be less expensive than direct observation • less likelihood of observation bias 	<ul style="list-style-type: none"> • Recall biases (patients may not accurately remember details of clinical procedures or counseling messages)

In addition to monitoring the quality of services at the clinic level based on national guidelines, the continued applicability and appropriateness of STI case management guidelines should also be monitored periodically. Specifically, the etiologies of the various STI syndromes and the continuing susceptibility of *Neisseria gonorrhoeae* (and *Hemophilus ducreyi* if applicable) to the recommended antibiotics should be monitored. National guidelines may need to be altered and field tested to incorporate new diagnostic tests that may become widely available or to adjust based on shifting syndrome etiologies or changing antibiotic susceptibility of STI pathogens. Data collection efforts to determine the ongoing appropriateness of national STI treatment guidelines require special studies with high-quality laboratory support.

Coverage

As important as high-quality preventive and curative services are, their contribution to STI control is limited by the extent to which they are used, particularly by those at highest risk of acquiring and transmitting STIs. Evaluating coverage is therefore as important as evaluating quality. Several approaches to evaluating coverage are possible and the choice of methods will depend on factors such as marginalization of target groups, extent and nature of the informal health care sector, and the structure of service delivery. Mapping and sampling methodologies are used to quantify the number of risk settings and estimate the size of hard-to-reach populations (see Chapter 8, “Uses of Behavioral Data for Program Evaluation,” Chapter 9, “Sampling Strategies for Monitoring HIV Risk Behaviors,” and Larson et al.⁵ for more information on these methodologies). Where special screening and treatment services for sex workers are part of the public sector health system, these can be evaluated using adapted versions of the clinical evaluation tools described above. Where sex work is more

covert and special outreach services have been developed, it may be more appropriate to evaluate those services using separate methods developed in the context of the project’s objectives.

Whatever the internal approach to evaluating services, it should be complemented by an external view of the intervention’s image and use. Community surveys should try to estimate the proportion of the target group that knows of and uses the services. Protocols and instruments designed for behavioral surveillance can include information about target groups’ knowledge of health services. Special surveys and qualitative research among people with recent STI symptoms would provide information on health services use and access issues. Quantitative surveys looking at coverage should always be supplemented by qualitative research to clarify issues around barriers to care (see Chapter 12, “The Role of Qualitative Data in Evaluating HIV Programs,” for a discussion of commonly used qualitative research methods).

Similarly, where identified bridge populations of men with occupational or other risks are included in interventions, evaluation should take into account both the quality of services offered as well as the extent to which those services are used. For example, in a city where migrant men are believed to be at risk, pharmacies may be a common source of care for treatment or prophylaxis. Evaluation of a program designed to improve such care should consider the quality of advice provided (through mystery shoppers or exit interviews), number and coverage of pharmacies, and service site knowledge and use through a survey of occupational risk groups supplemented by more in-depth qualitative research.

Effectiveness Evaluation: Outcomes and Impact

Evaluating the effectiveness of an STI control program involves estimating the impact of programs on the prevalence and incidence of STIs

and STI-related behaviors, including treatment seeking behaviors, condom use and sexual behaviors (such as number and type of sexual partners).

To guide STI control efforts, a program manager would ideally monitor in the general population and in specific target groups the prevalence and incidence of each STI pathogen and related STI morbidity, as well as the prevalence of risk and treatment-seeking behaviors. In reality, this ideal is difficult to attain for a number of reasons:

- data from health care facilities often do not reflect what occurs in the community;
- population-based data on STIs are difficult and expensive to obtain;
- most commonly available diagnostic methods for STIs are inaccurate, particularly for women;
- accurate laboratory data on STI pathogens are expensive to collect; and
- reporting systems, where operational, rely primarily on reports from public sector health care facilities, while most STIs are managed at private or informal sector facilities that rarely report.

Despite these obstacles, the outcome of STI control efforts on disease trends can often be estimated using a combination of passive data collection and special biologic and behavioral surveys, which together form a package known as STI surveillance. Table 4-3 summarizes some of the different components of STI surveillance. The definition and role of surveillance in STI control have been well established and refined over the years^{6,7,8}. Surveillance data in STI control programs are not only used to assess the effectiveness of programs but are used to determine the need for public health action. Additional critical program needs for which surveillance data are used include setting intervention priorities, planning and allocating resources, defining subgroups for targeted inter-

ventions, designing disease prevention programs, and conducting STI program process evaluation.

Passive Disease Surveillance

Most countries have some system in place for tracking numbers of STI diagnoses in public sector health care facilities, and reporting may include private practitioners as well. Despite major problems with completeness, reliability, and timeliness, passive STI surveillance has the advantage of being relatively inexpensive to implement. Common STI surveillance case reporting methods include:

- **Health information systems (HIS)**—Routine reporting of limited data, usually in the form of tally sheets, on a wide range of diseases seen at all public health care facilities.
- **Communicable disease reporting**—Mandatory reporting of certain diseases of public health importance (presumptive or laboratory confirmed) from clinical and/or laboratory sites.
- **Enhanced surveillance from sentinel sites**—More detailed data collection at a limited number of clinical and/or laboratory sites that are reinforced with training, regular supervision, and financial support.

The choice of passive surveillance methods or combined passive and enhanced surveillance methods will depend on a number of factors, including the organization of STI service provision and availability and reliability of laboratory diagnosis. Data may be inaccurate because of underreporting and poor quality control of laboratory services, or unreliable because of lack of consistent case definitions.

In practice, STI surveillance markers may be limited by feasibility considerations to a few common, curable STIs or easily identifiable syndromes. Table 4-4 presents several syndromic indicators commonly used in passive STI surveillance.

TABLE 4-3 The different types of STI surveillance

	Passive Surveillance—Routine Disease Notification	Enhanced Passive Surveillance—Sentinel Sites/Populations	Specific Surveys/Studies
Study Population	Entire population	Selected sites, such as: <ul style="list-style-type: none"> • reproductive health clinics (ANC, FP, Gyn) • blood banks • military • primary health care clinics • STI clinics • specific laboratories 	Various, including: <ul style="list-style-type: none"> • general population • specific target groups • convenience populations • clinic populations
Methods	<ul style="list-style-type: none"> • case notification 	<ul style="list-style-type: none"> • case notification • routine screening • laboratory reporting 	<ul style="list-style-type: none"> • STI prevalence studies • drug sensitivity studies • risk studies • behavioral studies • algorithm validation studies • syndrome etiology studies
Information obtained	<ul style="list-style-type: none"> • estimate of incidence in the entire population • data used in monitoring trends 	<ul style="list-style-type: none"> • estimate of incidence for sites • data used in monitoring trends in sites 	<ul style="list-style-type: none"> • STI prevalence at a specific time and in a specific population • rapid assessment • patterns of pathogens • drug sensitivity • health care-seeking/risk behavior
Advantages	<ul style="list-style-type: none"> • surveillance of entire population seeking care at services 	<ul style="list-style-type: none"> • manageable • flexible • specific studies (e.g., resistance patterns) can easily be added 	<ul style="list-style-type: none"> • flexible • possible in areas with extremely limited resources • studies in specific populations (e.g., high-risk groups) • facilitates interventions
Disadvantages	<ul style="list-style-type: none"> • unrepresentative due to underreporting, underdetection, misclassification • not flexible • hard to evaluate • legislative or regulatory requirements <p>As a result, trends are difficult to interpret</p>	<ul style="list-style-type: none"> • possible bias due to nonrepresentative sites, changes in structure, accessibility and acceptability of sites <p>Possible biases, such as changes in health care-seeking behavior, have to be taken into account when interpreting trends</p>	<ul style="list-style-type: none"> • random or repeatable samples are extremely difficult to obtain in the field • more costly <p>Representativeness, generalizability must be taken into account when interpreting trends</p>
Observations	<ul style="list-style-type: none"> • useful for rare diseases with a simple, clear, and unequivocal case definition • well-functioning passive surveillance systems rare 	<ul style="list-style-type: none"> • sentinel surveillance is an appropriate tool for the surveillance of STI when well-functioning passive surveillance systems do not exist • sentinel site data may be of higher quality and more cost-effective to obtain than passive surveillance data • sentinel sites with well-trained staff and good resources can serve as reference centers in the respective regions 	<ul style="list-style-type: none"> • useful for obtaining an estimate about frequency and distribution of STI s in a given population • complements behavioral and HIV surveillance activities • appropriate instrument to measure STI prevalence and monitor trends in high-risk/core groups • such studies can facilitate interventions • special studies are expensive

Source: Modified from Reference 4

In some situations, it may be useful to track etiologic markers of STI syndromes and include reporting from routine laboratory testing, such as antenatal syphilis screening. Cost and reliability of laboratory diagnosis are factors that may influence decisions about including laboratory surveillance.

Special Surveys and Studies

Where more reliable estimates of STI prevalence are needed and a supplement to the passive reporting system is desired, periodic population-based STI prevalence and behavioral surveys can be carried out on the general population or specific target groups. Although considerably more expensive than passive surveillance, such surveys can provide accurate estimates of disease prevalence in specific populations. Availability of highly accurate urine-based diagnostic tests for several common STIs makes population-based surveys feasible. Because of the high cost of such tests, surveys are most practical for monitoring trends among high-risk priority populations where initial prevalence is high (and sample sizes smaller). When combined with behavioral surveillance, STI prevalence surveys can provide valuable information about disease and risk behavior trends in priority populations, where most sexual transmission occurs.

These special STI surveillance activities, although complex and currently expensive, provide extremely valuable data for evaluating the effectiveness of STI and HIV prevention programs. The recommendations for Second Generation HIV Surveillance, which advise the collection of both biologic and behavioral data, include STI surveillance methods (syndromic reporting, syphilis seroprevalence, STI prevalence surveys among high-risk groups) as a complement to other HIV and behavioral surveillance activities². The role of behavioral surveillance is discussed in more detail in Chapters 8-13 of this Handbook. Other issues of importance for STI control as an HIV prevention intervention, such

as modeling the impact of STI interventions on HIV transmission and measuring the efficiency and cost-effectiveness of different STI control interventions, are discussed in Chapters 14-17.

Other surveillance activities that fall in the category of special studies include assessments of incidence and prevalence of STI-related complications (such as pelvic inflammatory disease [PID]

or cervical cancer), prevalence surveys of viral STIs (such as herpes simplex virus [HSV] type 2 or hepatitis), and estimation of the economic costs of STIs, to name a few. This information helps to refine and refocus mature STI control programs but are less essential when initial efforts should be directed at maximizing effective treatment opportunities for the bacterial STIs.

Controlled clinical trials, either community randomized or individual randomized, in assessing the impact of STI control on HIV incidence⁹ or treatment trials assessing the impact of chronic suppressive HSV therapy are difficult and expensive and not warranted in routine monitoring and evaluation of programs.

Prevalence rates of STI pathogens. As such, the type and the accuracy of diagnostic tests must also be taken into account when interpreting STI trends or comparing STI prevalence rates among different sites.

STIs as Biologic Proxies for HIV Transmission

The relationship between STIs and HIV is complex because they act both as confounders and effect modifiers (Figure 4-3). STIs facilitate HIV transmission through numerous biologic mechanisms that affect both infectiousness and susceptibility. HIV infection alters the natural history of some STIs. Moreover, the predominant mode of HIV transmission globally is through sexual contact; sexual behavior, partner type, and sexual network can affect both STI and HIV risk.

STI prevalence and incidence, both increasing and decreasing, are being promoted as a proxy measure for the effectiveness of HIV prevention programs because STIs, like HIV, are markers of unprotected sex. New cases of bacterial STIs, which are curable, ought to reflect high-risk sexual activity¹. An STI with high infectivity, short incubation period and period of infectiousness, and little resulting immunity would be a sensitive marker for unprotected sexual contact. Conversely, HIV prevention interventions should reduce STI prevalence and incidence rates more rapidly than the prevalence and incidence of HIV infection and, as such, bacterial STI measures would be a useful indicator of recent changes in sexual behavior for HIV¹.

This commonsense notion of STIs being a marker for the effectiveness of HIV prevention programs actually belies several complex interrelationships and processes¹¹. First, STI is a term that refers to a diverse group of about 30 pathogens—bacterial, protozoal and viral, each with unique biologic features of infectiousness, host response, and associated patterns of sexual contact. Second, STI transmission dynamics in a population are significantly influenced by subsets

OTHER CONSIDERATIONS

Laboratory Issues

The laboratory is an essential component of STI control programs and in evaluating STI control program interventions. The availability of quality laboratory services are severely limited in most settings due to constraints of cost (both initial and recurrent), expertise, inconsistencies in supplies, and erratic quality control measures and other support. Additionally, laboratory tests for specific STI pathogens vary widely in sensitivity and specificity to detect specific etiologic agents and each test has specific requirements in terms of specimen handling and processing¹⁰. As such, interpretation of laboratory-based STI trends must consider issues of test quality. For example, sudden declines in syphilis prevalence in antenatal clinic populations may be due to outdated reagents and lack of quality control procedures in place rather than an effective program. Similarly, a very high reported prevalence of chlamydial infection may be due to inexperienced technicians who over-interpret fluorescent slides rather than to a severe chlamydial infection problem.

The availability of urine-based DNA amplification testing has now made population-based STI surveys feasible and allows for surveys to be conducted at great distances from laboratory support. When conducting these surveys in the field, special attention needs to be paid to specimen handling to ensure quality laboratory data.

Finally, changing test methodologies can sometimes have a dramatic effect on measured preva-

TABLE 4-4

STI syndromes used as passive surveillance indicators

Syndrome	Advantages	Disadvantages
Genital ulcer disease (GUD) in men and women	<ul style="list-style-type: none"> marker of high transmission dynamics and/or poor STI control sensitive indicator to change when initially high 	<ul style="list-style-type: none"> GUD is less apparent in women, leading to potential underestimation of the problem in mature HIV epidemics, herpes simplex virus type 2 (HSV-2) appears as a major cause of GUD; HSV-2 reflects the immunosuppression in the population and not the proficiency of the STI control program
Urethral discharge (UD) in men	<ul style="list-style-type: none"> common STI syndrome in men clinical signs and symptoms are specific for STI 	<ul style="list-style-type: none"> men often seek treatment in informal sector, leading to underestimation of problem
Vaginal discharge in women	<ul style="list-style-type: none"> most common STI-related syndrome in women 	<ul style="list-style-type: none"> most vaginal discharge is not caused by STI-related pathogens poor sensitivity and specificity, leading to frequent diagnostic misclassification
Lower abdominal pain (LAP) or pelvic inflammatory disease (PID) in women	<ul style="list-style-type: none"> most common STI complication in women 	<ul style="list-style-type: none"> poor sensitivity and specificity, leading to frequent diagnostic misclassification
Congenital syndromes (congenital syphilis and ophthalmia neonatorum)	<ul style="list-style-type: none"> can be sensitive indicator of STI trends or of specific measures to reduce congenital transmission 	<ul style="list-style-type: none"> reliable case definitions and reporting systems may be difficult to establish and maintain

of the population, the so-called core and bridge populations. As such, changes in the incidence of bacterial STIs, such as gonorrhea, in a population might be due primarily to behavior changes, such as increased barrier method use, of these population subsets. It is conceivable that a decline in gonorrhea may occur in a population where HIV transmission would continue to take place, because people infected with HIV remain infectious for years. Third, effective treatment will have a dramatic impact on the incidence of treatable STIs. Consequently, one cannot attribute changes in STI incidence solely to changes in behavior. Finally, STI changes must be put in the context of longer-term trends, where such data are available.

While STI prevalence and incidence rates are important to monitor for STI and HIV control program purposes, the STI data need to be triangulated with behavioral information to fully understand the effect of HIV prevention efforts.

Renton and Whitaker distinguish two types of STI indicators, indirect and direct, for HIV control programs¹¹. An indirect indicator would be defined as an STI whose occurrence is used as a proxy measure of high-risk sexual activity. They propose that the ideal characteristics of such an STI would include high infectivity, short period of infectiousness, short incubation period, and absence of conferred host immunity. *Neisseria gonorrhoeae* might be an appropriate STI for an indirect indicator. On the other hand, a direct indicator would be an STI whose occurrence is used as a proxy measure of HIV infection itself. Characteristics of such an STI would include life-long infection and life-long infectiousness. Practically, a direct indicator of HIV might be used to assess the potential for HIV spread in populations or as a tool to monitor HIV programs without measuring HIV infection directly. Such a direct indicator has yet to be defined.

As STI, behavioral, and HIV data are collected on populations an understanding of their

interrelationship will increase. This increased understanding may help to improve surveillance systems and make interpretation of existing surveillance data stronger.

CONCLUSION

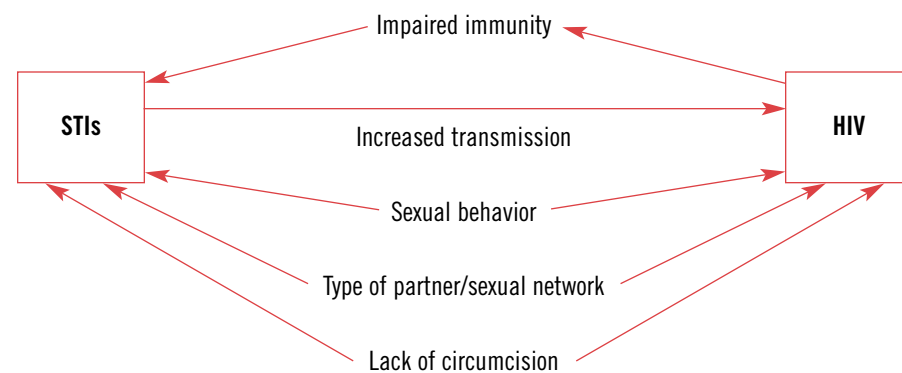
STI control programs must adapt strategies that respond to the heterogeneity of STI pathogens and transmission dynamics within their populations as well as the limitations of finance, infrastructure, and personnel. As such, STI control programs vary widely among countries and among sub-populations and, consequently, monitoring and evaluation of such programs will vary. There are, however, essential elements of monitoring and evaluation regardless of the program type:

The STI laboratory is an important component of STI control programs and in the evaluation of STI control program activities. The availability of laboratory services is severely limited in most settings due to constraints of money, technical expertise, commodity supply and quality control. As such, the cost of high-quality testing must be taken into account when planning laboratory-based evaluations, and the type and sensitivity of diagnostic tests must be considered when interpreting STI trends or comparing STI prevalence rates among different sites or populations. STI trends as biologic markers for HIV prevention programs are best interpreted with behavioral information in the population because treatment and specific interventions in sub-populations can affect STI prevalence in the general population.

- **Formative evaluation**—As a first step in the program cycle, information should be collected to describe the current situation, identify need, and design an appropriate response.
- **Process evaluation**—These data monitor how efficiently program activities are carried out and help keep services on track. Much of the data for process evaluation can be collected routinely as part of service delivery. Monitoring the quality and coverage of STI services is essential and requires special studies and surveys to be undertaken.
- **Effectiveness evaluation**—These data estimate the impact of the STI control program on the prevalence and incidence of STIs and STI-related behavior both in the general population and in specific target populations. Despite obstacles, outcomes of STI control efforts on disease trends can be estimated using a combination of passive data collection and special biologic and behavioral surveys—an STI surveillance system.

FIGURE 4-3

Interactions between STIs and HIV



REFERENCES

- 1 UNAIDS/00.17E. National AIDS programmes: a guide to monitoring and evaluation. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2000.
- 2 WHO/CDS/CSR/EDC/2000.5 and UNAIDS/00.03E. Guidelines for second generation HIV surveillance. Geneva: World Health Organization and Joint United Nations Programme on HIV/AIDS; 2000.
- 3 Saidel T, Vuylsteke B, Steen R, et al. Indicators and the measurement of STD case management in developing countries. *AIDS* 1998;12(Suppl 2):S57-S65.
- 4 Schwartländer B, Van Den Hoek A, Heyman D, Gerbase A. Surveillance. In: Dallabetta G, Laga M, Lamptey P, editors. *Control of sexually transmitted diseases*. Arlington (VA): Family Health International; 1998.
- 5 Larson A, Stevens A, Wardlaw G. Indirect estimates of "hidden" populations: capture-recapture methods to estimate the numbers of heroin users in the Australian capital territory. *Soc Sci Med* 1994;39(6):823-831.
- 6 WHO/CHS/HIS/99.2. Guidelines for sexually transmitted infections surveillance. Geneva: Joint United Nations Programme on HIV/AIDS and World Health Organization; 1999.
- 7 Ward JW, Greenspan JR. Public health surveillance for HIV/AIDS and other STDs: guideposts for prevention and care. In: Holmes K, Marling P, Sparling P, et al., editors. *Sexually transmitted diseases*. 3rd ed. New York: McGraw-Hill Inc.; 1999. p. 887-892.
- 8 Catchpole MA. The role of epidemiology and surveillance systems in the control of sexually transmitted diseases. *Genitourin Med* 1996;72(5):321-329.
- 9 Grosskurth H, Gray R, Hayes R, et al. Control of sexually transmitted diseases for HIV-1 prevention: understanding the implications of the Mwanza and Rakai trials. *Lancet* 2000;355(9219):1981-1987.
- 10 Van Dyck E, Meheus AZ, Piot P. Laboratory diagnosis of sexually transmitted diseases. Geneva: World Health Organization; 1999.
- 11 Renton AM, Whitaker L. Using the classical STDs to monitor HIV prevention programmes. In: Paccaud F, Fader JP, Gutzwiller F, editors. *Assessing AIDS prevention*. Basel (Switzerland): Birkhauser Verlag; 1992. p. 32-51.

5

EVALUATING CONDOM PROGRAMMING

SUSAN SMITH AND JOHN STOVER

The Use of Indicators in Measuring Condom Program Effectiveness	68
Evaluating Condom Use, Knowledge, and Attitudes Toward Condoms, and User Characteristics	68
Baseline and Tracking Surveys	70
Panel Studies	71
Omnibus or Continuing Surveys	71
Consumer Intercept Surveys	72
Evaluating Condom Availability, Condom Quality, and Quality of Customer Services	73
Outlet Checks, Retail Audits, and Mapping Studies	73
Condom Quality Testing	74
Mystery Shopper Studies	74
Analytic Techniques	75
Cost-effectiveness Analysis	75
Time Series Analysis	76
Discriminant Analysis	76
Conclusion	77
References	78

5

EVALUATING CONDOM PROGRAMMING

A key ingredient in preventing HIV/AIDS is condom programming. The three principal ways in which condoms are distributed are through government programs, subsidized social marketing programs, and the commercial sector. Government programs

typically involve donated products that are distributed through the established government health infrastructure. Under this approach, condoms are distributed free of charge through government hospitals, clinics, and rural health posts. The reach of public-sector programs is frequently extended through the use of outreach workers or distribution to key target groups, such as the military or schools.

Most countries with severe HIV/AIDS problems have also developed social marketing programs to provide supplemental means of condom distribution. These programs typically use subsidized products that are distributed primarily through traditional commercial outlets, such as pharmacies and small shops. Non-traditional commercial vendors, however, are increasingly being used to provide access to harder-to-reach groups and at venues closer to where sexual encounters are likely to take place. Less traditional outlets include bars, hotels, vending machines,

taxi drivers, truck stops, gas stations, and kiosks.

Social marketing programs use the tools and techniques of commercial marketing to promote products having a social benefit, in this case condoms for AIDS prevention. The framework most commonly used to think about the various program components is the four Ps marketing-mix framework. The four Ps comprising this framework are **products**, **place** (distribution), **promotion**, and **price**. Promotion, a major component of most condom social marketing programs, includes mass media advertising, public relations activities, trade promotions, and provider training. Because the programs are subsidized, social marketing prices fall in between the free government products and full commercial prices. Prices are typically set to be appropriate for middle to lower-middle socioeconomic (or C and D class) target markets. Finally, market research is typically used to design, monitor, and evaluate the various components of the program¹.

A third, less-developed approach for condom programming involves encouraging the commercial sector to become more actively engaged in promoting condom availability.

Commercial distributors do not make much profit on condom sales, so innovative ways, such as co-marketing, offering customer service training, or public relations events and contests, are often used to gain their participation.

Condom use has been dramatically increasing throughout the developing world. In Kenya, for example, condom distribution has increased from less than 10 million, in the mid-1980s, to more than 50 million, by the mid-1990s. Today, worldwide condom use could be as high as 600 million per year. To better understand the impact of condom programming on health outcomes and on preventing HIV/AIDS, it is important to evaluate how condoms are being used. Toward that aim, the following key questions are the most important to consider:

- How many condoms are being distributed?
- Who is using condoms?
- With which partners are people using condoms?
- With what consistency are people using condoms?
- What are the barriers to condom use?
- How effectively are condoms being used?
- How cost-effective are different condom distribution programs?

This chapter examines various methodologies and techniques that can be used to answer these questions.

THE USE OF INDICATORS IN MEASURING CONDOM PROGRAM EFFECTIVENESS

In the early 1990s, WHO/GPA developed a set

of Priority Prevention Indicators (PIs) to assist National AIDS Control Programmes evaluate their country programs. A few of these indicators dealt specifically with condom use and availability. WHO/GPA field-tested these indicators and prepared procedure manuals to guide country implementation. However, these indicators were not widely adopted for routine use by country programs.

Subsequently, under USAID's EVALUATION Project, another set of indicators for HIV and sexually transmitted infections (STIs) were developed and documented as a part of a handbook on reproductive health indicators². Recently, two new efforts prepared handbooks of indicators for HIV/AIDS/STI programs. One effort was led by USAID's MEASURE/Evaluation Project, the other by UNAIDS. Both used a series of working groups to develop, refine, document, and field test the indicators.

Table 5-1 shows the recommended UNAIDS indicators that are specific to condom programming for AIDS at the national level³. As in the previous cases, the working groups have developed protocols for assessing these indicators.

Other programs have also used condom service statistics or sales, consistency of condom use, use effectiveness, attitudes and misconceptions regarding condoms, and matching of target markets with actual user profiles as additional indicators of program performance.

EVALUATING CONDOM USE, KNOWLEDGE, AND ATTITUDES TOWARD CONDOMS, AND USER CHARACTERISTICS

According to the UNAIDS manual on monitoring and evaluating national HIV/AIDS prevention programs, "indicators of sexual behavior and condom use are probably the most important of all indicators in monitoring HIV prevention

UNAIDS condom programming indicators

TABLE 5-1

Program Area	Indicator	Definition
Condom use	Condom use at last higher-risk sex	The percent of respondents who say they used a condom the last time they had sex with a non-marital, non-cohabiting partner, of those who have had sex with such a partner in the last 12 months.
	Condom use at last commercial sex, client report	Proportion of men reporting condom use the last time they had sex with a sex worker, of those who report having had sex with a sex worker in the last 12 months.
	Condom use at last commercial sex, sex worker report	Percent of sex workers who report using a condom with their most recent client, of sex workers surveyed having sex with any clients in the last 12 months.
	Condom use at last anal sex between men.	Percent of men or their partners who used a condom at last anal sex with a male partner, of those who have had anal sex with a male partner in the last 6 months.
Condom use-youth	Young people using a condom during premarital sex.	Percent of young single people (15-24) who used a condom at last sex, of all young, single, sexually active people surveyed.
	Young people using a condom at last higher-risk sex	Percent of young people (aged 15-24) who have had sex in the last 12 months and used a condom at last sex with a non-marital, non-cohabiting partner, of all young people surveyed.
	Condom use at first sex	The proportion of young people (aged 15-24) who used a condom the first time they ever had sex, of those who have ever had sex.
Condom availability	Condoms available for distribution nationwide	Total number of condoms available for distribution nationwide during the preceding 12 months, divided by the total population aged 15-49.
	Retail outlets and services with condoms in stock	The proportion of randomly selected retail outlets and service delivery points that have condoms in stock at the time of survey, of all retail outlets and service delivery points selected for the survey.
Condom quality	Condoms that meet quality control measures	The percentage of condoms in central stock and in retail outlets that meet WHO quality specifications.

programmes and evaluating their success.” These indicators are best assessed through surveys. The following sections discuss various types of surveys and what information is best obtained from each.

Baseline and Tracking Surveys

Baseline and tracking surveys can be used to measure changes in condom use; use by type of partner; consistency of use; knowledge of correct use; attitudes toward condoms; barriers to condom use; and exposure to various program intervention, such as advertising and promotions. These surveys can be conducted among the general population or among target groups.

Baseline surveys are conducted before any program intervention commences. Their objective is two-fold: to collect background information useful in designing the program (formative research) and to collect baseline measures of key performance indicators, which will be used in subsequent evaluations.

Sampling and questionnaire design are the major research design issues. With these types of surveys, each wave involves drawing independent samples in each survey round (with the same set of parameters and from the same or an updated sample frame). This can be a costly and time-consuming exercise. Consequently, many surveys are narrowed in scope to include only those target markets that are of primary interest to the program. For example, surveys might be conducted among males in three urban areas as opposed to designing surveys that are nationally representative of all sexually active males. Furthermore, in order to make best use of HIV/AIDS surveillance data, attempts are frequently made to sample respondents around HIV sentinel sites.

Thought also needs to be given to the frequency of subsequent survey rounds. Experience has shown that, in the absence of a major and radically new program effort, condom use is unlikely to change significantly among the general adult population in a single year. Exceptions

might be found among smaller, well-targeted groups, such as youth or certain high-risk communities. The frequency with which surveys are repeated should be determined by the amount of time in which significant changes are expected.

In the area of questionnaire design, significant progress has been made in the last ten years. There is now ample evidence that people will answer detailed questions about sexual behavior in questionnaires and that the data collected are corroborated by trends identified from other information sources, such as condom sales and STI prevalence. Several known biases exist, however. For example, there is evidence that women tend to underreport sexual activity outside of marriage in standardized surveys. Responses by males and females are often inconsistent. It can be important to corroborate survey information with other sources, if possible. Qualitative methods are often better at uncovering the details of condom use behavior. This information, however, can only be used to help interpret survey results and cannot be used to monitor indicators.

Two condom use issues have been particularly difficult to measure in structured questionnaires. The first issue involves trying to categorize the degree of risk someone is exposed to, by the type of sexual partners. Two approaches are being used. One is a time-based approach, that is, “regular” versus “non-regular” partners as defined by a relationship that has existed more than a year. A second approach classifies partners on the basis of cohabitation, namely living together or not living together.

The second issue involves measuring consistency of condom use. This is important because many condom promotional campaigns aim at getting more consistent condom use among non-regular or casual partners. In general population surveys, the tendency has been to ask for use (or non-use) at last sexual contact. It is felt that this avoids the bias associated with recall and self-reporting issues. For sub-population

surveys, however, the trend has been to directly ask about consistent use, usually using some type of scaled answer such as always, sometimes, or never. Other lines of questioning, such as attitudinal issues and advertising recall, have been more successfully developed and are now usual components of market research tracking surveys.

The two key survey implementation issues include interviewer training and questionnaire pretesting. Both these elements can greatly improve the quality of results. A more detailed discussion of behavioral data collection, sampling issues, questionnaires, data validity, and qualitative evaluation approaches is presented in Chapter 8, “Uses of Behavioral Data for Program Evaluation.”

Panel Studies

Panel studies are designed to examine changes over time for a given sample or panel of users or outlets. These studies consist of a series of interviews or observations of the same respondents or outlets over an extended period of time. For example, a panel of individuals could be asked to record their condom usage over the course of some time period. The basic characteristics of panel research are that it: (1) retains the same (or nearly the same) sample throughout the life of the study, and (2) measures elements at periodic intervals.

In contrast, cross-sectional surveys offer only a “snapshot” view of users. When replicated with successive samples, cross-sectional surveys may reveal trends at the level of population aggregates; but only panel studies (amenable to time-series analyses) are able to trace change in the knowledge, attitudes, and behaviors of individual users and may determine causality.

In principle, selecting a panel is no different from selecting a sample for a cross-sectional survey. Both are designed to represent a universe of users, the general population, or of outlets. The first step involves recruiting respondents into the panel and obtaining reasonable assurance

that they will cooperate throughout the duration of the study. Most panel studies offer some type of incentive to help ensure continued participation. Examples of incentives include small gifts, cash, and special lotteries for panel members. Even the best-run panels, however, will suffer some respondent attrition. Some members will undoubtedly move, be temporarily away, or otherwise be inaccessible or difficult to find. This is a problem unique to panel designs and special efforts need to be made to minimize these losses.

Bias resulting from attrition can be taken into account when interpreting panel data by comparing the characteristics of respondents lost with the characteristics of those who remain. For example, if losses from mobility occur more frequently among younger people than among older ones, findings can be weighted accordingly.

Although panel studies are complex and often costly, they offer several advantages. Obtaining repeated information from respondents over time increases accuracy and permits a more precise determination of behavioral processes, such as use habits, which are derived from patterns over extended periods of time. Most important, panels make it possible to analyze shifts and changes and to link them to the prior events that have brought them about. At the same time, panels have their problems, such as the conditioning of members through repeated contact and loss of members over time. Yet the panel remains one of the more powerful tools of causal analysis.

Omnibus or Continuing Surveys

Omnibus or continuing surveys are another source of data for measuring changes in users over time. An omnibus survey is a survey that is already being conducted by an established market research firm, such as Gallup or Research International. Such surveys are likely to exist in the more commercially developed countries. These surveys usually contain nationally representative samples of households and are conducted at

regular intervals, such as monthly, bi-monthly, or quarterly. Each survey contains a core set of demographic and socioeconomic questions used for categorizing respondents. Commercial clients then pay to have a set of market-specific questions added (for example, questions pertaining to product or brand use, image, attitudes; reactions to pricing and availability; advertising recall). Clients only pay for the survey round in which they participate. They receive responses to the core questions as well as their market-specific questions. This format is popular among consumer products companies, such as Lever Brothers and Colgate Palmolive. Questions on condom knowledge and attitudes have been added to many omnibus surveys, although explicit questions on condom use might be considered less appropriate for this format. The main difference between panel and omnibus studies is that panel studies conduct repeated interview waves with the same set of respondents, whereas omnibus surveys involve repeated interview waves with different individuals, though they are drawn from the same sample.

The major advantage of participating in an omnibus survey is that it is usually less expensive than conducting a separate survey, since the costs are shared among several clients. Thus, it often provides an inexpensive way to collect survey data for several points in time in order to track changes in knowledge, attitudes, and practices. The major limitation of omnibus surveys is that the sample cannot be specified. If a country desires to sample only urban men, for example, this sub-sample will have to be extracted from the full omnibus sample. This may result in small sample sizes and correspondingly lower measurement precision.

Results from omnibus surveys conducted in Mexico and Jamaica show that knowledge of condoms as a method of AIDS prevention increased after the launch of AIDS media campaigns. Figure 5-1 shows the increase in knowledge of condoms as a prevention measure between waves 2 and 3 of the Mexican survey, which corresponds to both the launch and accompanying controversy surrounding that country's national AIDS media campaign⁴.

Consumer Intercept Surveys

A consumer intercept can take place in several ways. One way is to have an interviewer briefly screen clients/customers as they leave an outlet. If they qualify, a request for an interview can be made on the spot. Another approach is to have the staff or salesperson screen and indicate appropriate persons to interviewers. Alternatively, the staff person, after appropriate screening, can ask those who qualify if they would be willing to be contacted in their home for a subsequent interview. If prevalence were very low, it would be less time consuming and more cost-effective to record the names and addresses of qualifying respondents and arrange to interview them later in their homes than to intercept them for an immediate interview.

If seeking the cooperation of a commercial pharmacy or shopkeeper, it is usually necessary to

provide the proprietor with some incentive. Often, a small cash payment is provided for each interview. If interviews are to be conducted on the spot, sometimes incentives, such as a free pack of condoms, are also offered to respondents. An alternative approach is to provide the customer with a certificate that can be redeemed at the participating outlet. This single mechanism provides a benefit to both the proprietor and customer.

There are sampling issues involved in how to most appropriately select the sites where respondents are being intercepted and screened. If a sizeable number of outlets are to be selected, the preferred method is to choose outlets randomly. If only a modest number of sites are to be covered, the selection is often done purposely to ensure sufficient variability with regard to neighborhood characteristics, volume of customers, or other criteria.

Results can be used to profile users. The profiles are used to determine how well the program is reaching its target market. Results can also be used to better understand key attitudes and behaviors of condom users. It is conceivable that the intercept approach could also be used with clients of sex workers (SWs), although some type of special arrangement would have to be worked out with the SWs themselves.

EVALUATING CONDOM AVAILABILITY, CONDOM QUALITY, AND QUALITY OF CUSTOMER SERVICE

Condom availability is a prerequisite of condom use. Therefore, it is important to measure various elements of condom availability. The issue becomes more complex when availability is broken down into "available by whom" and "at what times," and when availability takes into account additional social barriers as well as issues of eligibility and affordability. The next sections outline standard procedures for measuring condom availability as well as the quality of the condoms themselves and the associated quality of service delivery.

Outlet Checks, Retail Audits, and Mapping Studies

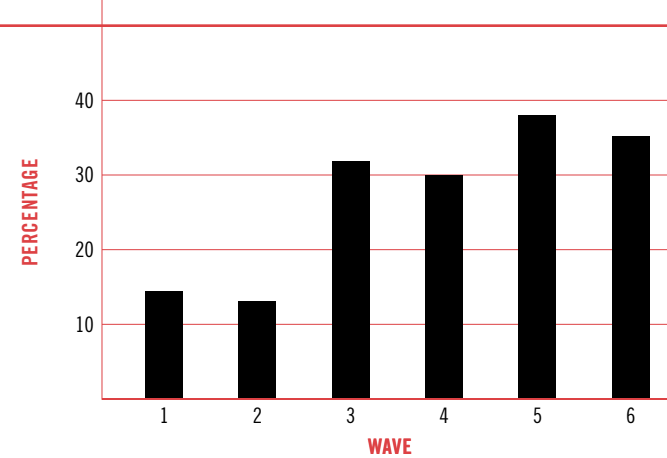
Condom availability at the national level is best determined by interviews with key informants, such as government procurement officers, commercial importers/distributors, other donor agencies, and non-governmental organizations (NGOs). Logistics management information systems can be useful in measuring condom flows to various distribution or wholesale points within the distribution system. They are generally inadequate, however, to measure condom availability or sales at the retail level.

Availability at the retail level is usually measured through some type of outlet check or audit. Outlet checks generally are used to determine the number of outlets of each type that have condoms of different brands. Retail audits provide information on the number of condoms sold in each outlet over time. Sampling is typically handled in one of two ways. With the more traditional outlets or with outlets that are relatively few in number, the first wave is frequently a census of all outlets, for example, pharmacies or government clinics in a certain region. This can be compared against a list of "official" outlets. Official lists may miss outlets that are not licensed or those that are closed, but may also indicate outlets that have been missed by the census. Samples are then pulled from this census for subsequent tracking waves. At the end of the project, a census often repeated to see if the total number of outlets has changed. Data from censuses give an indication of the percent penetration of condom distribution, by type of outlet.

For less traditional outlets or outlets that are too numerous to be covered on a census basis, sites are selected on based on the advice of those who know the local situation well. Data on availability are then reported as the number of outlets where condoms were found, divided by the number of sites visited. Regardless of the sampling approach, it is obviously important to identify and sample (in some manner) all the

FIGURE 5-1

Awareness of condoms as a means of AIDS prevention in Mexico



different types of outlets where condoms are likely to be found, including places such as bars, hotels, gas stations, and market stalls.

The biggest expense associated with an outlet check is getting interviewers to the outlets. Once they are there, they can relatively easily check other items of interest to the program as well, such as the availability of literature, condom prices, and waiting time for purchases. Once in the field, interviewers can also readily pick up other pertinent information, such as black market sales, leakage of public-sector goods into the commercial sector, and illicit cross-border trading of condoms.

Retail audits differ from outlet checks in that the protocol includes measuring sales. This is done by visiting the same set of outlets at regular intervals and collecting information on beginning and ending period inventories as well as intervening deliveries. Sales for the period can then be calculated by adding deliveries to beginning period inventory and subtracting ending period inventory. Some type of retailer training and incentive is usually offered to ensure successful participation in the audit. The principal problem encountered with projecting sales from this approach is that it is difficult to select a representative sample. Sales by outlet vary greatly and this information is not readily available by some proxy measure beforehand. Another approach is to purchase retail-audit data from an established commercial market research firm. For example, IMS HEALTH, a company that collects and sells pharmacy sales data in many countries, has spent considerable time working out sampling schemes for various countries. Condoms, however, are not covered in all IMS country surveys, and commercial data vendors do not typically include sales from less-traditional outlets.

Finally, some countries measure condom availability at the retail level through mapping studies, which plot outlet census data on maps, either manually or electronically using some type of geographic information system (GIS)

software. This provides spatial information and patterns pertaining to availability and accessibility. It is sometimes used in conjunction with national health surveys.

Condom Quality Testing

Condom quality can be measured at different points in the distribution system. The most important checkpoint, however, is the point closest to use—in other words, at the retail level. An outlet check or retail audit provides a good vehicle for collecting a sample of condoms for quality checking. WHO and USAID have standard protocols for checking condom quality^{5,6}. This requires a minimum of equipment and trained staff.

Mystery Shopper Studies

Mystery shopper studies involve the use of researchers posing as shoppers to observe performance at service delivery points. The researchers are disguised observers in the sense that they do not identify themselves as researchers, but instead appear to be normal customers or clients. These types of studies are used to monitor or assess the quality of staff-customer interaction.

Mystery shoppers are partially scripted and act out the role of a typical condom customer/client. Immediately after the “shopping experience” they record their observations and information gathered on a questionnaire. Many questionnaire items are factual in nature, such as, “How many condoms did they give you?”

Others, however, involve subjective judgment and are recorded on some type of scaled basis. For example, the question, “Did the pharmacist maintain a neutral, non-judgmental manner with you?” could be scored on a five-point scale from very judgmental (1) to not at all judgmental (5). This raises the issue of inter-rater reliability, or the variability due to different researchers’ subjective judgment of the factor being assessed. Good interviewer training can go a long way to reducing this problem. The training should

include both a thorough understanding of the factors being assessed and actual practice scoring these behaviors in the field. Sometimes two observers are used in mystery shopper studies. One researcher does the role playing and the other just observes. Both researchers record their experience independently. Another way to increase the accuracy of mystery shopper scoring is to conduct two baseline waves, by different researchers/shoppers, followed by a series of tracking waves. Sometimes different roles or scenarios are acted out in the tracking waves. For example, a young person might have a different experience obtaining condoms than would an adult. The greater number of measurement points should increase the accuracy of the assessment. Care must be taken, however, to ensure that the staff do not become familiar with the researchers. This can be managed either by using different researchers on tracking waves or by leaving a long enough time interval between visits.

In social marketing programs, outlets with continuing good customer service are often given some type of incentive or award. Mystery shoppers can be used to monitor and evaluate this performance. In addition, results, particularly improvements desired, can be incorporated into follow-up training programs.

ANALYTIC TECHNIQUES

Surveys are commonly analyzed using descriptive statistics, cross-tabulations, t-tests, and multiple regression. The strengths and limitations of each are well understood by researchers and will not be further discussed here. There are, however, three other analytic techniques that have not been as widely applied to this kind of research but probably should be used more frequently. These are cost-effectiveness analysis, time series analysis, and discriminant analysis. These are more sophisticated regression techniques that

involve different ways of statistically controlling for outside variables in order to isolate the effect of a particular program intervention. They are very briefly summarized here.

Cost-effectiveness Analysis

The relative cost-effectiveness of different condom distribution or delivery systems can be evaluated using this type of analysis. Because AIDS prevention funds are limited, it is important to periodically examine the cost-effectiveness of different condom programming alternatives.

As mentioned in the introduction to this chapter, condom distribution systems fall into one of three categories—government programs, social marketing programs, or commercial distribution. Public and commercial distribution systems are more conventional and, therefore, are somewhat easier to understand. Social marketing programs, on the other hand, can be organized in a variety of ways. There are two principal ways—organizations specially created to manage social marketing programs or existing commercial distribution companies who manage social marketing programs. In some countries, governmental or quasi-governmental organizations administer the program.

It is often assumed that social marketing programs offer a more cost-effective approach to delivering products because they have built-in cost recovery and harness existing commercial outlets. Within social marketing programs, it is further sometimes assumed that systems that piggyback on established commercial distribution systems, as opposed to those that create their own distribution fleet, are more cost effective, though they are likely to reach fewer people. In an era of resource constraints, it is important to understand the relative cost-effectiveness of different programming approaches. For this reason, evaluations of condom programs should include analysis of cost-effectiveness.

The cost-effectiveness methodology involves looking at program costs in relation to program output (typically the number of condoms distributed) for alternative approaches. Total costs can be determined by adding direct costs and applying a proportion of indirect costs. Net costs are found by subtracting revenue from total costs. Alternatively, program costs can be obtained by tracing all direct costs and shadow pricing (estimating the value of) goods and services donated to the program. The number of condoms distributed to the retailer is typically the output measure used. These data are routinely collected as a part of logistics management information systems^{7,8}.

Cost effectiveness can also be examined as costs per the number of infections averted from condom programs and compared against other condom promotion strategies and other intervention strategies, such as blood screening or STI treatment⁹. Simulation models can be used to estimate the number of condoms required to

avert one new HIV infection. Available simulation models range from easy-to-use models such as AVERT to large-scale simulation models such as iwgAIDS and STDSIM^{10,11,12}.

Results from the iwgAIDS model, with data from East African countries, have been used to rank the relative cost-effectiveness of various targeted intervention strategies. Table 5-2 shows that interventions that target adolescents and high-risk groups are the most cost-effective¹¹. Similar analyses might also be undertaken comparing social marketing with other types of interventions. For a more detailed discussion on the topics of impact modeling and cost-effectiveness analysis, see Chapter 17, “Guidelines for Performing Cost-effectiveness Analysis of HIV/AIDS Prevention and Care Programs.”

Time Series Analysis

Time series analysis is a way to examine changes in a trend after a particular program intervention has taken place. For example, the number of condoms distributed to specialty outlets, such as bars, could be examined over time to see if the timing of any change in sales from bars corresponds to the effort to expand distribution to less-traditional outlets.

To do this analysis, a long time series of data are needed both before and after the intervention. The technique essentially involves fitting a trend curve to the data before the intervention and fitting a second to the data that corresponds to the period after the intervention. If there is a kink in the curve (if the slopes are different) in the “before” and “after” cases, it suggests (though does not prove) that the intervention program has had an effect on sales.

Discriminant Analysis

Discriminant analysis is used to identify the characteristics that best distinguish or differentiate groups from each other. For example, this technique could be used to identify the distin-

guishing characteristics of consistent versus inconsistent condom users. It might be revealed that inconsistent users are younger, poorer, and less well educated. Interventions could then be designed and targeted to this specific group to improve consistency of condom use. Discriminant analysis is similar to multiple regression but is appropriate when the dependent variable is dichotomous. For example, if we want to know whether users of condoms from a social marketing program are different from users of other condoms, the dependent variable would be “Uses the social marketing brand? (Yes/No)” Like multiple regression, we would include in the analysis a number of other variables that might explain whether a person uses the social marketing brand or not. These might include income, urban residence, knowledge of many brands, or access to a public supply source. The discriminant analysis would provide an equation showing how each of these explanatory variables relates to the probability that a person uses the social marketing brand. Discriminant analysis can also be used when there are more than two groups. For example, it could be used to determine if there are significant social and economic differences among those who use social marketing, public, and other commercial sources. One of the outputs of a discriminant analysis is a classification table that shows how the equation classifies all the people in the sample. Some would be correctly classified as using the social marketing brand or using some other brand. Others would be incorrectly classified by the discriminant equation. Some would be incorrectly classified as users of the social marketing brand when, in fact, they do not use it. In some cases, these misclassifications may indicate the characteristics of people who are likely to adopt the social marketing brand in the future or who could be convinced to use it most easily.

CONCLUSION

The principal components of condom programming include condom distribution, condom promotion, and condom use. Indicators have been developed by different groups to measure various elements of availability and use. Availability can be measured through qualitative information interviews, management information systems (MIS) data, services statistics or sales, purchased commercial data, or some type of retail-level outlet check. The outlet check or retail audit typically provides the most reliable measure of condom availability. To thoroughly understand the issue of availability, however, additional issues, such as availability by whom and at what times, must be examined. This requires investigating relevant sub-groups as well as exploring questions of social barriers, accessibility, and cost. A combination of audits and surveys (or qualitative interviews) can be used to do this.

Elements of condom use are best measured through some type of survey, ideally repeated at different points in time. Survey protocols and questionnaire designs have been improving. In addition, good interviewer training and pretesting can help further improve and customize surveys. Sometimes insights gained from qualitative research can be helpful in interpreting quantitative findings.

Monitoring and evaluating condom programming can provide valuable guidance to HIV/AIDS prevention programs, in terms of making mid-course corrections as well in assessing overall program performance and in capturing lessons learned. The reliability of these assessments can be enhanced by cross-checking key findings with other available data sources.

TABLE 5-2

Relative cost-effectiveness of intervention strategies in a typical East African city

Intervention	Relative Cost-Effectiveness
Adolescents	5.6
High-risk populations	4.8
Family planning clients plus men	4.2
Family planning clients plus high-risk women	3.2
Family planning clients	3.0
Urban population	1.4
National population	1.0

REFERENCES

- 1 Bertrand JT, Stover J, Porter R. Methodologies for evaluating the impact of contraceptive social marketing programs. *Eval Rev* 1989;13:323-354.
- 2 Dallabetta G, Hassig S, editors. Indicators for reproductive health program evaluation: final report of the subcommittee on STD/HIV. Family Health International/AIDSCAP, with Carolina Population Center, Tulane University and The Futures Group International; 1995.
- 3 UNAIDS/00.17E. National AIDS programmes: a guide to monitoring and evaluation. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2000.
- 4 Stover J, Bravo M. The impact of AIDS on knowledge and attitudes about condoms as a contraceptive method in urban Mexico. *Int Fam Plann Perspect* 1991;17:61-73.
- 5 WHO/RHT/FPP/98.15 and UNAIDS/98.12. The male latex condom: specifications and guidelines for condom procurement. Geneva: World Health Organization and Joint United Nations Programme on HIV/AIDS; 1998.
- 6 McNeill ET, Gilmore CE, Finger WR, editors. The latex condom: recent advances, future directions. Research Triangle Park (NC): Family Health International; 1998.
- 7 Stover J, Heaton L. The costs of contraceptive social marketing programs implemented through the SOMARC Project. SOMARC/The Futures Group International; 1997.
- 8 Stallworthy G, Meekers D. An analysis of unit costs in selected condom social marketing programs, 1990-1996. Population Services International; 1998.
- 9 Stover J, Way PO. Impact of interventions on reducing the spread of HIV in Africa: computer simulation applications. *Afric J Med Pract* 1995;2:110-120.
- 10 Rehle T, Saidel T, Hassig S, et al. AVERT: a user-friendly model to estimate the impact of HIV/sexually transmitted disease prevention interventions on HIV transmission. *AIDS* 1998;12(Suppl 2):S27-S35.
- 11 Stanley EA, Seitz ST, Way PO, et al. The United States Interagency Working Group approach: The IWG model for the heterosexual spread of HIV and the demographic impact of the AIDS epidemic. In: *The AIDS epidemic and its demographic consequences. Proceedings of the UN/World Health Organization Workshop on Modelling the Demographic Impact of the AIDS Epidemic in Pattern II Countries: Progress to Date and Policies for the Future.* New York; 12-1Dec, 1989.
- 12 Van Vliet C, Homes KK, Singer B, Habbema JD. The effectiveness of HIV prevention strategies under alternative scenarios: evaluation with the STDSIM Model. In: Ainsworth M, Fransen L, Over M, eds. *Confronting AIDS: evidence from the developing world. Selected background papers for the World Bank policy research report, Confronting AIDS: public priorities in a global epidemic.* Brussels and Washington (DC): European Commission and World Bank; 1998. p. 207-238.

6

EVALUATING VOLUNTARY HIV COUNSELING AND TESTING PROGRAMS

CLAUDES KAMENGA, THOMAS COATES, THOMAS REHLE

Objectives of VCT Evaluation	81
Evaluating Service Delivery and Service Use	82
Counseling Protocol Adequacy	82
Testing Protocol Adequacy	83
Staff Performance	84
Service Accessibility and Barriers	84
Service Use	85
Sources of Data	85
Evaluating Outcomes and Impact (Effectiveness) of VCT Programs	87
Randomized Design	87
Pre-Post Intervention Client Surveys	87
Conclusion	89
References	90

6

EVALUATING VOLUNTARY HIV COUNSELING AND TESTING PROGRAMS

The World Health Organization (WHO) defines voluntary HIV counseling and testing (VCT) as a confidential dialogue between a client and a care provider aimed at enabling the client to cope with stress and take personal decisions related to HIV/AIDS¹.

Although the effectiveness of this intervention in changing people's behavior to reduce the risk for HIV infection had been under debate until recently, VCT is already a major component of HIV prevention and care programs of most developed countries and is being promoted in many developing countries^{2,3}.

The existing literature on counseling and testing is almost exclusively composed of reports on studies that have tried to assess the effects of VCT on behavior, with before and after intervention evaluation as the most commonly used study design³. Very little in the literature addresses issues such as how well the service is provided, how the service is perceived both by the clients and providers, or how cost-effective the service is as provided. This scarcity of information is unfortunate because people affected by HIV/AIDS want HIV counseling and testing services for future planning (including planning for marriage and children), emotional support,

medical services, and other referral services⁴. As such, VCT services require continued, comprehensive (when possible) evaluation to help adapt the service in response to evolving knowledge, client needs, and technology.

This chapter provides general guidelines for evaluating counseling and testing programs. Although the goals and objectives of counseling and testing programs may vary from one country to another and from one program to another, this chapter is intended to serve as a practical reference for service providers, program managers, and those called upon to evaluate HIV prevention programs.

OBJECTIVES OF VCT EVALUATION

In general, voluntary counseling and testing pursues two interdependent objectives:

- to enable clients to plan and to cope with issues related to HIV/AIDS; and
- to facilitate preventive behavior.

As stated above, the specific objectives of counseling and testing may vary based on the needs expressed or identified during the planning of HIV/AIDS prevention and care programs. For example, in the United States, HIV counseling and testing is used for surveillance, promoting behavior change, public education, and referring individuals into treatment and care systems⁵. In most developing countries, counseling and testing programs are essentially designed to influence clients' risk behavior and facilitate social and medical support for clients who test positive.

For an evaluation to produce results that will inform the design, implementation, and improvement of VCT programs, one must take into account the program's objectives. However, whatever the program objectives are, evaluation activities should address two main areas most relevant for service providers and policymakers:

- **Service delivery**—How well voluntary counseling and testing is provided.
- **Program effectiveness**—The intermediate outcomes and long-term impact that voluntary counseling and testing may have on the population receiving the service.

Once program goals and objectives have been clearly defined, the next critical step is selecting appropriate indicators to monitor and evaluate the VCT intervention. Table 6-1 lists examples of program indicators that may be useful in addressing the different levels of evaluation for voluntary counseling and testing services: service delivery/program outputs, intermediate program outcomes, and expected program impact.

EVALUATING SERVICE DELIVERY AND SERVICE USE

Depending on the goals and objectives of the program and the interests of the program managers, an evaluation of service proficiency may cover all aspects related to providing the service or it may focus on one or more specific aspects. The sources of data will also be a function of the aspect of the service to be evaluated. For example, client interviews are an appropriate means to measure clients' satisfaction with the service. Key aspects of VCT services that should be evaluated include counseling and testing protocol adequacy, staff performance, and service accessibility and barriers. Each is discussed below.

Counseling Protocol Adequacy

Counseling and testing protocols may vary from one program to another based on the goals and objectives of the program. However, whatever the approach taken, the VCT intervention must be regularly evaluated to determine whether it is provided in accordance with the pre-determined protocol and whether it satisfies clients needs. Results can be used to improve the quality of the service provided. Counseling adequacy is defined by the main components and characteristics of voluntary HIV counseling shown in Box 6-1. Questions that must be answered include:

- How well do the counselors follow the counseling protocol?
- Do the clients feel their confidentiality is protected?
- Is risk assessment conducted? If so, how well is it done?
- Is information provided on HIV transmission and risk factors?
- Is a risk reduction plan discussed?
- Is the meaning of the HIV test explained?
- Is the HIV test result clearly given?
- Is emotional support provided?

- Are referrals for medical and social support provided?
- What is the waiting time at the VCT site?
- Is partner notification conducted? If so, how is it done?

An analysis of the answers to these questions will provide feedback to be used in improving the quality of the service provided.

Testing Protocol Adequacy

The testing protocol for a VCT service must be designed to reach maximum reliability and validity in accordance with local conditions, such as the type of equipment available, local HIV seroprevalence, and the resources available to acquire the recommended test kits. The testing protocols used in VCT programs must be examined against the testing strategies for HIV

Illustrative voluntary counseling and testing (VCT) program indicators

TABLE 6-1

Process indicators

SERVICE DELIVERY/PROGRAM OUTPUTS

- proportion of people in the community who know about the VCT service
- number of people counseled and tested at the VCT site (per month, per year)
- proportion of people counseled and tested who have returned to receive their test results*
- proportion of people testing HIV positive who have been referred to appropriate care and support services
- proportion of people counseled and tested who have informed their partners (partner notification)
- composite score of service aspects:
 - adequacy of counseling protocol
 - adequacy of testing protocol
 - adequacy of counseling and testing that is actually provided (staff performance)
 - proportion of clients counseled who are also tested for HIV (linkage between VCT)
 - accessibility of the service to persons who need or want it
 - existence of barriers to clients seeking counseling and testing

Effectiveness indicators

OUTCOMES (INTERMEDIATE PROGRAM EFFECTS)

- changes in HIV/sexually transmitted infection (STI)-related risk behaviors among VCT clients and their partners
- changes in STI trends in sub-populations reached by the program
- reduced stigmatization of, and discrimination against, people in the community affected by HIV/AIDS
- increased community support for people living with HIV/AIDS

IMPACT (LONG-TERM PROGRAM EFFECTS)

- changes in trends in HIV incidence/prevalence in the population or sub-populations reached by the program
- reduced mother-to-child transmission of HIV infection in women of childbearing age targeted by the program
- sustained changes in societal norms in the community reached by the program

* This indicator applies to VCT sites that do not perform rapid, on-site HIV testing.

diagnosis recommended by UNAIDS and the World Health Organization to ensure that they are adequate for the local context⁶. The evaluation of the testing protocol must provide answers to the following questions:

- How consistently is the protocol used?
- How valid is the testing algorithm in terms of specificity and sensitivity?
- How long must clients wait to receive their test result? Are clients comfortable with the waiting period?

- How much does the testing protocol cost?
- Is the testing protocol the most appropriate given local conditions? If not, how can it be improved?

Staff Performance

VCT service requires well-trained and motivated personnel. Regular monitoring of their performance is essential to ensure quality and may help to prevent staff burnout. Focus must be placed on such questions as:

- How well trained are the counselors?
- How well do counselors deliver the protocol?
- Are counselors well informed about other issues relevant to VCT services, including testing technology, options for HIV-positive women who are pregnant, and possible referrals to care and support services?
- How well do counselors meet clients' needs?
- Are counselors appropriately supervised?
- Are the counselors appropriately used?
- What mechanisms are in place to help counselors solve problems and deal with stress?

Service Accessibility and Barriers

It is important to identify factors that affect accessibility and create potential barriers to service use. Parameters to be evaluated include:

- How far must the intended population travel to reach the service?
- Is public transportation to the VCT site available?
- How much does it cost for clients to receive VCT services?

Because clients must pay for medical services in many developing countries, it is essential that cost does not become a barrier to individuals using the service, especially those who might need it the most. When assessing accessibility, it is also important for those who are planning an

evaluation to keep in mind that being near to a VCT site does not always guarantee easy access to the service. In fact, in areas where there are strong stigmas attached to HIV/AIDS, proximity can be a barrier to service use because potential clients may prefer to go to a VCT site far away from the sight of their neighbors, who may suspect them of being infected just because they visited a VCT center. In this context, it is also important to assess who is being reached by the VCT site:

- Are those at highest risk obtaining the services?
- Are significant populations or groups not being reached?

Service Use

The extent to which services are used is an important factor in determining the viability of a program. A VCT service with a minimal level of use by the target population is not cost-effective and therefore unlikely to receive support and continued funding, even if it is effective in other ways. An evaluation of service use must answer the following questions:

- Who uses the service?
- How many clients are served?
- Why do people seek the service?
- Do clients complete all procedures involved in using the service?
- Is the level of use sufficient to justify sustaining the service?

Sources of Data

Evaluation data on VCT service delivery and use can be obtained from various sources. Table 6-2, adapted from Coyle et al., summarizes the sources of data that can be used to evaluate different service aspects⁵.

Service records and the **staff** of counseling and testing sites are an important source of information for the evaluation of VCT services. The sites' records may be used to collect data on the level of service use, the characteristics of clients attending the sites, reasons for using counseling and testing service, and the testing protocol being used. **Standard forms** containing relevant information must be developed and filled out on a regular basis by the VCT staff. These can be used later for evaluation purposes. The staff of VCT sites may provide useful information about

Possible sources of data on VCT service aspects

Service Aspects	Source of Data			
	VCT Sites	VCT Clients	Observers	Population
Protocol adequacy		•	•	
Testing protocol	•	•	•	
Service accessibility	•	•		•
Service use	•	•		•
Staff performance	•	•	•	•

Source: Modified from Reference 5

BOX 6-1

Components and characteristics of voluntary HIV counseling

Components

- information giving
- risk assessment
- development of a risk reduction plan
- explanation of the meaning of HIV test
- informed consent for HIV test
- psychological support
- emotional support
- appropriate referral

Characteristics

- confidential
- personalized
- empowering
- enabling (facilitates the exercise of power, i.e., a woman may be empowered to negotiate condom use, but condoms must be available if she is to exercise this new skill)

their perception of the quality of the service provided to clients, possible barriers to service use by potential clients, the level and quality of supervision provided to counselors, but also information on the impact—physical, emotional, and otherwise—that counseling and testing has on its providers. Information on these issues may be obtained through periodic **key informant interviews or focus group discussions**, if appropriate.

Clients can provide information on all aspects of VCT service delivery and use. These data may be obtained through **exit interviews with a subsample of randomly selected consenting clients** after a counseling session using a standardized questionnaire that focuses on the clients' perceptions of the quality of the session and the counselor's performance, or their impressions of the VCT site in general (for example, accessibility of the site, perceived barriers

Direct observation can help in assessing the adequacy of counseling and testing protocols and the adequacy of the counseling and testing actually provided (staff performance). Some practical issues related to direct observation must be addressed before it is used, however. For example, clients may be concerned about the confidentiality of the information they reveal if there is a third person in the room during the counseling session. The presence of an observer may also alter the natural way a counselor interacts with his/her client. Efforts must be made to reassure clients and minimize pressure on the staff being monitored.

Surveys of the population at large, which include former, current, and potential VCT service users, or **surveys of selected populations**, such as hard-to-reach groups in defined catchment areas of the site, provide complementary informa-

Direct observation can help in assessing the adequacy of counseling and testing protocols and the adequacy of the counseling and testing actually provided.

to use, organization of the service, ability of the service to meet client needs, length of waiting time, cost of the service). **In-depth interviews and focus group discussions** with a selected number of clients may be used to collect additional contextual information.

Non-participant observers as well as “**professional customers**” (or mystery clients) may be used to conduct quality assurance activities through **direct observation** of the VCT procedures.

tion about the performance of the VCT service as perceived by the community. In addition, **population-based surveys** may help to identify what the population expects from the service as well as barriers to service use. Such surveys will also provide information that can be used to characterize people who use and do not use the service, and identify ways to improve the service and make it more accessible and attractive to the population.

EVALUATING OUTCOMES AND IMPACT (EFFECTIVENESS) OF VCT PROGRAMS

Effectiveness evaluation of VCT programs aims to determine how well voluntary counseling and testing services have achieved their intermediate and long-term program goals (see Table 6-1). Intermediate outcome indicators may measure the extent to which the VCT intervention has encouraged behavior change among clients and their partners and changes in STI rates as a biological proxy indicator for adopting preventive behaviors. Other important outcome indicators should measure the reduction in stigma of, and discrimination against, HIV/AIDS-affected people in the community. Measures of long-term program impact attempt to determine whether VCT intervention activities have affected the rate of HIV transmission in the community, including mother-to-child transmission. They should also include an assessment of the impact of VCT on societal norms in the community reached by the program. To approach this difficult task and be able to make meaningful inferences on program effectiveness, evaluators must analyze VCT process data together with other types of data that are collected in the catchment area of VCT services. These data include behavioral survey data, HIV sentinel data, and ethnographic research data.

A recent multicenter randomized trial conducted by the AIDSCAP Project of Family Health International and UNAIDS/WHO in three developing countries has demonstrated the effectiveness of VCT in changing sexual behavior of those counseled and tested¹⁷⁸. However, although VCT has demonstrated its effectiveness in changing risk behavior in these selected study sites, VCT program managers still must determine whether the VCT service they provide makes a difference for those who receive it. An effective counseling model in one community may not be as effective in another community.

Different models may have to be tried to identify the most effective one for a given setting. The following descriptions of different outcome and impact evaluation approaches show how the effectiveness of particular VCT services can be assessed.

Randomized Design

A randomized controlled design—by far the most rigorous way to measure VCT effectiveness—may be used so long as it is assured that participants in the control group also receive a beneficial intervention (see example in Box 6-2). Different VCT protocols may be tested using this design. For example, same-day testing may be compared to the standard protocol requiring the client to return in a week or two for the test results. Or, a two-session counseling protocol (pre- and post-test counseling) may be compared to an open protocol in which a client uses as much counseling as needed. Counseling and testing programs may also be compared with other prevention interventions. Given the high cost of this experimental design and the scarce resources in most developing countries, program managers must carefully weigh the relevance of this design before embarking on a randomized controlled intervention. It should be used only when there is an important conceptual question to be answered that will have regional or international significance. In reality, few programs will ever conduct such trials because of their expense and methodologic complexity.

Pre-Post Intervention Client Surveys

With this approach, a random sample of clients seen at the VCT site(s) is selected to be followed for a given time period. A standardized behavioral survey questionnaire is administered to the selected clients before they receive any intervention and the same questionnaire is administered to them some time (1, 3, or 6 months) after the intervention. To enhance this evaluation approach, data on STI status should be collected from the clients at intake and follow-up to support the

Illustrative example: The AIDSCAP/UNAIDS/WHO Multicenter Randomized Trial

Objectives To assess the impact of voluntary HIV counseling and testing (VCT) on behavior change among persons seeking such services and to determine the cost-effectiveness of the VCT intervention.

Design and Implementation Consenting people were randomized to receive either voluntary HIV counseling and testing intervention (VCT) or standard health information (HI). This study was conducted in three countries (Kenya, Tanzania, and Trinidad) and was advertised by using different strategies, including radio, television, posters, pamphlets, and community outreach. At baseline, a standardized questionnaire was administered to obtain data on sexual behavior, psychological status, knowledge and attitudes about HIV/AIDS and other STIs, and care-seeking behaviors for STIs. Subjects in the VCT group received pre-test counseling and had blood taken after providing consent for HIV antibody testing. They were provided with condoms and asked to return 2 weeks later for HIV serostatus notification and post-test counseling. Those in the HI group were shown an informational videotape containing culturally appropriate information on the prevention of HIV and other STIs and were provided with a stock of condoms.

All subjects were invited to come back to the study center at 6 and 12 months and all were administered a standardized follow-up questionnaire and offered voluntary HIV counseling and testing at this point. At 6 months, study participants were also offered physical examination and laboratory testing for sexually transmitted infections (and offered treatment if found positive for any STI).

To corroborate the self-reported behavior and estimate incidence of STIs, data on STIs were collected. Two

urine samples, one at baseline and a second at 6 months were obtained from all consenting participants. To estimate the incidence of gonococcal and chlamydial infections, a Ligase Chain Reaction (LCR) test was performed on all 6-month urine samples. Then the baseline urine samples of subjects with a positive LCR at six months were subjected to LCR. Those with a negative test at baseline and a positive test at 6 months were considered incident (new) cases.

Key Findings A total of 4,293 participants were recruited (1,515 in Kenya, 1,427 in Tanzania, and 1,351 in Trinidad) with an overall retention rate of more than 75 percent at the 6-month follow-up. Results from this study suggest the following:

- HIV VCT intervention was more effective than health information in reducing risk behavior among individuals recruited in the study. A 35 percent reduction in the proportion reporting unprotected intercourse with their non-regular partners was noted among participants who received VCT as compared to only 15 percent among those who received health information.
- HIV VCT intervention was not associated with an added risk of negative psycho-social outcomes.
- VCT is a cost-effective intervention. The cost to provide the service is \$US 27 per client and the cost per HIV infection averted is estimated to be between \$US 249-\$346 in the two African countries with high HIV prevalence (see Chapter 17 for more details on cost-effectiveness analysis).

survey findings. Although less expensive and complex than a randomized controlled trial design, such surveys are not easy to conduct in resource poor settings and the results are often considerably biased due to substantial follow-up losses. Moreover, a pre-post intervention design with no comparison group does not allow evaluators to control for behavior-modifying effects that are unrelated to expected intervention effects (see Chapter 1, “Conceptual Approach and Framework for Monitoring and Evaluation,” for more information on this issue).

Nevertheless, the evaluation of the AIDS Information Center in Kampala, Uganda, is an example of this approach⁴. Another example is the work by Kamenga and colleagues in Kinshasa, Congo (former Zaire)⁹, where married couples were interviewed on sexual behavior before receiving voluntary HIV counseling and testing. Discordant couples (one partner HIV-positive

and the other HIV-negative) were then followed and assessed monthly for behavior change (questionnaire) and STI incidence (laboratory testing). Figure 6-1 shows some of the outcomes from this work. However, as noted above, these results have to be interpreted with caution because of the inherent methodological weakness of study designs with no comparison groups.

CONCLUSION

There is now increasing support for expanded programs of voluntary counseling and testing to enable people to cope with issues related to HIV/AIDS, to encourage preventive behaviors, and to facilitate access to care and support services for people who test positive for HIV. To ensure continued quality and inform programmatic improvement, evaluating VCT services

Reported condom use and abstinence following voluntary HIV counseling and testing

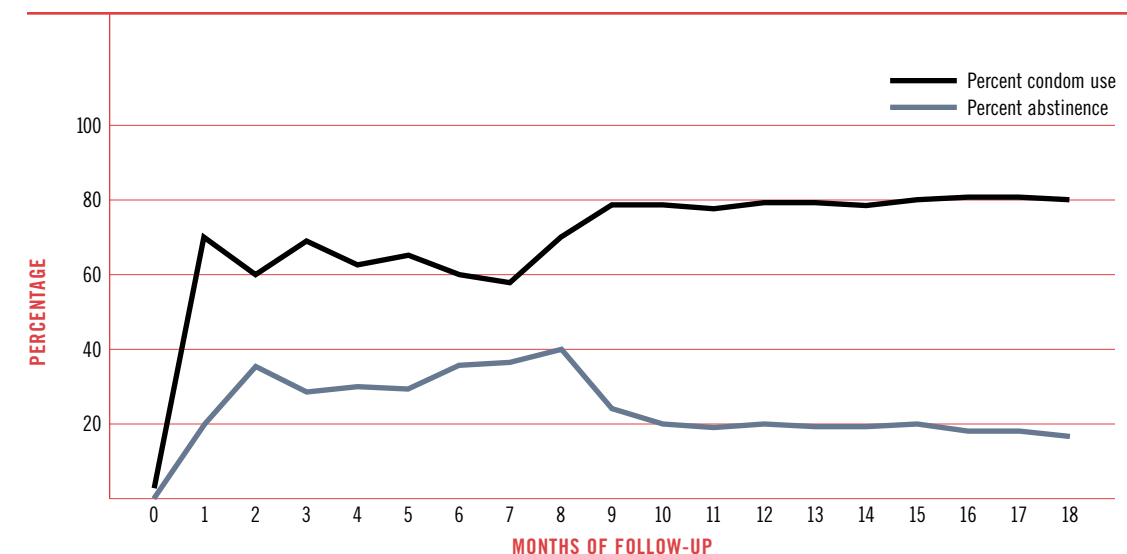


FIGURE 6-1

must be an ongoing process that is integrated into the implementation of the service from the beginning. Evaluative activities will be determined on the basis of program objectives and the available funds for evaluation.

For practical and operational purposes, the evaluation of VCT interventions should focus on key service aspects, such as service use, the adequacy of counseling and testing protocols, staff performance, and service accessibility, and should use complementary sources of information that provide different perspectives on the various service performance aspects. These sources include VCT staff and service records, client surveys, direct observation of VCT service provision, and population surveys in the community reached by the program. Special emphasis must be given to ensuring the confidentiality of sensitive information revealed by clients or

VCT staff. The data collected must be analyzed and used to ultimately provide feedback to all interested parties at different levels (from the staff to the central authorities) and the methodology used must be carefully selected taking in account program priorities and available resources.

Program outcomes related to behavior change, stigma reduction, and community support should be assessed periodically to determine the extent to which voluntary counseling and testing services have achieved their intermediate program goals and objectives. Measures of the long-term program impact should include trends in mother-to-child transmission of HIV in women of childbearing age because voluntary counseling and testing services play an essential role in interventions designed to reduce this mode of HIV transmission.

REFERENCES

- 1 WHO/GPA/TCO/HCS/95.15. Counseling for HIV/AIDS: a key to caring. World Health Organization; 1995.
- 2 De Zoysa I, Philips KA, Kamenga MC, et al. Role of counseling and testing in changing risk behavior in developing countries. *AIDS* 1995;9(Suppl A):S95-S101.
- 3 Higgins, DL, Galavotti, G, O'Reilly KR, et al. Evidence for effects of HIV antibody counseling and testing on risk behaviors. *JAMA* 1991;266(17):2419-2429.
- 4 Campbell CH Jr, Marum EM, Alwano-Edyegu MG, et al. The role of HIV counseling and testing in the developing world. *AIDS Educ Prev* 1997;9(3Suppl):92-104.
- 5 Coyle S, Boruch R, Turner C, editors. *Evaluating AIDS prevention programs*. Washington (DC): National Academy Press; 1991. p. 15-32.
- 6 Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO). Revised recommendations for the selection and use of HIV antibody tests. *Wkly Epidemiol Rec* 1997;72(12):81-87.
- 7 The Center for AIDS Prevention Studies. The voluntary HIV counseling and testing efficacy study. Final report. Arlington (VA): Family Health International; 1998.
- 8 The Voluntary HIV-1 Counseling and Testing Efficacy Study Group. Efficacy of voluntary HIV-1 counseling and testing in individuals and couples in Kenya, Tanzania, and Trinidad: a randomized trial. *Lancet* 2000;356(9224):103-112.
- 9 Kamenga M, Ryder R, Jingu M, et al. Evidence of marked sexual behavior change associated with low HIV-1 seroconversion in 149 married couples with discordant HIV-1 serostatus: experience at an HIV counseling center in Zaire. *AIDS* 1991;5(1):61-67.

7

EVALUATING CARE PROGRAMS FOR PEOPLE LIVING WITH HIV/AIDS

ERIC VAN PRAAG AND DANIEL TARANTOLA

M&E Planning Step 1: Defining the Elements of Comprehensive HIV Care That Need to Be Evaluated	95
A Strategy for HIV and AIDS Care	95
The Continuum of HIV/AIDS Care	96
M&E Planning Step 2: Agreeing on a Framework	98
M&E Planning Step 3: Setting Standards for Comprehensive Care	99
M&E Planning Step 4: Defining Indicators to Measure Comprehensive Care	100
Cross-cutting Issues	100
Combining Quantitative and Qualitative Measurement Tools is Necessary to Capture the Context of Providing Care	100
Standards Will Evolve Over Time	101
Suggested Methods to Measure Indicators and Address Cross-cutting Issues	102
Conclusion	103
References	104

7

EVALUATING CARE PROGRAMS FOR PEOPLE LIVING WITH HIV/AIDS

The responses to the care needs of People Living with HIV/AIDS (PLHA) in highly affected and resource-constrained settings have been very diverse. Numerous enthusiastic small-scale projects with an holistic approach have been developed, while the already

overburdened and fragile general health services continue with a “business as usual” approach because they have not yet coped with additional and new demands. Responses at national referral institutions often focus on externally funded clinical or epidemiological research projects more than resource-constrained service delivery development. Community responses have indeed been developed to meet particular social and prevention needs, while medical needs have not been met. A few, but increasing, number of programs in heavily affected countries have tried to respond to comprehensive care needs. These programs include clinical, psychosocial, and preventive interventions in an integral way and are covering a growing audience^{1,2,3}.

Support required to meet AIDS care needs, either from national governments or international bodies, has been difficult to obtain and limited in financial terms. Indeed, care has never scored high on the priority list of national

HIV/AIDS programs in the developing world because public resources for health have had to be spread so thin between competing public health needs. There was—and continues to be—a fear among donors and decision-makers that resources invested in AIDS care fall into a bottomless pit. Their commitment to funding care is limited by doubts that supporting supposedly fragile medical systems that are already overwhelmed by an increasing number of incurable AIDS patients would divert resources and attention away from HIV prevention or from more pressing care priorities, such as malaria or tuberculosis⁴. Only recently, through the overwhelming humanitarian need to provide care and the pressures from NGOs and people living with HIV in particular, is HIV care back on the donor agenda. Indeed, the need to involve people infected with HIV is seen and understood as critical in planning care programs. The dramatic result of antiretroviral drugs in settings with optimal care

has generated national and international responses to address pricing and patent issues and although there are no real breakthroughs in pricing as yet, the emphasis on access to treatment as a right and a responsibility is being seen in rich and poor countries alike. This emphasis, if translated into greater access, will have consequences for the necessary care delivery systems to ensure safe and effective use of the drugs.

Because the number of comprehensive HIV/AIDS care programs is limited, monitoring and evaluating such a comprehensive approach has not yet been able to provide project implementers and donors with sufficient information on successes or failures. This chapter will look into the “what” and the “how” of HIV care monitoring and evaluation. Monitoring and evaluating care cover a very broad range, similar to that shown in Chapter 1, “Conceptual Approach and Framework for Monitoring and Evaluation,”

immediate results of care activities, such as the number of people in need of palliative care receiving care from a family care provider. Other **outcomes** are the end results of care for the individual, family, and community, such as reduced need to attend hospitals. Finally, **impact information** refers to the long-term effect of care on economic productivity, quality of life, and survival time. Although impact is important to understand, and there is certainly a need to look for cost-effective and sustainable options that link the effectiveness of care systems to the availability of resources, those broader economic and developmental issues are beyond the scope of this chapter, which is geared to practical care providers in the field. We will focus mainly on the inputs and outputs of service delivery programs and lay out a framework for determining the aspects of care that must be monitored to ensure the delivery of quality services.

Monitoring and evaluating HIV care and support is still new, and the development of effective monitoring tools and methodologies is in its early stages.

for prevention. In care, these levels of evaluation range from **information collection** about inputs, such as the infrastructure of care facilities, human capacity, staff attitudes, and socio-economic context in which care is being delivered, to the **activities of actual care delivery** measured against acceptable standards of care, such as the proportion of attenders with an oral Candida infection offered HIV testing, to **output information** or

Monitoring and evaluating HIV care and support is still new, and the development of effective monitoring tools and methodologies is in its early stages⁵. Recently, an international working group coordinated by UNAIDS, WHO, and the MEASURE evaluation project of the University of North Carolina issued a generic HIV monitoring and evaluation guide aimed at the national program level, which

includes a set of proposed indicators for care. These indicators are designed to better guide National AIDS Program Managers and funding sources⁶. WHO is currently field-testing these indicators in five countries in Africa and Asia. However, there is increasing recognition that additional tools are necessary for monitoring and evaluating the ways in which HIV care activities are being implemented at the local level. An additional advantage of locally relevant monitoring and evaluation (M&E) is that it allows for self-monitoring. In that way, an opportunity for the implementing staff will be created to use monitoring and evaluation results directly to refine, adapt, and strengthen their activities.

The process of having those who implement care programs collect and analyze data and use the findings is slowly gaining general appreciation and is essential for improving service performance^{2,7,8}. Such a participatory approach is in line with ongoing public service and health care reforms, which emphasize that decision-making, executive responsibilities, and accountability should be vested at implementing levels, for example, the district or the catchment area of a community-based non-governmental organization (NGO). Capacity building among care providers to monitor and evaluate will be a key requirement.

**M&E PLANNING STEP 1:
DEFINING THE ELEMENTS OF COMPREHENSIVE
HIV CARE THAT NEED TO BE EVALUATED**

A Strategy for HIV and AIDS Care

Care programs should be developed with the active involvement of PLHA while also taking into account capacities and needs of the local health system. Several studies in developing countries have assessed and recently compiled the needs of PLHA and health care providers¹. These needs form the basis of the essential care elements as follows:

- **clinical management**—providing early diagnosis, including HIV testing, rational treatment, and follow up care;
- **nursing care**—promoting and maintaining hygiene and nutrition, palliative care, home care and education to care providers at home and family, and promoting observance of universal precautions;
- **counseling**—providing psychosocial and spiritual support, including stress and anxiety reduction, risk reduction planning and enabling coping, accepting serostatus and disclosure to important others, positive living, and planning of the future for the family; and
- **social support**—providing welfare services, spiritual support, and legal advice; providing information and referrals to such services; and facilitating peer support.

The composition of and emphasis within such a **comprehensive** approach will change over time between the phase of suspicion of infection or early HIV-related illnesses and late stage disease. For example, when infection is suspected, emphasis will be on meeting anxiety needs through counseling, followed at a later stage by meeting clinical needs through managing opportunistic infections. At later stages of the disease, palliative care, combined with social support to ensure a future for the family, become the most pressing needs.

At any moment in time, however, providing the different elements of comprehensive care complement and even potentiate each other and if provided well will result in a better quality of life for affected people. Managing a clinical condition, for example, will be easier and better appreciated if worries about being infected and what will happen in the future can be dealt with through referral to a counselor and to services dealing with social or legal support.

For care to be comprehensive, it must comprise a wide range of support services^{1,9,10}.

These need not necessarily be provided by the same individual or institution but can be provided through networking in partnership with other services, institutions, or projects. Comprehensive care should also include timely referrals between home or community and the hospital and vice versa, effective discharge planning, and follow up at each level¹.

The Continuum of HIV/AIDS Care

The essence of “comprehensive care across a continuum” (Figure 7-1) provides a unique

opportunity to respond to the wide range of medical, emotional, and social support needs of people living with HIV/AIDS and their families^{1,9,11}. To enter a care continuum, a diagnosis has to be made in such a way that it allows further care seeking and support. Experience so far has shown that voluntary counseling and testing (VCT) sites at hospitals or clinics—or even free-standing sites—serve the purpose of helping people know about and accept their serostatus or diagnosis and learn where to go for care¹². During the course of HIV illness, PLHA will

have different needs, and care may need to be provided away from hospitals and closer to home or within the home. Team work and referral mechanisms will allow providers to offer a continuum of care and thus avoid parallel and isolated care efforts.

The principle of a care continuum, applicable to most if not all health issues, is particularly critical to HIV/AIDS for several reasons. First, HIV infection is a chronic state, which may extend over a long period of time if opportunistic infections are prevented or promptly treated. This situation requires that there be a strong synergy between various levels of the care systems, as some infections may be diagnosed and treated at the local level, while others may require effective referral to care institutions. Second, because of the ongoing fear and stigma associated with HIV infection, many people are forced to seek care far from home. This occurs despite the fact that optimal care involves access to care where it is most convenient and cost-effective, and where people benefiting from such care can pursue their professional and personal lives with minimum disruption and without fear. As a general rule, the farther a person has to travel from home to receive care the greater the personal cost (for example, economic cost of arranging travel time away from home and work and costs of time lost).

Self care may not always be possible or sufficient. Skilled health personnel, drawn from primary health care facilities or from the community itself, may be trained to provide such forms of care as palliative and nursing care. Some forms of home-based care will require that outreach programs be created. Mobility of care providers and the efficiency of communication will be essential ingredients in these programs. Other forms of home-based care will build on the strength of community support structures with trained volunteers supervised by the primary health care level. At the home level, social support is often provided by the community itself, with or without external help.

Primary Health Care

Depending on the strength of the national or regional health system and the private and traditional sector, primary care facilities may include clinics and pharmacies, traditional healers, health posts or dispensaries operated by health assistants, health centers managed by trained health care providers, or medically trained personnel, including clinical nurses and physicians. The human resources, skills, collaboration, equipment, supplies and other resources available at this level will determine the type and quality of services locally available, and the criteria for referral of patients to secondary or tertiary care levels.

Secondary Health Care

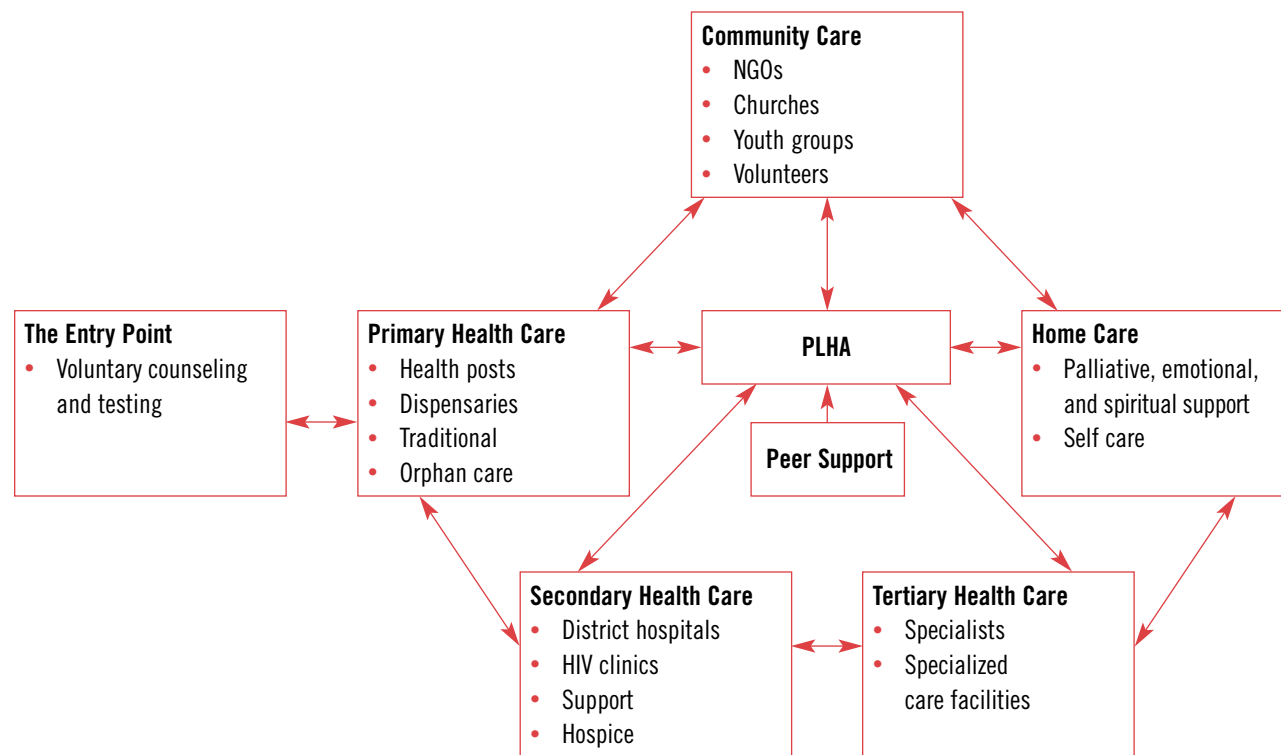
This typically includes a district or regional hospital (public, private, or non-profit), such as tuberculosis treatment centers, HIV clinics, VCT centers, rehabilitation centers, or social and welfare services. Often, NGOs run comprehensive health and social services either on their own or linked with public facilities. PLHA support groups often assist at these facilities. At this level, human resources, biomedical technology, and physical facilities are available to provide complex inpatient, outpatient, and social services. Inpatient facilities for terminal care for those who do not

The various stages across the continuum are:

Home and Community Care

Self care requires skills in prevention and early presumption and treatment of some opportunistic infections and other conditions, such as diarrhea, headaches, or dermatosis. These skills may be acquired by the person living with HIV, and/or by his/her relatives or friends who may also provide needed social support. Peer support groups are very important in supporting PLHA and helping them update and exchange information on optimal self care and care seeking.

FIGURE 7-1 The HIV/AIDS continuum of care



© Eric van Praag/Daniel Tarantola HIV/AIDS and Sexually Transmitted Diseases/WHO

benefit from such support at home may be available at this level in separate units such as day care centers or hospices for the chronically ill.

Tertiary Health Care

Regional, national, or university hospitals (public or private) are expected to offer the highest level of care. This care is often provided by multidisciplinary teams and combined with research. Attendance at tertiary care facilities may evolve with the stage of maturity of the HIV/AIDS epidemic. Such facilities may see and admit most people diagnosed with HIV/AIDS at an early stage of the epidemic. A shift of care functions from tertiary facilities to other levels of the care continuum usually occurs as the epidemic matures, the patient load increases, skilled personnel are more widely available, and case management becomes more systematic. These facilities often have a research agenda in both basic and applied research. National policies, norms, and standards for HIV/AIDS care are often influenced by the knowledge and experience gained within these facilities.

Living with HIV/AIDS implies obtaining medical and other types of support from the most convenient and effective sources of care. The continuum of care will therefore require defining roles and functions within each of the elements of the care continuum; creating services and securing the resources adequate to perform these roles and functions; and building the bridges between each of the elements of the continuum so that referral and return home coincide with the corresponding level of care provision and capacity for any point of the evolution of HIV infection.

M&E PLANNING STEP 2: AGREEING ON A FRAMEWORK

Monitoring and evaluating HIV/AIDS care at the implementation level must be modeled on the concept of the care continuum that is relevant

to the situation one is working in. It should seek to collect, analyze, and use data that reflect the extent and quality of care provided at all levels of the health care system, and identify gaps requiring remedial actions. At the outset, an agreement has to be reached on what the norms or standards of care provision should be in that setting. Standards can be defined in public health terms as something that serves as a basis for comparison and is usually described in a guideline, protocol, or operational procedure for that setting (this definition is adapted from Stedman's Medical Dictionary¹³). Once the standards are set, programs can develop indicators and other measurement tools to assess the scope (how comprehensive are they?), quality (how good are they?), and coverage (how many in need are reached?) of care services at each level.

A fourth, equally important, dimension is monitoring and evaluating the organization and efficiency of the health services provided. The levels of evaluation mentioned earlier in this chapter, which deal with inputs, such as structures (is there privacy to conduct counseling?), supply lines (are drugs regularly available?), capacity (is time allocated to health staff to conduct counseling?), staff attitudes (how much stigma exists within a health facility?) are examples of monitoring efficiency.

For example, a standard for quality of palliative care could be that oral codeine be used in the treatment arsenal of care providers, while the standard for coverage would be that codeine is regularly available from public or private sources by PLHA who need it. The indicator for palliative care would then be the proportion of PLHA within a home care program who have access to codeine.

Table 7-1 provides examples of possible standards that can be put into operation to monitor whether or not a comprehensive approach incorporating the continuum of care is functioning.

M&E PLANNING STEP 3: SETTING STANDARDS FOR COMPREHENSIVE CARE

It is important to emphasize that standards are not necessarily universally applicable, but reflect the best care obtainable in current circumstances in a particular environment or setting. In theory, standards should be formulated for minimum, optimum, or maximum levels of care, taking into account possible variations in the resources or skills available, development of new and cheaper technologies, and the ease of access and affordability in different geographic areas within a particular country. The ultimate goal of many care programs, however, is to achieve equity in the provision of care across countries and populations. Experience has shown that as countries develop their care capacity, inequities are significant and long-lasting features of most care programs. Thus, in many countries, the design of HIV/AIDS care programs and their monitoring and evaluation elements are based on minimum standards that all implementing actors are expected to abide by and use as references to evaluate their performances. Those responsible for setting standards for appropriate care provision in any particular situation need to consider three different dimensions that influence the choice of standards:

- The first dimension deals with the technical aspects of the intervention to be provided and is determined by the efficacy and effectiveness of the specific intervention. For example, isoniazid (INH) prophylaxis for people dually infected with HIV and tuberculosis (TB) is the “universal” standard.
- The second dimension is determined by the social and contextual factors that allow efficacious interventions to function under operational conditions. For example, low demand for HIV testing due to stigma and unavailability of TB drugs for AIDS

care units makes a “universal” standard not relevant for many settings.

- The third dimension is determined by the level at which interventions are provided. Such levels include the community level, the clinic and hospital level, or the tertiary referral level. For example, INH prophylaxis is only a realistic standard at referral sites where active TB exclusion and monitoring can be guaranteed.

Very few care projects have actually developed standards, although there is a current move to define national standards for care in many heavily

Setting standards for monitoring the continuum of care

TABLE 7-1

- Counseling services are provided before meeting clinical needs of patient (scope).
- Clinical care is focused on identifying opportunistic infections during the early part of clinical management (quality).
- Palliative care combined with social support is the focus of clinical care during the later stages of clinical management (scope).
- Ongoing counseling is provided to the patient during clinical management to alleviate his/her concerns regarding prognosis or other personal matters (scope and quality).
- Appropriate and timely referral and networking takes place between home, community services, hospital, etc. (efficiency).
- Synergy exists between levels of health care (home and community, primary, secondary, tertiary) (efficiency and quality).
- Appropriate resources (such as equipment, supplies, skilled staff) are available at each level (efficiency).
- Roles and functions within each element of the care continuum are well-defined (quality).

1 Voluntary Counseling and Testing (VCT)

- 1.1 At least one VCT site accessible in each district.
- 1.2 Pretest counseling (individual or group) available to clinical patients, pregnant women, young people, and general public.
- 1.3 Rapid and confirmatory HIV testing operational at all VCT sites.
- 1.4 Trained counselors authorized to break the news of HIV diagnosis and provide post-test counseling.
- 1.5 Follow-up counselling to be accessible at community or institutional level for each client, if needed.

2 Clinical Management

- 2.1 Management of opportunistic infections and HIV-related illnesses with essential drugs provided at all levels of health system
- 2.2 Preventive therapy with INH for TB/HIV dually-infected people after exclusion of active TB provided through close collaboration between VCT sites, HIV clinical care service, and TB control program staff.
- 2.3 Preventive therapy with cotrimoxazole for people with systematic HIV infection provided in areas with low/moderate resistance to cotrimoxazole with careful monitoring for side effects and effectiveness.
- 2.4 Where affordable, referral to sites where access to the safe and effective use of anti-retroviral drugs (ARVs) is ensured.

3 Social Support

- 3.1 Active referral of all people diagnosed with HIV to social support services from public NGO programs.
- 3.2 Printed directories of available VCT, clinical, social and home care services, provided at all health care sites for each district for active referral and discharge planning across a continuum of care.

4 Home Care

- 4.1 At least one home care service per district to provide palliative care, including pain management, nutritional support, and follow-up services for orphans.

5 Involvement of People Living with HIV/AIDS

- 5.1 Involvement of PLHA and/or PLHA support groups in essential and comprehensive HIV care for each district.

affected countries. Uganda is now completing work on a set of minimal or essential standards for care that will apply to all care-providing institutions, together with an enhanced set of standards applicable where resources and capacities allow. Box 7-1 shows how Uganda has put these principles into practice.

M&E PLANNING STEP 4: DEFINING INDICATORS TO MEASURE COMPREHENSIVE CARE

Once standards have been agreed upon and formulated in a protocol, guideline, or set of performance procedures, indicators or measurement tools must be developed for the monitoring or evaluation exercises. Quantitative HIV care indicators have been developed for programmatic purposes to measure progress in HIV care delivery countrywide in order to improve national planning and resource allocation⁶.

At the implementation level, more detailed quantitative indicators and qualitative information are needed to aim for site-specific performance enhancement. These measurements will assist implementers to detect gaps in scope, quality, coverage, or efficiency and allows for planning remedial action. The examples in Box 7-2 show sets of indicators for the various elements of comprehensive care.

CROSS-CUTTING ISSUES**Combining Quantitative and Qualitative Measurement Tools is Necessary to Capture the Context of Providing Care**

Quantitative assessment through indicators may not convey the full picture needed to assess the results of care provision. An understanding of these facts can only be achieved by using complementary qualitative approaches to help put the quantitative data in context and allow for a

full appreciation of the wide range of technical, social, and psychological aspects that care entails.

Particularly with HIV programs, contextual factors that determine illness perception and care-seeking behavior must also be considered. Contextual factors can affect interactions between the care provider, community, and patient. For example, the occurrence of tuberculosis in an adult may be strongly perceived by family and community as HIV-related and thus induce fear, denial, and non-care seeking, all of which may result in increased risk of spreading tuberculosis as well as HIV. Alternatively, terminal illness in an environment where home care is operational and the family is supportive may result in disease progression with fewer complications. Monitoring and evaluating care activities must therefore measure, as well as possible, changes in these contextual factors as they influence planning, inputs, and program outcomes.

Standards Will Evolve Over Time

Over time, each element relevant to HIV/AIDS care interventions may be subjected to changes as a result of internal (such as program performances) or external (such as social and economic variations) factors. When applied to HIV/AIDS care, the monitoring and evaluation process, particularly with regard to indicators, should therefore be flexible and adapt the indicators to these changes.

Care program inputs will evolve depending on the amount of acceptance within families and communities, the disease burden, priorities, and the resources available.

Program outputs will change with the application of technical innovations, the strength of partnerships addressing comprehensive needs, and the level of efficiency within care delivery institutions.

Several key principles, which transcend the domain of HIV/AIDS care but determine needs and demands, should also be taken into account in devising monitoring and evaluation methods applied to HIV/AIDS care. These principles

include gender sensitivity, age sensitivity, respect for human rights and ethics, and sensitivity to target groups such as drug injecting communities.

Gender sensitivity requires that programs be devised so as to recognize and respond to specific needs, demands, aspirations, and capacities of

Clinical management

- Percent of health units that have the staff trained to manage HIV-related illnesses, manage referrals for HIV-infected patients, and supervise palliative care.
- Percent of health units that are currently stocked with drugs to treat and prevent common opportunistic infections and drugs for palliation, and that have not reported running out of drugs in the past 12 months.

Voluntary Counseling and Testing in relation to HIV care*

- Percent of health care providing sites that have an accessible VCT service staffed by trained counselors providing HIV counseling and testing services free or at affordable rates.
- Percent of people tested HIV positive who have been referred to appropriate clinical care and social support services.

Stigma/discrimination

- Percent of health care providers showing acceptable attitudes during their clinical work toward people with HIV.

Support

- Percent of volunteers in home-based care program able to provide full range of activities as per home care manual or protocol.
- Percent of households in home-based care program receiving full range of locally-agreed home care activities.
- Percent of households with a chronically ill adult aged 15-49 who have received external help in the last 12 months in caring for the patient or replacing lost income.
- Percent of households currently caring for orphans that have received free help with care from outside the family.

* For more on these services, see Chapter 6, "Voluntary HIV Counseling and Testing Programs."

women and men and the constraints imposed on them. Likewise, monitoring and evaluation methods should aim to measure and describe these specifics, and measure the extent to which these are reflected and attended to in HIV/AIDS care initiatives.

Age sensitivity implies that HIV/AIDS care initiatives recognize the specific needs of boys, girls, men, and women at various stages of their lives. For example, HIV/AIDS care services

Factors such as gender and age and respect for non-discrimination and other human rights should provide a lens through which every monitoring and evaluation indicator is examined.

specifically designed for young people may have greater health and social impact than if they were incorporated into health services catering to the needs of all age groups. Age sensitivity also has a role to play when monitoring and evaluation systems are designed. Defining “adults” as 15-49 year old men and women may obscure the needs of 15-18 year olds, who as “children,” may require different care than adults, and dismiss those of adults older than 49 as non-existent or irrelevant.

Respect for human rights is reflected in the extent to which governments create the conditions that enable people to have access to care, and ensure that HIV/AIDS care initiatives do not discriminate on the basis of such factors as race, ethnicity, sex, age, or sexual identity.

Access to information and education, to free movement, and to the right of privacy (through the protection of confidentiality) are examples of human rights particularly relevant to HIV/AIDS care¹⁴. The ethics of HIV/AIDS care, which may in some cases find its roots in human rights law, may add to the obligations of care providers. For example, the ethics of informed consent regarding care (including HIV testing) is critical to HIV/AIDS care initiatives and, as such, should be considered as relevant to the monitoring and evaluation of these initiatives. Likewise, respect for confidentiality and the client’s wish to accept or reject any care-related intervention are also relevant for monitoring and evaluation initiatives.

Considered as the attributes of effective health programs, the combination of sensitivity to factors such as gender and age and respect for non-discrimination and other human rights should provide a lens through which every monitoring and evaluation indicator is examined. For example, if an indicator is the proportion of people having access to tuberculosis prophylaxis, or to HIV testing and counseling, a set of corollary questions should include: Are there differences in this regard between men and women? Between youth and adults? Between races? And among them, which mix of attributes is more likely to be associated with lowest access to these services?

SUGGESTED METHODS TO MEASURE INDICATORS AND ADDRESS CROSS-CUTTING ISSUES

Ideally, indicators such as those shown in Box 7-2 should be made routinely available to guide service performance. However, the complexity of indicators, the cross-cutting issues mentioned above, and, in particular, the confidentiality of all patient-linked information make a routine monitoring approach impractical. If they are carefully planned using quantitative and qualitative approaches, surveys or small studies at regular

intervals can provide necessary information to identify ways to improve services and learn about sensitive issues. However, it will remain difficult to measure coverage because all infected people in need of care cannot easily be identified at the community level.

A health facility-based observational survey and staff interviews can address the scope of comprehensive care elements and capacity issues such as trained staff, drug supply status, and referral capacity (as mentioned in Box 7-2). Focus group discussions with care providers can provide very useful additional information on attitudes, gender or age biases, and personal fears and concerns. When repeated yearly, these surveys, interviews, and discussions can detect trends in both the quantitative and qualitative sense. When it comes to measuring perceptions or experiences of patients or clients receiving care, inpatient or exit interviews at the care facility may not be the best source of information. People with HIV usually do not want to risk being identified through interviews as HIV positive. However, where support groups for people living with HIV exist, innovative ways to gather members and others for focus group discussions are feasible. Accepting that such a group is not randomly selected, conclusions about coverage of care cannot be made, but very useful information on the quality of care and insights about care provision as well as information on stigma and gender issues can be obtained. Household interviews within home care programs for people with chronic illnesses usually identify those with late-stage disease. Again, qualitative methods such as in-depth interviews or focus groups can provide information on quality of home care and care-seeking behavior, while survey methods using questionnaires can elicit data on coverage and scope.

CONCLUSION

This chapter has attempted to propose approaches to developing monitoring and evaluation mechanisms for HIV/AIDS care programs. The proposed approach recognizes that such systems should develop along several guiding principles:

- The standards and norms of care should be defined in each country, for each level of services, and for each population affected, and these norms and standards should be used as baseline reference for monitoring and evaluation purposes.
- The efficiency and effectiveness of the care continuum should be monitored and evaluated in order to explore and document the possible occurrence of gaps in the referral system.
- Monitoring and evaluation systems must be geared to respond to questions that are relevant to decision-making at the local level, where care has been provided.
- Cross-cutting issues should be considered both in designing HIV/AIDS care programs and in drawing up and implementing monitoring and evaluation processes. Monitoring and evaluation processes should build such factors as gender, age, and human rights into their initial design and draw on quantitative and qualitative methods.

By feeding into periodic program planning, the application of findings arising from monitoring and evaluation will affect the equitable access to quality care. The capacity of care systems to add more effective and costly HIV/AIDS drugs to their treatment panoply will depend, to a large extent, on their ability to demonstrate that they can effectively and safely deliver increasingly complex services.

REFERENCES

- 1 Osborne CM, van Praag E, Jackson H. Models of care for patients with HIV/AIDS. *AIDS* 1999;11(Suppl B):S135-S141.
- 2 Kaleeba N, Kalibala S, Kaseje M, et al. Participatory evaluation of counselling, medical and social services of The AIDS Support Organization (TASO) in Uganda. *AIDS Care* 1997;9(1):13-26.
- 3 WHO/GPA/TCO/HCS/95.1. TASO Uganda: the inside story. Geneva: World Health Organization; 1995.
- 4 Mann J, Kay K. Confronting the pandemic: The World Health Organisation's Global Programme on AIDS. *AIDS* 1991;5(Suppl 2):S221-S229.
- 5 Mertens TE, Caraël M. Evaluation of HIV/STD prevention, care and support: an update on WHO's approaches. *AIDS Educ Prev* 1997;9(2):133-145.
- 6 UNAIDS/00.17E. National AIDS programmes: a guide to monitoring and evaluation. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2000.
- 7 WHO/HST/97.2. Co-operation in strengthening national health information systems. Geneva: World Health Organization; 1997.
- 8 Bodart C, Sapirie S. Defining essential information needs and indicators. *World Health Forum* 1998;19(3):303-308.
- 9 WHO/GPA/TCA/HCS/95.14. Provision of HIV/AIDS care in resource constrained settings. Geneva: World Health Organization; 1995.
- 10 Gilks C, Floyd K, Haran D, Kemp J, Squire B, Wilkinson D. Sexual health and health care: care and support for people with HIV/AIDS in resource poor settings. London: Department of Development (DFID); 1998.
- 11 Martin AL, van Praag E, Msiska R. An African model of home-based care (HBC): Zambia. In: Mann J, Tarantola D, editors. *AIDS in the World II*. New York: Oxford University Press; 1996.
- 12 Schopper D, van Praag E, Kalibala S. Psycho-social care for AIDS patients in developing countries. In: Carr SC, Schumacher JF, editors. *Psychology and the developing world*. Westport (CT): Praeger Publishers; 1996.
- 13 Stedmans' Medical Dictionary, 26th edition, Baltimore (MD): Williams & Wilkins; 1995.
- 14 Mann J, Gostin L, Gruskin S, Brennan T, Lazzarini JD, Fineberg HV. Health and human rights. *Health and Human Rights* 1994;1(1):6-23.

III

METHODOLOGIES FOR MEASURING BEHAVIORAL TRENDS

Uses of Behavioral Data for Program Evaluation	109
Sampling Strategies for Monitoring HIV Risk Behaviors	125
Indicators and Questionnaires for Behavioral Surveys	149
Assessing the Validity and Reliability of Self-Reported Behavioral Data	165
The Role of Qualitative Data in Evaluating HIV Programs	173
Effective Dissemination of Data Collection Results	195

8

USES OF BEHAVIORAL DATA FOR PROGRAM EVALUATION

TOBI SAIDEL, ELIZABETH PISANI, STEPHEN MILLS, TIM BROWN, THOMAS REHLE, AND MICHEL CARAËL

The Role of Behavioral Research in Monitoring and Evaluation	110
Informing Effective Program Design	110
Diagnosing the Problem	111
Serving as an Evaluation and Advocacy Tool	111
Explaining Changes in HIV Prevalence Through Data Triangulation	111
Methodological Issues in Behavioral Surveillance	112
Planning for Behavioral Data Collection	112
Choosing Populations to Monitor	112
Considering Measurement Issues	117
Meeting Special Challenges	122
Conclusion	123
References	124

8

USES OF BEHAVIORAL DATA FOR PROGRAM EVALUATION

Control of the HIV epidemic differs from that of other infectious diseases because of the complex and personal nature of the risk behaviors that drive its spread. An understanding of these behaviors is the key to an appropriate response, and tracking them

over time is one of the most crucial elements of an effective monitoring and evaluation system for HIV prevention and care programs. The purpose of this chapter is to describe the contribution of behavioral data to monitoring and evaluation, with an emphasis on outcome evaluation. Methodologies that lead to the availability of better quantitative indicators of behavior change will be discussed in detail.

Many different kinds of behavioral research are needed to guide HIV/AIDS programs in directing prevention resources and helping project implementers plan for appropriate action. Formative behavioral research is needed to help explain the multiple complex factors that influence risk behaviors and to help with the design of effective programs for specific communities. In addition to helping frame the context for prevention efforts, behavioral research also provides a firm understanding of the patterns and distribution of risk in the population. The systems that

are established to monitor these risks feed not only into the design and direction of prevention activities but also into their evaluation.

Seemingly, it would be desirable to have valid and reliable behavioral indicators to evaluate every project that is carried out with a specific sub-population. However, the limited resources for conducting evaluation research and the methodological difficulties involved in gathering high-quality data make this unfeasible in many situations. From the perspective of a national program, it may not be practical, or even necessary, to assess behavior change for every individual project, **especially when those projects are using strategies with already proven effectiveness.** This is especially true in light of the fact that it is usually not possible to attribute to particular interventions changes that have occurred, unless control groups are used. Quite apart from the logistical difficulties of controlled studies, it is usually not ethically feasible to deny

an intervention to a control group when that prevention strategy has already been proven effective. Only in the case of a demonstration project to test a new intervention or study an unanswered research question would there be justification for such a rigorous evaluation design.

In situations where there are multiple interventions addressing overlapping sub-populations, it is often more appropriate to combine the resources of comprehensive HIV prevention and care programs to monitor national or regional trends in behaviors, and to conclude that the changes are attributable to the sum total of program effort. With regard to individual projects, although there is a need for some behavioral research to help implementing agencies evaluate their efforts in a meaningful way, such research should be consonant with the capacity of the agencies that are carrying out the interventions. It may involve the use of qualitative studies or rapid assessments as opposed to representative, large-scale quantitative surveys.

THE ROLE OF BEHAVIORAL RESEARCH IN MONITORING AND EVALUATION

In Chapter 1, “Conceptual Approach and Framework for Monitoring and Evaluation,” a framework for evaluation was described that included a discussion of the input/output level, as well as that of outcome and impact. The importance of monitoring the resources put into a program (such as budget and human resources) and the “products” of those inputs (such as availability of condoms and sexually transmitted infection [STI] treatment) were discussed. However, the evaluation of the effectiveness of HIV prevention programs will almost always require information that informs us about the extent to which the ultimate goals of the program were achieved. The ultimate goal of most HIV prevention programs is to reduce the transmission

of HIV. So prevention programs are likely to focus on interventions that will affect the rate of transmission. Among other things, these include encouraging safer sexual behavior, treating STIs more effectively, and reducing needle sharing between drug injectors. Reductions in the risk behaviors that spread HIV necessarily lead to a reduction in the spread of HIV. For those who are familiar with monitoring family planning programs, risk behaviors act as the “proximate determinants” for HIV infection. They can therefore act as a proxy for measuring program impact.

One of the most powerful ways to measure proximate determinants like behavior change is through repeated quantitative behavioral surveys designed to measure behavioral indicators. The process, also known as **behavioral surveillance**, is used to track different populations who are exposed to the risk of HIV and whose behavior may contribute significantly to the spread of HIV. These surveys are conducted to systematically monitor changes in HIV/STI risk behaviors over time, and are conducted at regular intervals (every 1-2 years). There are several reasons for tracking risk behaviors, which vary in importance depending on the type of epidemic in the geographic area of interest, the type of sub-populations that are affected, and the need for data to advocate for an appropriate response. Whether behavioral indicators are measured at the country, district, or project level, they play a role in describing pathways for spread of the virus, guiding program interventions, evaluating effects of programs, and increasing the understanding of the epidemic among policymakers and the general public.

Informing Effective Program Design

Effective program design requires not only a knowledge of who is at risk, but also an understanding of the levels of different risk behaviors. Knowing how much and what kind of risky

behavior people are engaging in helps in setting priorities for intervention needs, and also in gauging how these priorities may change over time. However, knowing levels of risk behavior alone does not provide the understanding of why people engage in those behaviors and what might motivate them to reduce their risk. Such an understanding is also a critical part of program design. Each new prevention effort requires careful evaluation of many factors, including social, demographic, and contextual factors. Behavioral research, especially qualitative research, can help provide that understanding, and ultimately help program managers design interventions that are better targeted to the specific needs of the communities they are addressing. Methods for qualitative research are described in more detail in Chapter 12, “The Role of Qualitative Data in Evaluating HIV Programs.” They include such techniques as rapid assessment, social and geographic mapping, focus group discussions, and in-depth interviews.

Diagnosing the Problem

One of the most important roles of behavioral surveillance is to provide a “reading” on the levels of risk behavior existing in communities that are already affected by HIV and to suggest pathways for the spread of the virus. Without some notion of existing levels of HIV and a basic understanding of the potential for spread through high-risk behaviors, public health officials and program planners would be at a loss to respond, or at least would lack the tools to do so in an efficient manner.

Repeated quantitative surveys can help identify those sub-populations who are most vulnerable to infection, and also indicate the levels of risk behavior in the general population. Certain sub-populations may interact with people at high risk of HIV infection as well as with people at low risk of infection, thus serving as a behavioral “bridge,” potentially carrying the HIV virus from one population to another. Behavioral data

collection systems inform public health planners about these networks of risk, helping them to make better decisions about which interventions are most appropriate for various groups, and how to prioritize resources.

Serving as an Evaluation and an Advocacy Tool

A good behavioral data collection system can give a picture of changes in sexual and drug-taking behavior over time, both in the general population and in vulnerable sub-populations. The system will record a reduction in risky sex just as it will record persistent risk behavior or shifts in the pattern of risk. These changes can provide an indication of the success of the overall package of activities aimed at promoting safe behavior and reducing the spread of HIV. Likewise, they can also indicate areas where current strategies appear to be inadequate or misguided, indicating that alternate approaches or renewed efforts are needed. Showing that behavior can and does change following efforts to reduce risky sex and drug taking is essential to building support for ongoing prevention activities.

Explaining Changes in HIV Prevalence Through Data Triangulation

Changes in HIV prevalence may indicate the long-term impact of multiple HIV/AIDS prevention interventions, but it is very difficult to prove that observed decreases in prevalence trends are the result of HIV prevention programs. Other factors such as mortality, migration, and saturation of the population at risk can also account for such changes.

A consensus is emerging among decision-makers that prevention programs need to investigate both trends in HIV infection and trends in behavior that may lead to that infection. Indeed, analyzed together with other types of information, behavioral indicators can contribute not only to the understanding of trends in the HIV epidemic but also to the understanding of the relationship

between program effort and impact. When behavioral data are analyzed in combination with output or process data (such as increased condom availability and improved STI services), and other proximate determinants (for example, decreasing STI rates [other than HIV] and changes in risk behaviors), the evidence that prevention efforts may indeed be having the desired effect is reinforced.

METHODOLOGICAL ISSUES IN BEHAVIORAL SURVEILLANCE

With this background information in mind, a discussion of current recommendations in measuring quantitative behavioral indicators follows. These recommendations have evolved out of the experience of initiating multiple behavioral surveys in both general population groups and key sub-populations in many different countries.

Planning for Behavioral Data Collection

Consensus Building

Although it has been discussed already in this book, enough cannot be said about the importance of building consensus and gaining support among various partners, including members of the communities involved, local and national health officials, non-governmental organizations (NGOs), and bilateral and international donors/institutions, before proceeding with large (or even small-scale) behavioral data collection exercises. The goals of consensus building are to:

- make clear to the communities who will be the focus of the information gathering how they will benefit from the research and how their cooperation is needed;
- coordinate among government agencies and various donor agencies on plans for

large-scale behavioral data collection, seeking input from all the relevant players on decisions about what data are needed, which populations should be studied, and which geographic areas should be covered; this participation will help to use resources efficiently and avoid duplication;

- ensure that the best local researchers are involved and that an appropriate institutional base is established for ongoing behavioral data collection.

Rapid Assessment Through Use of Existing Data

Before planning large-scale data collection, it is particularly important to conduct a rapid assessment of risk behaviors and to map out the location and size of risk groups. Care should be taken to use existing data. Previously existing qualitative research can indicate which sub-populations are most at risk in a society, and can provide a better understanding of these populations that will greatly enhance the planning of the quantitative surveillance activities. More information about rapid assessment and mapping of risk groups is described in Box 8-1.

Choosing Populations to Monitor

Repeated cross-sectional surveys can be conducted in samples randomly drawn from the general population or thought to be representative of the general population. They can also be carried out in selected sub-populations whose behavior may lead to a disproportionate risk of contracting or passing on HIV infection.

General Population

General population surveys using household-based sampling frames can provide a credible picture of the extent of risk behavior in the general population and of the links between the general population and groups with higher-risk

behavior, such as sex workers or drug injectors. It is generally recommended that surveys among general population adults take place every 4-5 years. These surveys are important because they provide an understanding of the magnitude of the links between the general population and higher-risk groups that is essential to planning an effective national program and directing resources. If behavioral data collected in the general population show that links to populations with higher-risk behavior are limited, then prevention resources can be concentrated largely in more vulnerable populations, with general population efforts being developed more gradually. Should these links, however, be more extensive, then prevention programs will need to expand coverage of their efforts more immediately.

General population surveys do not necessarily need to be nationally representative in order to be useful to national programs. Large, geographically stratified samples in specific regions can provide as much useful information and at a much lower cost. This type of limited survey is recommended unless there are huge behavioral differences between regions, in which case it will be necessary to adjust data collection strategies so that these differences are adequately covered.

While it may be easier to conduct behavioral surveys among sub-populations thought to be proxy groups for the general population, such as factory workers, there will always be various sources of bias associated with such populations. Therefore, it is recommended that behavioral surveys in general population groups select respondents randomly from sampling frames composed of households.

Sub-populations with High-risk Behavior

Household surveys are not adequate for tracking risk behaviors that are not widespread in the general population but which may contribute disproportionately to the spread of HIV, such as injecting drugs, male to male sex, or selling sex.

To obtain data from individuals with these high-risk behaviors, special sub-population surveys must be conducted.

Bridge Groups

Household surveys also do not reach individuals who are mobile and who tend to spend extended periods of time away from home, such as the military or migrant workers, long-distance truck drivers, or other frequent travelers. These groups also may be disproportionately likely to contract or pass on HIV, and specific surveys are required to reach them in sufficient numbers. These groups have the potential to drive the growth of an epidemic, especially in the early stages. If they become infected in large enough numbers, they can act as a conduit of infection between high-risk sub-populations and the general population. Given this situation, it is important to gather information about risk-taking behavior in these groups, and to design interventions to meet their specific needs.

Youth

Young people are particularly vulnerable, especially in populations where sexual activity begins at an early age. Young people are especially susceptible to HIV in part because they have only recently become sexually active and tend not to be in stable partnerships.

It is appropriate to carry out repeated behavioral surveys only in sub-populations that are the target beneficiaries of prevention programs. Besides being unethical to collect information from a population one has no intention of supporting, it is also pointless. Firstly, behavior is unlikely to change substantially in the absence of prevention programs. Secondly, the purpose of collecting any data is ultimately to inform and/or improve programming. If no programming is planned, data collection is a waste of time, money, and effort.

Identifying and mapping sub-populations

In most places, only limited knowledge exists on the extent and actual forms of risk behavior in communities that are at higher risk or particularly vulnerable to HIV. In many countries, this is true of populations such as sex workers and clients, men who have sex with men (MSM), injecting drug users (IDUs), and children in difficult circumstances. These groups are often stigmatized populations as well. Unfortunately, many people assume that all members of each of these populations are the same and that the extent of and forms of risk behavior are uniform across the population. Prevention programs formulated under these types of simplistic assumptions risk being ineffective and failing to meet the needs or special circumstances of the members of these communities.

In general, risk in any sub-population is not distributed uniformly. For example, depending upon the cultural, legal, and economic environment, there may be several identifiable sub-populations of MSM, each with very different prevention needs. There may be groups of men who visit saunas, others who seek sexual partners in parks or public places, still others who attend local bars, men who sell sex, and men who primarily meet others through occupational networks. Any individual man who has sex with other men may be part of one or more of these sub-groupings. Similarly, sex workers are never a monolithic population. In any given place, sex work will occur in a number of different settings that might include freelance work on the streets, in brothels, in bars, in discos, around lakes or near rivers, in parks, or work through call-in and Internet services.

Marketing language defines each of these identifiable sub-populations with specific characteristics as a “segment.” If a marketer were developing a product for sale, it would produce versions of that product that fit the specific preferences and needs of a particular segment.

Prevention programs need to be defined the same way—that is, they need to be developed as a product that fits the needs of a particular segment of an highly exposed or vulnerable population. In each of the locally defined segments of men having sex with men or sex workers outlined above, different levels of risk behavior (for example, varying rates of condom use, different levels of anal and oral sex, or varying rates of partner exchange) and different prevention needs (for example, access to condoms in places where risk behavior occurs, social support for safer behaviors, or protection from sexual violence) are likely to exist. Unless prevention programs reflect these realities, they will not be as effective as they might be.

In most places, far too little is known of at-risk and vulnerable sub-populations and the way they group into particular segments. We usually do not know the number and distribution of these populations, the levels and determinants of their risk behaviors, or the most appropriate avenues and approaches for designing and implementing effective programs to reduce HIV transmission among them. The segments they might fall into have not been identified systematically nor have their characteristics and needs been determined.

Rapid assessment methodologies have been developed to address these information deficits. Two important components of these approaches are:

- *Geographic mapping*—determining the location and characterization of places where risk behaviors for HIV transmission occur. Geographic mapping approaches involve systematically collecting data on the places and locations where risk behavior in any particular sub-population occurs. Through qualitative methods, such as key informant and in-depth

interviews, ethnographic observation, or focus groups, the types of places where risk behaviors may occur are identified. Once this identification is completed, field teams systematically work their way through a specified geographic area to identify each location of this type, characterize it, and, if possible, determine the number of people engaging in risk behavior at that particular site. (These approaches are described in more detail in the manual *Geographic and Social Mapping of Commercial Sex: A Manual of Procedures*¹.) This information provides a quantitative assessment of the size of the various important segments in a given sub-population.

The data collected during this process are used to:

- elucidate the local distribution of risk settings;
- establish listings for each different type of setting, which can later be used to direct the implementation of prevention programs and ensure adequate coverage of the populations; and
- provide a baseline to examine changes in the distribution of settings that may occur as the HIV/AIDS epidemic and behavioral risk evolve in the country.

- *Social mapping*—rapidly assessing key determinants of risk in a segment and identifying effective messages and approaches for reducing HIV transmission. The idea of social mapping is to use qualitative research techniques, including focus groups and in-depth interviews with sex workers and key informants, and quantitative approaches, such as rapid surveys, to collect information on the current situation, risk behaviors, and factors influencing risk

in different segments of an at-risk or vulnerable population. This is a bottom-up approach that seeks to actively involve members of the at-risk or vulnerable community and other key actors with knowledge of and influence on that community in identifying prevention needs and developing effective avenues for reducing the risk among community members. During the process, detailed information on social, environmental, and economic factors that affect risk in that sub-population and its important segments is gathered.

The findings are then used to:

- set priorities among prevention efforts within the particular sub-population (in terms of number of people in a given segment, levels and forms of risk behavior, or ability to access a particular segment’s members);
- determine what factors need to be addressed by prevention efforts; and
- identify important messages and effective prevention approaches that meet the needs of each segment.

If resources were unlimited, a large number of segments could be defined to cover all the possible ways in which any group or sector of interest might be broken down. Targeted prevention programs might then be developed for each of those segments. However, in practice this would be prohibitively expensive. Designing and implementing prevention programs takes time and effort. Thus, a more realistic approach is to determine the most important factors on which to segment the target population and then to design activities for the most important segments in terms of overall risk.

TABLE 8-1

Recommendations on choosing target groups for repeated cross-sectional behavioral surveys, according to stages of the epidemic

Epidemic Stage	Recommended Survey Groups	Frequency
Low-level	Highly-exposed groups (female sex workers, injecting drug users [IDUs], men who have sex with men [MSM])	Every 1-2 years
	General population	Every 4-5 years
Concentrated and generalized	Highly-exposed groups (female sex workers, IDUs, MSM)	Every 1-2 years
	Bridge groups (military, mobile groups such as miners, long distance truck drivers, agricultural workers)	Every 1-2 years
	Youth	Every 2-3 years
	General population	Every 4-5 years

The choice of which target groups to survey should be driven by the stage of the epidemic. WHO/UNAIDS² defines the stages of the epidemic as follows:

- **Low-level epidemics**—Those with an HIV prevalence assumed to be less than 5 percent in all known sub-populations presumed to practice higher-risk behaviors.
- **Concentrated epidemics**—Epidemics with an HIV prevalence that has surpassed 5 percent in one or more sub-populations presumed to practice higher-risk behaviors but that remains below 1 percent among proxy groups for the general population, such as pregnant women.
- **Generalized epidemics**—Epidemics in which HIV has spread far beyond the sub-populations with higher-risk behaviors,

which are now heavily infected, and is higher than 1 percent among proxy groups for the general population, such as pregnant women.

Table 8-1 summarizes the recommendations about which groups to survey in the various stages of the epidemic.

Groups at high risk of contracting or passing on HIV are important to track relatively often at any stage of the epidemic, and bridge groups should be monitored as soon as HIV is detectable in any significant amount among those groups with high-risk behavior. It is recommended that youth populations be surveyed every 2-3 years once a country has entered into a concentrated epidemic stage in which HIV is present in high-risk groups with whom young people might be interacting. General population household surveys are recommended on a less frequent basis because they are logistically difficult and expensive to carry out.

Some of the advantages of non-household surveys are that by going straight to the groups with the higher-risk behaviors, surveys can be done with smaller samples and less movement of survey teams, and therefore can be done more frequently. However, these advantages are counterbalanced by the fact that high-risk groups outside of households can be harder to reach, in the sense of being more difficult to locate and identify. A great deal of care must be exercised to ensure that systematic and repeatable sampling approaches are used that will help to minimize bias. People often compromise in carrying out behavioral surveys with high-risk groups that are hidden and difficult to identify by using convenience sampling techniques to reach them. The price to be paid for convenience sampling is that one can never be certain about who the data really represent, or whether apparent changes in behavior are real. Making the effort to be systematic about sampling groups that seem to

be disorganized almost always pays major benefits. More details on this subject are included later in this chapter in the section on “Whom to Survey.”

Table 8-2 provides examples of some mixes of target groups that have been chosen for behavioral surveillance surveys (BSS) in different countries.

Just as there are groups that are important to survey, it is important to point out that certain groups are **inappropriate** for behavioral surveillance. Notably, some of the groups that make sense to monitor for HIV surveillance, such as pregnant women and STI clients, make little sense for behavioral surveillance. Asking women in the later stages of pregnancy about their sexual behavior and condom use will **not** generally yield results in any way typical of the female

population at large. Similarly, if one is trying to measure increases in condom use as a way to monitor reductions in risk behavior in a given high-risk “source” population, then sampling STI clients to represent the high-risk population produces a built-in selection bias because STI clients by definition are not likely to use condoms consistently. (More information about general population and target group surveys can be found in previously published documents³⁻⁵.)

Considering Measurement Issues

It is important to adhere to a minimum standard of **rigor** in the conduct of behavioral surveys, if **high-quality** data are to be obtained. If it is expected that behavioral indicators will shed

Examples of respondent groups selected for BSS in various countries

TABLE 8-2

Cambodia	Kenya	Indonesia	Senegal	Thailand
Female	Female	Female	Female	Female
<ul style="list-style-type: none"> • Sex workers • Beer vendors • Working women 	<ul style="list-style-type: none"> • High-paid sex workers • Low-paid sex workers • Youth 	<ul style="list-style-type: none"> • Brothel-based sex workers • Non brothel-based sex workers 	<ul style="list-style-type: none"> • Registered sex workers • University students • Domestic housekeepers • Women in income-generating groups • Office workers 	<ul style="list-style-type: none"> • Direct sex workers • Indirect sex workers • Factory workers • Vocational students
Male	Male	Male	Male	Male
<ul style="list-style-type: none"> • Military/police • Motorcycle drivers • Vocational students 	<ul style="list-style-type: none"> • Bus drivers • Youth 	<ul style="list-style-type: none"> • Truck drivers • Sailors/seaport workers 	<ul style="list-style-type: none"> • University students • Secondary school students • Truck drivers • Apprentices in the informal sector • Office workers 	<ul style="list-style-type: none"> • Army conscripts • Factory workers • Vocational students

light on understanding the dynamics and likely future course of localized HIV epidemics, then it must be recognized that it takes time and effort to do good work. Cutting corners and oversimplifying the process in the name of convenience and ease of implementation is ultimately not helpful because the information produced from such research is not meaningful for monitoring behavior change and evaluating the effects of interventions. In the discussion that follows, three important issues related to collecting high-quality information are discussed. These include:

- what to measure (indicators);
- whom to survey (sampling); and
- how to obtain valid results (validity).

What to Measure

The goal of tracking behaviors in the context of HIV prevention is to learn about behaviors that are critical to the spread of the epidemic. It is very useful to define these behaviors as indicators, so that they may be tracked in a consistent manner over time. Because prevention programs aim to reduce unsafe sexual and drug-taking practices, indicators are designed to track such key indicators as:

- reductions in multiple sexual partnerships;
- increase in condom use with multiple sex partners;
- delays in the onset of sexual activity (for youth);
- reduction in sharing of needles (for injecting drug users).

It is worth mentioning that each of these areas represent the very end-stage of behavior change in most target groups, and that there are many other important knowledge and attitudinal changes that must take place before individuals get to the point of changing these behaviors. Program planners need to be aware of all the

intermediate steps that individuals need to take so that communities can get to these endpoints. Many other behavioral research techniques are available as tools to help assess the needs and progress of communities targeted by interventions. As alluded to earlier in the chapter, these include formative research using qualitative techniques and rapid quantitative surveys. However, with regard to quantitative indicators, it is important to think carefully about the limited set that are worth measuring, and the hard work required to measure them well.

One of the challenges faced in measuring indicators is in knowing how to define such concepts as “high-risk” partners and “condom use.” There are many different ways to define each of these. For example, high-risk partners could include paid sex partners or casual partners that one does not know very well. But a supposed “non high-risk partner,” such as a spouse, could also turn out to be high-risk, if that person is engaging in unprotected sex with an HIV-infected person. Similarly, there are many different ways to define condom use, including condom use at last sex and “consistent” condom use, which can be thought of as using condoms each and every time there is sexual contact. Condom use is also likely to vary depending on who the sexual partner is, so this must also be factored into the indicator.

Because there are so many different ways to define indicators, it is useful to have a standardized set that can be shared among different groups involved in HIV monitoring and evaluation. Working together to define indicators, it is possible to use collective experience to learn what works well and what does not. Evaluation professionals around the world are pursuing ongoing efforts to refine and improve the recommended set of indicators for monitoring HIV prevention programs. More information on recent recommendations is available in Chapter 10, “Indicators and Questionnaires for

Behavioral Surveys,” as well as in other published documents^{5,6}.

Chapter 1 discussed the issue of **setting targets** for indicators. Targets refer both to the magnitude of effect that it is possible to measure and to the level of attainment to be measured by the indicator. The smaller the change it is desired to measure, the larger the sample size that will be needed to do so. This can sometimes pose a problem for evaluators, because it is usually not practical or affordable to work with large samples. This means that if small changes are occurring, such as a 5 percent increase in the percent of men who used a condom the last time they had sexual intercourse with a high-risk partner, or a 5 percent decrease in the proportion of men visiting commercial sex workers in the past 12 months, these changes might not be detectable with the usual sample sizes. Or, if they were detected, they might not be considered significant in a statistical analysis. It is easier to measure larger changes, but prevention programs may not be capable of achieving those large changes, especially during short periods of time. In these instances, what may look like program failure may really be a failure of the methodology to be sensitive enough to measure the change. In addition, despite our best efforts, we do not always know how much behavior change is reasonable to expect in different settings and among different populations because of the complex and unpredictable nature of human sexual behavior.

We must therefore exercise caution in not over-interpreting behavioral indicators, and also in setting reasonable measurement goals in terms of feasibility and cost. It is generally recommended that repeat surveys attempt to measure change on the order of 10-15 percentage points. However, it is most useful to look at the trends over several rounds of data collection rather than to focus on the difference between any two specific data points. This is because of

the strong possibility that any individual data point might well be the result of random fluctuation or chance. It is better to rely on several data points to tell the story than just one or two. This is where the real power of repeated measures comes into play.

It must also be kept in mind that at a given point in time, indicators may reach their maximum levels of change, at which point the focus must switch to the equally important event of **sustaining** behaviors at their present level. When indicators are close to their optimum value, it becomes futile and impractical to keep trying to measure smaller and smaller increments of change. If, after several rounds of behavioral surveillance, behavior changes seem to be leveling off, it may be time to reduce frequency from once a year to once every 2 years. If, on the other hand, risky behaviors persist at high levels, then renewed intervention efforts and continued behavioral surveillance are called for.

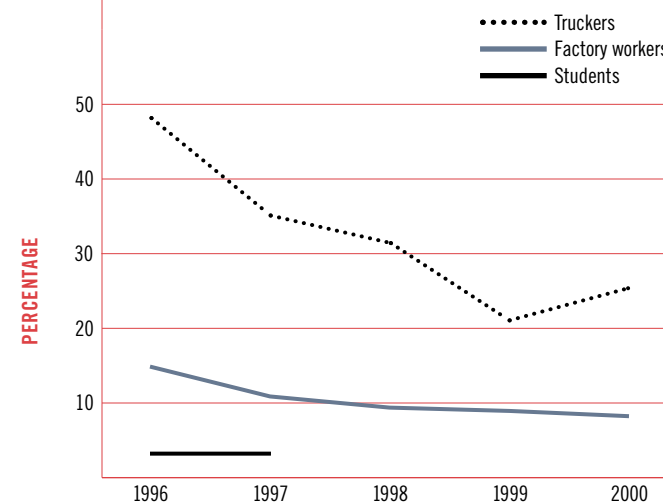
Figures 8-1 and 8-2 illustrate repeated measures of behavioral indicators in which positive results were achieved. One is among sub-populations in Tamil Nadu, India, and the other among female sex workers in Cambodia^{7,8}.

Whom to Survey

The issue of how to sample “hard-to-reach” target groups, such as injecting drug users, sex workers, and in particular, mobile and migrating populations, is one of the most difficult challenges that exists for those involved in behavioral surveillance. Although it is frequently said that it is not possible to do random sampling with these “moving targets,” it is definitely possible to use sound, systematic sampling approaches. Such approaches, if used consistently from one round to the next, increase the likelihood of obtaining reliable estimates of indicators with a minimum of bias. Given the urgent need for scientifically defensible data on behavioral trends in the groups most affected by the epidemic, there is a need to

FIGURE 8-1

Percent of males reporting sex with a non-regular partner in the past year: APAC Tamil Nadu 1996-1998



Source: Reference 7

move from non probability-based to probability-based sampling to the extent feasible.

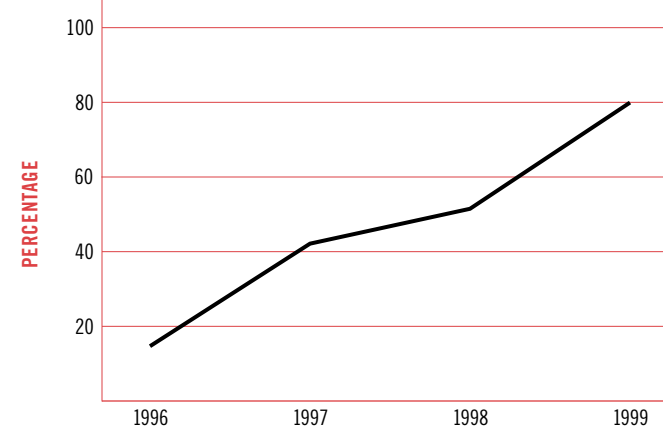
It is not always necessary to have a comprehensive list of all sub-population members to do probability sampling. In fact, because of the sensitive nature of the data being collected, which in many instances involves identifying individuals engaging in illicit behaviors and interviewing them about those behaviors, it is, in fact, preferable not to have a list. This is because of the imperative need to respect the privacy and ensure the confidentiality of all respondents who give their consent to participate in behavioral surveys. All that is needed to do probability sampling is information (or maps) of sites where individuals from the sub-population in question can be accessed. Probability methods can be used for all groups for which a sampling frame of sites or locations where group members congregate can be constructed. For groups for which sampling frames of sites/locations cannot feasibly be created, network or snowball sampling approaches can still be used in such a way as to improve the reliability of estimates. Detailed information on how to conduct sampling for surveillance can be found in Chapter 9, “Sampling Strategies for Monitoring HIV Risk Behaviors,” and also in published guidelines for behavioral surveillance⁵.

How to Obtain Valid Results

Equally as challenging as sampling hard-to-reach groups is the quest for valid results. If it cannot be assured that the data collected are meaningful, then it does not matter how perfect the sampling may be. The results will still not be useful. Apart from sampling error and selection bias, two main sources of error interfere with the ability to gather valid data. One comes from the people collecting the data and the other from the survey respondents. Both contribute to systematic error and both are avoidable if care is taken to exercise quality control during fieldwork.

FIGURE 8-2

Percent of female sex workers reporting always condom use with clients: Cambodia 1996-1999



Source: Reference 8

It is commonly said that people do not tell the truth about their sexual behaviors, and that they exaggerate, withhold information, or refuse to admit to behaviors that are culturally unacceptable. Women, especially, are thought to be reluctant to talk about sex. However, interviewing techniques exist that increase the likelihood of honest sharing of information. Well-trained interviewers who are well-equipped to handle situations that arise in the field are generally able to encourage honest responses and collect data that are credible and consistent with evidence from other sources.

To increase the likelihood of honest responses, interviewers must be thoroughly trained in open and non-judgmental questioning techniques and in accurate recording of responses. The amount of training required will vary depending on who is carrying out the survey. Where peers of those in the respondent group are selected as interviewers, they may be less likely than professional researchers to appear judgmental. Without adequate training they may, on the other hand, also be more prone to recording or coding responses in a way that reflects their own opinions or behavior. Sometimes, using same-sex interviewers or members of the sub-population, such as MSM or IDUs, to interview other members of those sub-populations can make the total difference between being able to obtain valid information or not. There is some evidence that use of computer-generated questionnaires also improves the likelihood of valid responses⁹. During the stage of adapting the questionnaire and organizing the fieldwork, these aspects of the survey, such as the profile of interviewers for obtaining the most valid results, the proper environment for conducting interviews, and the time it takes to complete the interview should be pretested. At this stage, it is sometimes necessary to conduct some focus group discussions or in-depth interviews with members of the target group to understand ways to improve validity of responses.

Building rapport with sub-populations with whom behavioral research is being conducted is

an essential element of success. When researchers have not worked closely with marginalized communities or taken the time to build support for the research process, the result is very often a lack of cooperation and ultimately an inability to obtain high-quality information. There are great benefits to building the support of the communities being studied by involving them in the research process. In addition to getting more meaningful information for monitoring behavior change and evaluating the effects of prevention efforts, learning to access and build trust in the communities to be surveyed will open the doors for later prevention efforts. Failure to do so can backfire severely and contribute to an atmosphere of hostility and distrust that can last for several years.

Among the numerous issues that can help to obtain valid survey results, many relate to the survey instrument itself. Some things that can be done to improve questionnaires are:

- conducting qualitative research before the survey to learn about some of the characteristics of the target group, and how best to approach them;
- comprehensively adapting and pretesting questionnaires so that they are suited to the local context;
- verifying that the language in the questionnaires is clear to the people being interviewed and that the questions are answerable;
- taking the time to do translation and back-translation to make sure that complex concepts are interpretable in a commonly understood manner; and
- using self-or computer-administered questionnaires when dealing with literate populations.

There is some evidence to suggest that people do respond more truthfully when self-administered questionnaires are used, as opposed to

face-to-face interview⁹. However, it is not always necessary that the whole questionnaire be self-administered. Sometimes it is enough for only the most sensitive questions to be self-administered. The important thing is that the same method of questioning, whether it be self-administered or face-to-face, be used for each respondent in a given sample population.

Other quality control issues relate to interactions with the groups being surveyed. Often these surveys are conducted among communities that are on the margins of society or the law, and therefore reluctant to open up to strangers. Working through NGOs who have relationships with the populations in question is an essential component of this work. In some cases, such as with MSM or IDUs, it will be necessary to use members of the community itself (or those working with them) to do the interviewing.

In terms of the actual data collection, care must be taken to ensure that interviews are conducted in strict privacy and out of the earshot of friends and family members. Intensive supervision of interviewers by experienced staff is also a necessity.

Despite the best efforts to control the fieldwork, the nature of self-reported data is such that it is not possible to verify the results objectively. Therefore, it is necessary to use more than one method to assess behavior change—the concept of multiple method triangulation. In the case of behavioral surveillance, it involves the use of qualitative methods after the survey to help interpret the findings. It also involves the use of biological data (HIV and STI prevalence) and other data, such as recorded condom sales or increased use of STI treatment facilities, to validate the findings of behavioral surveys.

Meeting Special Challenges

Although household surveys with general populations have been conducted for a number of years in many countries in the context of evalu-

ating HIV prevention programs, the experience of repeating surveys among hard-to-reach populations with higher levels of HIV risk is new to most countries. Such surveys require skill, sensitivity, and the backing of the communities involved, all of which take time to develop. Some of the special challenges involved in doing these surveys are discussed below.

Ethical Issues for Hard-to-Reach Groups

Ethical issues are always a concern, but may be even more of an issue when dealing with some of the high-risk groups that are most at risk of HIV. Confidentiality is important for all survey participants, but when the participant of the survey is a group involved in illegal activities, such as sex work, injecting drug use, or illegal migration into another country for work, the importance of protecting privacy is magnified. Sometimes, the factors that make these groups so hard to reach are the very ones that put them at elevated risk of HIV infection in the first place. Yet if they were to be sidestepped for collecting information simply because they are difficult to access and complicated to deal with, then this would be tantamount to sidestepping the epidemic itself.

Researchers must be highly sensitive to the reality that for some of the groups being studied, the survey itself could pose a danger because it might expose members of the group to authorities who would fine or imprison them. Social discrimination as a side-effect of the survey also cannot be easily controlled. Even if results cannot be linked to individuals, if they are linked to a community of people, they can still be quite damaging, and provide the impetus for increased stigmatization.

Special efforts must be made to ensure that survey participants understand their rights and the risks involved, and that every effort is made to ensure that the community will benefit from the data collection effort. One of the simplest, but often overlooked ways to do this, is to involve

the community in planning the survey and in disseminating its findings. This can help reduce the perception, frequently held by target groups, that they are merely being used as sources of information that will ultimately benefit other groups.

Who Should Conduct the Research?

Developing the technical capacity of a local institutional base to conduct behavioral surveillance is of paramount importance to ensure continuity over time. The involvement of independent private research firms without a vested interest in the continued use of the data can be problematic and can threaten the sustainability of the system. While this does not necessarily have to be the case, it is clear that without sufficient capacity building and commitments from the national government and international donors, a high-quality surveillance system cannot be maintained. In addition, although many countries are increasing their focus on decentralization and participation at the regional, district, or provincial level, including supporting surveillance, a need still exists for a solid institutional base at the central level to maintain national standards. In some countries, this is being handled by making provincial AIDS committees central to the data collection process and by training a team of interviewers at the local level who can be involved in surveillance on a continual basis from one survey round to the next.

Coordination of Behavioral and Biological Data Collection

It is a good idea to collect behavioral data in the same catchment areas where HIV surveillance is occurring, so that both behavioral and biological trends from the same locations can be observed over time. For the purposes of regular monitoring and evaluation it is not, however, advisable to attempt to collect biological specimens and behavioral data from the same individuals. Some of the reasons for this are that:

- HIV surveillance and behavioral surveillance have different measurement objectives, usually with different sample sizes and some differing sub-populations;
- refusal rates for biological testing and behavioral surveys can be very different; and
- collecting biological specimens accompanied by detailed behavioral information makes it more difficult to ensure confidentiality.

When all is said and done, the logistical and ethical difficulties involved in combining these surveys should be considered in light of the value added, which may be negligible in comparison to the benefits of doing them separately.

CONCLUSION

It is fair to say that behavioral data provide some of the most useful information available in the fight against AIDS, especially because at present, behavior change is the only weapon available for breaking the transmission cycle. However, good behavioral data are not easy to acquire. Without carefully planned data collection strategies and a strong commitment to establishing high-quality data collection systems, programs will continue to collect volumes of useless data.

Although many different types of behavioral research are needed by HIV/AIDS prevention programs, collecting high-quality behavioral indicators fulfills many objectives, including monitoring the dynamics of the epidemic, identifying the groups most at risk and helping describe the patterns of their behavior, and guiding program planning and evaluating program effects. One of the most useful and powerful ways to collect behavioral data is through repeated cross-sectional surveys in groups that are important to the spread of the epidemic.

Repeated behavioral surveys should be conducted for the general population every 4-5

years and more frequently among high-risk groups that are the focus of HIV prevention efforts. Selecting which target groups to monitor should be carefully thought through, taking into account the stage of the epidemic, the presence of various high risk and vulnerable groups, and the data needs of the country.

Because many of the populations that are in the center of HIV spread are difficult to identify and access, attention should be given to working in partnership with the communities at risk, and using systematic, repeatable approaches to sampling and surveying that will produce reliable, valid and unbiased data.

Ethical considerations are particularly important because many of the target groups that are most useful to monitor are also highly marginalized groups that are sometimes engaged in illegal activities. Without their trust and cooperation, successful survey work cannot be achieved.

Behavioral surveillance is now recognized as an essential component of second generation surveillance that supports the improved interpretation of epidemic trends. Therefore, every effort should be made by countries to establish an institutional base for conducting behavioral surveillance and steps should be taken to strengthen the technical capacity to sustain high-quality data collection systems.

REFERENCES

- 1 Brown T, Sittitrai W, Carl G, et al. Geographic and social mapping of commercial sex: a manual of procedures. Honolulu and Bangkok: Program on Population, East-West Center, and Thai Red Cross AIDS Research Center; 2000.
- 2 WHO/CDS/CSR/EDC/2000.5 and UNAIDS/00.03E. Guidelines for second generation HIV surveillance. Geneva: World Health Organization and Joint United Nations Programme on HIV/AIDS; 2000.
- 3 Family Health International/IMPACT and Joint United Nations Programme on HIV/AIDS (UNAIDS). Meeting the behavioral data collection needs of national HIV/AIDS and STD programmes. Proceedings from a joint IMPACT/FHI/UNAIDS workshop. Arlington (VA) and Geneva: Family Health International/IMPACT and UNAIDS; 1998.
- 4 Mills S, Saidel T, Bennett A, et al. HIV risk behavioral surveillance: a methodology for monitoring behavioral trends. AIDS 1998;12(Suppl 2):S37-S46.
- 5 The BSS Development and Implementation Group. Behavioral surveillance surveys: guidelines for repeated behavioral surveys in populations at risk of HIV. Arlington (VA): Family Health International; July, 2000.
- 6 UNAIDS/00.17E. National AIDS programmes: a guide to monitoring and evaluation. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2000.
- 7 AIDS Prevention and Control Project (APAC), Voluntary Health Services, Chennai, India. HIV risk behavior surveillance survey in Tamil Nadu: report on third wave, 1998; Supported by USAID.
- 8 National Center for HIV/AIDS, Dermatology and STDs (NCHADS). Cambodia Ministry of Health. Cambodia's behavioral surveillance survey 1997-1999 (BSS I-III).
- 9 Des Jarlais DC, Paone D, Milliken J, et al. Audio-computer interviewing to measure HIV-risk behavior among injecting drug users: a quasi-randomised trial. Lancet 1999;353(9165):1657-1661.

9

SAMPLING STRATEGIES FOR MONITORING HIV RISK BEHAVIORS

ROBERT MAGNANI, TOBI SAIDEL, AND THOMAS REHLE

Major Issues	128
Prototype Sampling Schemes	129
Cluster Sample Design	129
Targeted Snowball Sampling	132
Applications of the Prototype Designs to Selected Target Groups	133
Sex Workers	133
"Broker"-based Sex Workers	134
Men Who Have Sex with Men (MSM)	134
Injecting Drug Users (IDUs)	136
Youth	136
Mobile Populations	140
Determining Sample Size Requirements	140
Determining the Magnitude of Change to Measure	143
Considering Statistical Power	144
Other Sample and Survey Design Issues	144
Retaining or Replacing Sample Sites or Clusters in Each Survey Round	144
Dealing with Duplicate Observations	145
Dealing with Insufficient Numbers of Target Group Members at Sample Sites	145
Ensuring Replicability Through Documentation	146
Conclusion	146
References	147
Technical Appendix	148

9

SAMPLING STRATEGIES FOR MONITORING HIV RISK BEHAVIORS

As described in the preceding chapter, “Uses of Behavioral Data for Program Evaluation,” using repeated surveys to monitor trends in behaviors that put people at risk of contracting or passing on HIV infection and other sexually transmitted infections (STIs) is an

important aspect of evaluation strategies for HIV/AIDS prevention and care programs. Consensus has coalesced around the idea that monitoring risk behaviors (“behavioral surveillance”) should be undertaken for both the general population and for selected subpopulations whose behaviors or life circumstances put them at risk for HIV transmission (sex workers, men who have sex with men, injecting drug users, mobile populations) or because of their vulnerability to risk (youth)¹. Guidance about undertaking target group surveys and discussions of relevant indicators, questionnaire design, and validity and reliability issues are presented in other chapters of this book. Procedures for undertaking general population HIV risk behavior surveys are described elsewhere².

The method used to choose respondents or subjects is a crucial aspect of any survey effort. Indeed, the credibility of efforts to monitor

trends in HIV/AIDS risk and protective behaviors on the basis of repeated surveys will depend as much as anything on the method(s) used to choose survey subjects. Because most of the groups of interest for HIV target group surveys are difficult to locate or enumerate, sampling presents a formidable challenge.

This chapter provides guidance on sampling for repeated surveys designed to monitor trends on HIV risk-related behaviors in key subpopulations. The chapter begins with a discussion of major sampling issues and problems for target group surveys. Two prototype designs that should cover most target group survey needs are described, followed by illustrative applications of the prototype designs. Sample size requirements for repeated target group surveys are then considered. The final section considers several related survey and sample design issues for sub-population surveys.

MAJOR ISSUES

An important challenge in conducting meaningful target group surveys is to devise sampling plans that are both feasible and capable of producing unbiased estimates (or, more realistically, estimates with acceptably small levels of bias) for population subgroups that are not easily captured in conventional household surveys. First generation HIV risk behavioral surveys in target groups by and large resorted to informal or non-probability sampling approaches. However, recent efforts have attempted to put risk behavior monitoring on a more solid scientific footing by adopting more rigorous sampling methods.

Perhaps the central sampling issue to be addressed is the desirability and feasibility of using probability sampling methods in such undertakings. The major advantages of probability over non-probability sampling are twofold: It is less prone to bias, and it permits the application of statistical theory to estimate sampling error from the survey data themselves³. It is these features that make probability sampling methods the preferred choice whenever feasible. The major disadvantage is that a list or sampling frame is needed. While there are ways to make the task of developing sampling frames less costly and time consuming, it will nevertheless involve greater time and expense than would the adoption of a sampling approach not requiring a sampling frame.

The primary attraction of non-probability sampling methods is that they are less time consuming and costly to implement. However, there are several important drawbacks. The first is the risk of **sampling bias**. Where a list of sampling units is not available from which to select a sample following fixed rules, there is the danger that certain types of subjects will be disproportionately included in and others disproportionately excluded from the sample. The second is the issue of **replicability**, which is of key importance for surveys intended to monitor behavioral trends over time.

Where sample selection criteria are not defined in operationally precise terms such that they may be replicated in subsequent survey rounds, there is the danger that measured trends may be confounded by changes in sampling methodology. Finally, there is the problem that because such methods are not driven by statistical theory, there is no objective basis for assessing the **precision** or **reliability** of survey estimates.

In the final analysis, the issue reduces to one of the relative importance of “defensible” survey findings for the purposes for which the data are sought. In the event of unexpected findings, the use of non-probability sampling methods may leave a program vulnerable to questions about the “representativeness” or “unbiasedness” of the data. This is not to say that in any particular undertaking, a survey based upon non-probability sampling methods will not produce the same results as a probability survey. There is, however, greater **credibility risk** associated with the use of non-probability sampling methods.

In view of the need for accurate information on behavioral trends for HIV/AIDS prevention programs, a case may be made for moving from non-probability to probability sampling methods to the extent feasible. However, is probability sampling feasible for the population subgroups of interest for HIV/AIDS programs? Although probability sampling is more demanding, recent experience indicates that with modest levels of technical support, several national HIV/AIDS programs have been able to make the transition to the use of more rigorous sampling.

As will be demonstrated in this chapter, the basic ideas of probability sampling may be extended in a fairly straightforward manner to cover most of the population subgroups of interest for HIV risk behavior surveys. However, the use of probability sampling methods will not be feasible for some types of target groups, notably those whose members do not tend to congregate in fixed locations and for

whom it is thus difficult to develop a sampling frame. For such groups, the use of non-probability sampling methods is the only alternative.

As a practical matter, sampling for target group surveys will require:

- the use of different sampling strategies for different target groups;
- the collection of data in non-household settings for most target groups;
- the use of conventional probability sampling approaches in non-conventional ways; and
- the occasional use of non-probability sampling methods (in situations where probability methods are infeasible).

PROTOTYPE SAMPLING SCHEMES

Two prototype sampling schemes for HIV target group surveys are described in this section. The first is an extension of conventional **cluster sampling** to groups who are difficult to enumerate⁴. The second is a more rigorous form of snowball sampling known as **targeted sampling**⁵.

Cluster Sample Design

The prototype probability sample design for subpopulation surveys is a **two-stage cluster design**. Adaptations on the basic design should satisfy the sampling requirements for a majority of target groups surveys.

Defining Clusters

Central to the extension of cluster sampling methods to surveys of difficult-to-enumerate population subgroups is a flexible definition of a “cluster.” A cluster, or more precisely a **primary sampling unit** or **PSU**, is any aggregation of elements of interest (such as persons, households, or target group members) that can be unambiguously defined and used as a sampling unit from which to select a sample of elements of interest. Many readers will be familiar with the use of geographic areas as clusters or PSUs in household surveys.

For the purposes of target group surveys, **PSUs may be defined as any identifiable site or location where target group members congregate or may be found**. Box 9-1 provides some illustrative examples of possible operational definitions of PSUs for some of the target groups of interest for repeated HIV risk behavior surveys.

Illustrative primary sampling unit (PSU) or cluster definitions for selected sub-populations

BOX 9-1

Target group	Possible cluster definition
Sex workers	Brothels, massage parlors, bars, restaurants, city blocks, street corners, parks
Men who have sex with men	Bars, nightclubs, parks
Youth	Census enumeration areas, villages, city blocks, schools, workplaces, other locations where youth gather
Mobile populations	Truck stops, depots, customs checkpoints, markets, hotels, factories, plantations, geographic areas in mining towns

Developing Sample Frames

A **sampling frame** is simply a list of subjects of interest for a particular survey or, in the event that such a list does not exist, a list of sites or locations where members of a target group of interest are known to congregate. Sampling frames are an integral part of probability sampling. Indeed, the applicability of probability sampling methods to HIV risk behavior surveys hinges upon it being possible to construct meaningful sampling frames.

Except for the case of youth (who may be covered through household surveys), sampling frame development for subpopulation surveys will require preliminary fieldwork to identify for use as PSUs the locations where members of target

including such things as main streets, main features of the landscape or other identifiable features, and most importantly, main places where target group members “hang out” will suffice. These techniques are explained in Chapter 12, “The Role of Qualitative Data in Evaluating HIV Programs,” in Hogle and Sweat⁶, and in Brown and colleagues⁷.

Occasionally, lists and maps of locations of key gathering points for target group members, such as brothels, bars, massage parlors, truck stops, hotels, recreation sites, schools, or other locations, may already exist. Sometimes, non-governmental organizations (NGOs) who have been working with a subpopulation may have already created maps of their catchment areas.

Sampling frames are an integral part of probability sampling.

Indeed, the applicability of probability sampling methods

to HIV risk behavior surveys hinges upon it being possible

to construct meaningful sampling frames.

groups tend to gather. The process of gathering this information is known as **ethnographic mapping**, which simply means that basic ethnographic techniques are used to create the maps—specifically, participation observation, key informant interviews, and spending time “walking the community.” Creating the sampling frame may only involve creating lists of sites. In other instances, it may be necessary to prepare sketch maps. The maps need not have precise dimensions and distances; rough drawings

It is important that sampling frames cover the entire geographic universe defined for a given survey effort and include the large majority of sites or locations where target group members congregate. If not, the resulting survey estimates will be prone to bias if the behaviors of target group members excluded from the possibility of selection for the survey differ from those who were surveyed. Where the creation of a sampling frame is infeasible for the intended universe for a survey effort, the only alternative under probability

sampling is to restrict the universe to that for which it is possible to create a sampling frame.

Selecting Sample Clusters

Once a sampling frame of relevant PSUs has been created, unless the number of sites or locations is sufficiently small that they may all be covered in a given survey, a sample will need to be chosen. The recommended procedure for doing so will depend upon whether any information on the size of clusters (in other words, the number of target group members associated with each site or cluster) is available before the selection of sample clusters.

Statistically, the most efficient procedure is one in which PSUs or clusters are selected using **systematic sampling with probability-proportional-to-size (PPS)** at the first stage and a constant number of target group members chosen from each PSU at the second stage. Such a design results in a sample in which each target group member has the same overall probability of selection. This is known as a self-weighted sample. In addition to being relatively efficient in terms of sampling precision, this design eliminates the need to weight the data during analysis.

PPS selection should be used when establishments vary significantly in terms of numbers of sex workers associated with them; for example, when some establishments have five times or so as many sex workers as other establishments. Where the numbers of sex workers associated with establishments are roughly comparable, selecting clusters with equal probability will suffice.

However, PPS selection requires that a sampling frame with **measures of size (MOS)** be available or developed before the sample is selected. A measure of size is simply a count or estimate of the number of elements, or target group members, associated with each PSU. Exact counts are not necessary for use as measures of size—rough approximations will suffice.

Unless the errors in measures of size are quite large, the bias introduced into the survey estimates generally will be modest.

When measures of size for clusters are not available, sample clusters will have to be chosen with **equal probability**. Depending upon how sample elements are chosen at the second stage of sample selection, selecting PSUs with equal probability may result in sample elements having different probabilities of selection. In other words, the sample may be **non-self-weighting**, and it will be necessary to apply sampling weights to the data at the analysis stage if unbiased survey estimates are to be obtained.

Selecting Target Group Members within Sample PSUs

In conventional cluster sampling, sample elements are chosen from a list of elements associated with each sample cluster using either simple random or systematic sampling. For many target groups, however, developing a relatively complete list of elements associated with each sample site is likely to be problematic.

Two alternative approaches to second stage sample selection are proposed. The first option, a **quota sampling** approach, entails interviewing target group members as they come into contact with sample sites or locations until a target sample size has been achieved. For example, a sample of truck drivers might be selected by interviewing all truck drivers who happen to appear at the truck stop until the target sample size has been achieved. Note that under this strategy, the length of time required to achieve the target sample size will vary depending upon the volume of contacts with sample sites.

Alternatively, a **“take-all” strategy** could be adopted in which all target group members who come into contact with a sample site during a specified data collection interval (for example, on a particular day or night) would be included in the sample irrespective of their number. The

keys to the “take-all” strategy are that (1) data be collected at each sample site for the same amount of time at each site, and (2) data are obtained from all target group members that come into contact with each sample site during the designated data collection period. Thus, the “take-all” strategy is not recommended when large numbers of target group members congregate at the sites to be used as PSUs or when it is not possible for other reasons to capture all target group members who appear at sample sites during a specified data collection period. In such situations, the quota sampling approach is instead recommended.

For the “take-all” strategy to be workable, it will be necessary to have at least rough estimates of the typical number of target group members associated with each site. This information is needed to determine both how many sites need to be included in the sample and how many interviewers need to be assigned to each site in order to “capture” all of the target groups members who come into contact with each site on the randomly chosen day.

Targeted Snowball Sampling

The primary role envisioned for non-probability sampling in target group surveys is as a substitute for probability methods in situations where the latter prove to be infeasible. This occurs largely in instances where constructing an adequate sampling frame of sites or locations where target group members congregate is not possible. Target groups for which non-probability sampling methods may have to be used include injecting drug users (IDU), some types of sex workers, and possibly men who have sex with men (MSM).

The basic form of non-probability sampling recommended for target group survey efforts is a modified snowball sampling referred to as **targeted sampling**⁵. The basic idea in snowball sampling is to compensate for the lack of a sampling frame by learning the identities and/or

locations of members of a given network of target group members through interviews with informants and other target group members themselves. Snowball sampling is an inherently iterative process. Typically, the data collection process begins by interviewing informants and target group members known to the researchers in order to learn the identities of other target group members. The researchers then contact these persons, collect the data, and obtain information on where additional target group members might be found. Leads from each wave of referrals are followed up until a sample of pre-determined size has been achieved.

An important limitation of snowball sampling is that sample target group members are likely to provide information only on other target group members who are in their own social, economic, and/or sexual network. To the extent that risk-taking and/or protective behaviors differ across networks, this poses a potential bias problem for target group surveys. Research in San Francisco, California, for example, revealed the existence of social networks that differ in terms of race, ethnicity, and type of drug used, even in relatively compact geographic areas. Thus, in order for the snowball sampling approach to yield meaningful monitoring data, it is necessary to ensure that target group members from different networks in a given setting are included in the sample.

Targeted sampling is a combination of street ethnography, stratified sampling, quota sampling, and snowball or chain referral sampling. Watters and Biernacki⁵ describe this approach as being “a purposeful, systematic method by which controlled lists of specified populations within geographic districts are developed and detailed plans are designed to recruit adequate numbers of cases within each of the targets. While they are not random samples, it is particularly important to emphasize

that targeted samples are not convenience samples. They entail, rather, a strategy to obtain systematic information when true random sampling is not feasible and when convenience sampling is not rigorous enough to meet the assumptions of the research design.”

Three basic steps are involved in taking a targeted sample:

- initial geographic mapping;
- ethnographic mapping and stratification; and
- recruitment of quotas of target group members in specified subcategories through snowball sampling.

APPLICATIONS OF THE PROTOTYPE DESIGNS TO SELECTED TARGET GROUPS

The following sections contain examples of how the various sampling designs described above may be used with sex workers, men who have sex with men, injecting drug users, and youth.

Sex Workers

Domains and Stratification

An initial issue to be addressed in undertaking surveys of sex workers is whether different types of sex workers in a given setting differ with regard to risk-taking and protective behaviors. For example, in Senegal a distinction is made between registered and clandestine sex workers; in India, between brothel-based and freelance sex workers; in Kenya, between high- and low-paid sex workers; and in Thailand, between “direct” and “indirect” sex workers (namely, sex workers working in massage parlors or brothels versus those working as bartenders or waitresses in bars or restaurants who also engage in commercial sex). If behaviors are thought or known to differ, it would be advisable to treat the types of sex workers as separate survey

domains. If not, then they can be treated as a single domain (although as separate sampling strata).

Sampling Frame Development

In most settings, at least some sex workers will work from fixed establishments, such as brothels, massage parlors, or bars. For sex workers who do not work from fixed establishments, city blocks, public parks, and other locations where sex workers congregate may be used as sample sites.

Once a list of sites has been created, it can be used to construct a sampling frame consisting of **time-location segments**, which are used as PSUs. To illustrate, suppose that preliminary research in a given setting revealed 20 commercial sex establishments. If establishments were open 7 days per week, a total of 140 PSUs would be formed (20 sites x 7 days). If sample PSUs are to be chosen with probability proportional to size, the listing or sampling frame of establishments should also include a measure of size for each PSU. The appropriate measure of size is the expected number of sex workers at a given site on a given day.

The rationale for doing this is to try to spread out the sample over different times/days of the week in the event that sex workers with differing behaviors work on different days of the week. For example, it might be the case that “part-time” sex workers whose behaviors differ from “full-time” sex workers work only on weekends.

To ensure an adequate distribution of sample PSUs with respect to such characteristics as geographic location and type of establishment, researchers typically order the sampling frame according to such factors. For example, commercial sex establishments might be ordered by first listing establishments located in the northwest quadrant of the city, followed by establishments in the southwest quadrant, and so on. Within each quadrant, establishments would be ordered by type of establishment. If two or more cities are included in a target group survey, geographic

stratification could be accomplished by listing all establishments in the first city, then those in the second city, and so on. The sampling frame development process for a target group survey of sex workers is illustrated in Box 9-2.

BOX 9-2

Illustrative example of the sampling frame development process

In an East African city, outreach workers undertaking HIV prevention activities with sex workers had identified the majority of female sex workers as being either bar-based or club-based. Although there were also home-based and migrating sex workers, the majority fell into the first two categories. An NGO had already established a good rapport with the women, and was able to help establish a sampling frame. Separate sampling frames were created for “high-paid” and “low-paid” sex workers, with each establishment (bars and clubs) being classified as one or the other. The lists of establishments were geographically ordered to assure adequate spread with respect to differing characteristics in different parts of the city.

The ordered lists of establishments were then expanded into “time-location” PSUs. Because there were so many bars in the city, the “time-location” listing for the “low-paid” sex workers did not include all seven nights of the week. Only one weeknight was included, plus Friday and Saturday nights. Thus, for the approximately 200 bars in the “low-paid” sampling frame, there were $200 \times 3 = 600$ time-location PSUs.

As there were only 8 “high-class” clubs and a handful of “high-class” bars, five nights per week (Monday, Tuesday, Thursday, Friday, and Saturday) were included in the sampling frame. Wednesday night was eliminated because this was popularly known as ladies night, when ladies could enter for free. Conducting the survey on Wednesday nights was thought to run the risk of diluting “high-paid” with “low-paid” sex workers.

Selection of Sample Clusters and Sex Workers

Once the sampling frame has been developed, a sample of PSUs can be chosen either with probability-proportional-to-size or with equal probability, and a sample of sex workers using either a quota sampling or a “take-all” approach. Box 9-3 provides an illustrative application of this two-stage cluster sampling scheme for establishment-based sex workers.

“Broker”-based Sex Workers

In some settings, sex workers who are not based in establishments may not congregate in public places, and thus, the cluster sampling approach described above will be infeasible. In India, for example, encounters with sex workers are sometimes arranged through “brokers.” In other settings, arrangements are made by telephone. If a significant portion of the commercial sex trade operates in this fashion in a given setting, then probability sampling methods will not be feasible and the targeted snowball sampling approach will be necessary.

Men Who Have Sex With Men (MSM)

Men who have sex with men (MSM) are difficult to enumerate in sample surveys. However, in many settings, MSM tend to congregate in certain types of establishments or locations (for example, certain bars, nightclubs, parks, or neighborhoods) in sufficient numbers that such locations may be used as PSUs for cluster sampling. In many settings, this may be the only feasible means of gathering behavioral data on MSM. It should be recognized, however, that because not all MSM frequent such locations, this approach is prone to bias to the extent that the behaviors of MSM who frequent such locations differs from those who do not. Alternatively, the targeted snowball sampling approach could be used.

The proposed cluster sampling approach for MSM is quite similar to that used for sex

Illustrative example of a two-stage cluster sampling scheme for establishment-based sex workers

BOX 9-3

Selecting sample PSUs clusters

In the example of the East African city described in Box 9-2, planners decided to use the “take-all” approach to select sample sex workers within sample clusters, and thus sample clusters were chosen with equal probability so that the sample would be self-weighted. The target sample sizes were 300 high-paid and 300 low-paid sex workers. Because the average number of women in “low-paid” establishments was 8, it was decided to select $300/8 = 38$ clusters with equal probability. In “high-paid” establishments, there was an average of 20 women per site. Therefore $300/20 = 15$ clusters were required. However, 20 clusters were chosen in order to ensure adequate distribution. Thirty-eight time-location clusters for “low-paid” sex workers and 20 time-location clusters for “high-paid” sex workers were selected using systematic random sampling with equal probability.

Selecting sample respondents

A “take-all” approach was used to select sample respondents. On the night each site was visited, all sex workers who were present, or who entered the site while the interviewers were working, were contacted for an interview. Each “time-location” cluster was visited only once between 10:00 p.m. and midnight. The interviewer or team

of interviewers spent the same amount of time in each chosen cluster. The hours of 10 p.m. to midnight were chosen as the most likely time when women would arrive to seek clients. If the interviewers had come later, it may have been more difficult to talk with the women, because they would already be engaged with potential clients. In bars, where it was anticipated that only a few sex workers would be present, only one interviewer was sent to the site. Where a large number of sex workers were expected, a team of several interviewers was deployed.

It was difficult to interview women in the bars because they were busy and distracted searching for clients. Thus, when the interviewer contacted a woman, he/she offered her the option of either being interviewed on the spot or of making an appointment to meet at another place and time. An incentive was offered. For woman wishing to be interviewed on the spot, the interviewer offered her a drink in return for the time she was taking away from her work. For women wishing to make an appointment to be interviewed later, the interviewer offered to pay the woman’s transport and buy her lunch. Because some bars/clubs had multiple clusters and could be chosen more than once, interviewers always verified that the woman had not been previously interviewed by one of the other interviewers.

workers who are not based in establishments. The initial step is the development of a sampling frame of locations where MSM congregate. In compiling the list of establishments, attention should be paid to ensuring that the frame covers all geographic parts of the survey universe and that all relevant networks are included, such as those defined by specific ethnic or socioeconomic characteristics.

Once a list of establishments/locations has been developed, time-location sampling units should be created for use as PSUs. For example, if 10 establishments or locations were identified and establishments were open 7 days per week, a total of 70 PSUs would be created. Note, however, that if preliminary research indicated that MSM tended to frequent such establishments only on certain nights, the sampling frame

In 1992-1993, an HIV seroprevalence and risk behavior survey was undertaken among homosexual and bisexual men in the San Francisco Bay area. A variant of cluster sampling was used in conducting the survey. Clusters or PSUs for the survey consisted of locations or sites frequented by young homosexual or bisexual men or by men who had sex with men but did not identify themselves as homosexual or bisexual. The sites considered included street corners or sidewalks, dance clubs, bars, and parks known to be frequented by MSM. On the basis of key community informant interviews and focus groups with groups of MSM, a total of 49 sites were identified. Based upon direct observation and interviews with MSM found at each site, the final list of sites to be included in the survey was reduced to 26 “high volume” sites, defined as sites that yielded two or more eligible persons per hour during “peak” hours.

Interviewing was conducted during afternoons, evenings, and late at night. Intercept methods were used to obtain interviews with eligible respondents. Men appearing to be 22 years of age or younger were approached at each site as they walked on the sidewalk, entered a venue, or waited in line. Interviews and blood drawing were conducted in a specially equipped van parked nearby.

The target sample size for the survey was $n = 500$ men. The sample was allocated to the 26 sites proportionally to the expected number of eligible respondents per hour based upon information gathered during the preliminary research phase. Data were collected during 96 visits to the 26 sample sites.

Comment on the example Because blood samples are not required in target group behavior surveys, a van is not necessary to conduct interviews. Also note that proportional allocation also would have resulted had each site simply been visited for the same amount of time (for example, 2-3 visits per site).

Source: Reference 8

might be limited to such nights. The list of PSUs should be ordered geographically and by establishment type before sample selection. Box 9-4 provides an illustrative application of cluster sampling to behavioral surveys of MSM.

Injecting Drug Users (IDUs)

Of the groups to be covered by target group surveys, IDUs may well be the most difficult to survey. Among the problems likely to be encountered are difficulties in locating sufficient numbers of IDUs and in obtaining cooperation in responding to the survey. There is an absolute need to safeguard the identity, location, and confidentiality of anyone cooperating in the effort to obtain data from potential informants, as well as IDUs themselves.

With regard to sampling, IDUs may not congregate in sufficient numbers for a cluster sampling approach to be effective. However, in some settings it may be possible to identify areas of cities where higher than average concentrations of IDUs may be found. For example, in HIV/AIDS-related research in San Francisco, it has been feasible to use key informant interviews and consultations with police and medical authorities to identify neighborhoods or districts with significant numbers of IDUs. Even if a sufficient number of such areas can be identified, it will still be necessary to identify the different social networks operating. Accordingly, the targeted snowball sampling approach is likely to be the most feasible alternative in most settings. Box 9-5 provides an illustrative example of the use of the targeted snowball sampling approach for collecting survey data on IDUs.

Youth

Youth differ from the other groups that might be covered in target group surveys in that household surveys may be the preferred way to go about monitoring behavioral trends. Only youth who reside at school, who are

institutionalized, or who have no fixed place of residence (for example, homeless or street children) would be excluded from the universe of a household survey.

However, in some settings it may not be acceptable to survey youth at their place of residence about sensitive topics. If so, it will be necessary to identify segments of the general population of youth for whom it is feasible to locate and interview outside of their homes. For example, one might consider for inclusion as proxy groups youth in schools, youth working in the informal sector of the economy (such as street hawkers), and youth working in low-skill occupations in the formal sector (such as domestic workers or apprentices). Finally, special categories of youth, such as homeless or street children, might be considered.

Household Surveys of Youth

When household surveys are to be used to enumerate youth, the conventional two-stage cluster sample design proposed for general population surveys by WHO is the recommended sampling approach². As this sampling scheme is well documented elsewhere, it will not be discussed here. However, a comment on the procedure used to select a sample of youth within sample PSUs is in order.

The preferred procedure is to first create a list or sampling frame of all households containing one or more youth located within each sample PSU, and then choose a sample of households using either simple random or systematic sampling. However, because creating complete lists of households with youth tends to be costly and time consuming, shortcut procedures are often used, which sometimes introduce substantial bias. A more robust shortcut method, referred to as the segmentation method, has seen increasing use in recent years. The basic approach is to divide sample clusters into smaller segments of approximately equal size, choose one segment at random from each cluster, and interview

Researchers in San Francisco, California, sought to gather data on sexual risk-taking behaviors among IDUs for use as baseline measures in evaluating behavior change interventions. The up-front research activities began with an attempt to identify parts of the city (districts and neighborhoods) that contained high concentrations of IDUs and drug-related activity. To do this, direct observations of neighborhoods were made for signs of drug activity. Key informants (drug treatment program staff, police, and residential hotel desk clerks) were interviewed, and police arrest records and emergency room admission data were reviewed. Neighborhoods were then ranked in terms of potential density of IDUs, and maps were developed showing their locations.

In each potential high-activity district, ethnographic mapping exercises were carried out to identify the predominant networks of IDUs. This research indicated that although some subgroups lived in close proximity to each other, their social interaction was minimal or non-existent. It was thus necessary to gain entry into several different networks differentiated along racial, ethnic, and gender lines. Snowball sampling was then used to obtain data for pre-determined sample sizes for each of the various subgroups.

As the project field office was located near a predominantly black housing project, it was easy to recruit black study subjects. In fact, a quota ceiling had to be placed on black study subjects to ensure adequate representation by Latino subjects. Recruiting Latino men required the hiring of a male Latino outreach worker in order to gain access into the network of Latin IDUs and transporting study subjects to and from the project office. Recruiting a sufficient number of female study subjects required the hiring of a female former drug user who knew many of the female IDUs in the study area and brought them to the project office.

Source: Reference 5

all youth found in households in the chosen segment. The advantages of this approach are twofold in that it (1) avoids the household listing operation, and (2) results in a self-weighting probability sample. The method is described in detail elsewhere^{9,10}.

A key issue in household surveys of youth concerns the way in which sample youth who are not available to be interviewed should be handled. In some surveys, fieldworkers are instructed to merely substitute other respondents, such as in a neighboring household. For target group surveys of youth, this practice should be discouraged because of the potential bias that may be introduced. For example, youth who engage in high-risk behaviors may be more likely to live in single-parent households and/or to be at

home less regularly, thus making it more difficult to locate them for a survey interview. If such persons are systematically excluded from target group surveys, the survey data will underestimate the extent of risk behavior. The recommended course of action is to require return visits (“call-backs”) to each sample household in order to obtain an interview from each sample respondent.

School Surveys of Youth

In settings where a sizeable proportion of youth remain in school at the intermediate and secondary levels, conducting surveys in schools represents a cost-effective way of reaching youth. Two cluster sampling schemes for undertaking school surveys are described below. The first is for use when surveys can be conducted in school classrooms using self-administered questionnaires; the second is used when data collection has to take place outside of classroom settings.

The logistically simplest approach is to have students’ complete self-administered questionnaires during class sessions. In addition, the low cost of self-administered questionnaires might enable data to be obtained for larger samples of students than will be feasible if personal interviews are used to collect the data. When “in-class” data collection is possible, a two-stage cluster sample design should satisfy most target group survey needs. Under this design, a sample of schools is first chosen from an ordered list of schools, then a sample of classes is chosen from an ordered list of classes in the identified sample schools, and data are gathered from all students in sample classes. Because measures of size (in this case, the number of school enrollees) are likely to be available before sample selection in most settings, schools should be chosen using systematic sampling with probability-proportional-to-size. Box 9-6 provides an illustrative example of a school survey with in-class data collection.

If in-class data collection in schools is not

possible, it will be necessary to obtain data from students in non-classroom settings. Although it may be possible to schedule appointments with individuals or groups of students to be interviewed either before or after school (see Box 9-7 for an example), it also may be necessary to conduct intercept interviews with individual students at strategically chosen locations, such as outside of classrooms or in cafeterias, lunch rooms, or other common areas where students congregate.

Irrespective of the strategy used, it is important that steps be taken to ensure that the sample is sufficiently well spread out across students of different grades or levels. If students are to be interviewed as they enter or leave class, the classes or sections from which sample students are to be drawn should be chosen using a systematic-random selection procedure similar to that used in selecting classes or sections for in-class data collection.

Workplace Surveys of Youth

In order to obtain behavioral survey data on out-of-school youth, it is first necessary to determine where such youth may be found. One possibility is to interview youth at business establishments that typically employ youth. Examples of workplace sampling frames for youth in the informal sector include businesses employing apprentices, helpers of truck/bus/van drivers, and motorcycle taxi drivers.

As the types of businesses or occupations with significant numbers of youth are likely to vary from setting to setting, a generic sampling approach is proposed here. The recommended approach is a cluster sample design, with business establishments employing youth being chosen at the first stage of sample selection. As with most target group surveys, the sampling frame development process will begin with consultations with key informants and target group members themselves. The purpose of these consultations is to determine businesses

Illustrative example of an out-of-class school survey of youth

BOX 9-7

In a large African city, a behavioral surveillance survey was conducted using a three-stage cluster sample of secondary school youth. The overall sample size was 420 males and 735 females. The primary sampling units were schools, with classrooms within schools used as secondary sampling units. The number of students to be selected from each school was set at 20, meaning that 21 schools were needed from which to choose male respondents and 37 schools for female respondents.

A sampling frame was created of all secondary schools in the city that included an approximate measure of size. Separate sampling frames were created for males and females. Schools educating both female and male students appeared in both lists with different measures of size. Sample schools were chosen using systematic sampling with probability-proportional-to-size (PPS). Within each sample school, four classrooms were selected from ordered lists of classes (by grade level) with measures of size again using systematic sampling with PPS.

Finally, in each sample class, survey supervisors counted the number of students (either male or female, depending on the survey), calculated a sampling interval, and chose a systematic sample of five students per sample class. Arrangements were then made to interview the students outside of class.

that employ youth and the number of youth who are typically found at such businesses.

Once the sampling frame has been developed, a sample of workplaces can be chosen. If measures of size are available, workplaces should be chosen with probability-proportional-to-size and a fixed

BOX 9-6

Illustrative example of an in-class school survey of youth

As part of an evaluation of the national family life education program implemented in public schools in Peru, a KABP-type survey was conducted in a sample of secondary schools. A total of 34 schools were included in the survey, 10 from the capital, Lima, and 24 from cities located in eight Departments in the interior of the country. Schools were chosen using a systematic-random selection procedure with probability proportional to size.

In each sample school, one class or section from each of the five grades at the secondary level of the education system in Peru was randomly chosen to be included in the survey. The estimated average class size was 30 students, and thus the expected sample size per school was $n = 150$ (5 classes x 30 students per class). The total expected sample size was $n = 5,100$ students (34 schools x 150 students per school). Self-administered questionnaires were used in the survey.

Source: Reference 11

number of workers chosen per workplace using systematic sampling at the second stage of selection. However, if the number of workers present at workplaces varies significantly from day to day, it is instead recommended that workplaces be chosen with equal probability and the take-all strategy for selecting sample subjects within sample sites be used. Under this strategy, all youth workers present on the day and time each sample site is visited should be included in the sample. This approach eliminates the need to sub-sample workers in the event that the number present exceeds the target sample size for a given site, or conversely having to return to the site on another occasion in the event that the sample size quota is not met on a single visit. It also results in a self-weighting sample.

Surveys of Youth with No Fixed Residence

For youth who do not have a fixed place of residence, a modified cluster sampling approach in which neighborhoods, city blocks, public parks, and other locations where youth with no fixed residence are known to congregate are used as PSUs. The number of sites of PSUs to be chosen will depend upon how many youth are expected to be found per PSU per interval of fieldwork. If

only small numbers of youth are typically found on a given day or night, more PSUs or clusters will need to be included in the sample to reach the target sample size. Alternatively, the same sites could be visited on more than one night, although this may well result in many duplicate interviews. Note, however, that if this strategy is followed, the number of nights that each site was visited needs to be documented so that the sampling weights can be adjusted accordingly. Additionally, if the number of sites where street youth congregate is small (for example, fewer than 10), it may be necessary to include all sites in the sample.

Mobile Populations

Individuals in mobile populations are of concern for HIV/AIDS programs because they spend considerable periods of time away from home, and in many settings and cultures they tend to engage in casual sexual relationships and use the services of sex workers on a more frequent basis than is observed in the general population. In some cases, mobility may involve crossing national borders. Examples of mobile populations include transportation workers, merchants, and migrant laborers.

The basic cluster sampling approach described above is easily extendable to these groups. The major variation lies in the nature of the sites or clusters to be used in cluster sampling. Possible sites for constructing sampling frames for the various subgroups within the broad heading of mobile populations are indicated in Box 9-8.

DETERMINING SAMPLE SIZE REQUIREMENTS

The primary objective of repeated behavioral surveys is to measure and compare changes in behavioral indicators over time. The size of the sample is a key design parameter in any survey because it is crucial in ensuring sufficient statistical power to detect and measure such changes.

The sample size required **per survey round** to measure change on a given indicator will depend upon five factors:

- The initial or starting level of the indicator:
 - the magnitude of change that evaluators wish to be able to reliably detect;
 - the probability with which evaluators wish to be certain that an observed change of the magnitude specified did not occur by chance (that is, the level of significance);
 - the probability with which evaluators wish to be certain that the actual change of the magnitude specified will be detected.
- The relative frequency with which persons with the characteristics specified in a given indicator may be found in the target group population.

An expression for the required sample size (n) for an indicator measured as a proportion for a given target group in each survey round is given by:

$$n = D \left[\frac{Z_{1-\alpha} \sqrt{2\bar{P}(1-\bar{P})} + Z_{1-\beta} \sqrt{P_1(1-P_1) + P_2(1-P_2)}}{(P_2 - P_1)^2} \right]^2$$

where:

- D = design effect;
- P_1 = the estimated proportion at the time of the first survey;
- P_2 = the proportion at some future date such that the quantity $(P_2 - P_1)$ is the size of the magnitude of change that one wishes to be able to reliably detect;
- $\bar{P} = (P_1 + P_2) / 2$;
- $Z_{1-\alpha}$ = the z-score corresponding to the probability with which one can be certain that an observed change of size $(P_2 - P_1)$ did not occur by chance (that is, the level of significance);

$Z_{1-\beta}$ = the z-score corresponding to the probability with which one wishes to be certain that a change of size $(P_2 - P_1)$ will be detected (that is, the power of the survey).

The design effect (D) is the factor by which the sample size has to increase in order to produce survey estimates with the same precision as a simple random sample. It is based on the homogeneity or similarity **within** and **between** the clusters. In short, the greater the differences between the clusters compared to within the clusters, the greater the sample size must be to compensate for these differences. Assuming that the number of cluster sample sizes can be moderately small in a given survey (not more than 20-25 individuals), the use of a standard value of $D = 2.0$ should adequately compensate for the loss of accuracy resulting from two-stage sampling designs.

The use of this formula is illustrated in Box 9-9.

A table based upon this formula that permits readers to determine final sample sizes without having to perform calculations is provided in a Technical Appendix at the end of this chapter. The table provides sample sizes needed to measure changes in behavioral indicators of a magnitude of 10 and 15 percentage points for different initial values of a given indicator and for different combinations of significance (α) and power (β). The sample sizes provided in the Appendix table are based on one-tailed values of $Z_{1-\alpha}$ (one-sided significance test), assuming a rationale exists for anticipating the direction of change in behavioral indicators in settings where HIV/AIDS prevention interventions have been introduced. This will result in smaller sample sizes than if corresponding two-tailed z-score values of $Z_{1-\alpha/2}$ (two-sided significance test) were to be used. Two-tailed z-score values are appropriate when the direction of change cannot reasonably be predicted and/or if programs wish to take a more cautious stance with regard to

BOX 9-8

Illustrative cluster definitions for subgroups of the mobile population target group

Subgroup	Possible cluster definition
Transportation workers	Truck stops, depots, customs checkpoints
Merchants	Customs checkpoints, markets, hotels
Migrant workers	Factories, plantations, geographic areas in mining towns

Example 1

Suppose that a program desired to detect an increase of 10 percentage points in the proportion of female sex workers (FSWs) who always used a condom with partners and be 90 percent confident of being able to detect a change of this magnitude at the 95 percent level of significance. At the time of the first survey, it is thought that about 30 percent of FSWs always use condoms. Thus, we set $P_1 = .30$ and $P_2 = .40$, and use the one-tailed z-score value for $Z_{1-\alpha} = 1.645$ (95%) and the z-score value for $Z_{1-\beta} = 1.282$ (90%). Inserting these values into the formula, we obtain:

$$\begin{aligned} n &= 2[1.645 \sqrt{2(.35)(.65)} + 1.282 \sqrt{(.3)(.7) + (.4)(.6)}]^2 / (.4 - .3)^2 \\ &= 2[(1.1096 + .8600)^2 / .01] \\ &= 776 \text{ FSWs in each survey round} \end{aligned}$$

Example 2

Suppose that a program desired to detect an increase of 15 percentage points in the proportion of male vocational students who used a condom during last sex in the past 12 months. Levels of significance of 95 percent and power of 80 percent are desired. On the basis of earlier survey data, it is thought that the appropriate “baseline” value on the indicator would be 40 percent. Thus, we set $P_1 = .40$ and $P_2 = .55$, and use z-score values of $Z_{1-\alpha} = 1.645$ (95% significance level for a one-sided test) and $Z_{1-\beta} = 0.84$ (corresponding to 80% power) and obtain:

$$\begin{aligned} n &= 2[1.645 \sqrt{2(.475)(.525)} + 0.84 \sqrt{(.4)(.6) + (.55)(.45)}]^2 / (.55 - .40)^2 \\ &= 2[(1.1617 + .5865)^2 / .0225] \\ &= 271 \text{ vocational students in each survey round} \end{aligned}$$

Note: In the above examples, a design effect of 2.0 was used.

sample size requirements.

For some indicators, a second sample size computation step will be needed. Take, for example, the indicator “proportion of male vocational students

who used a condom during their last encounter with a female sex worker.” In this case, the first step in calculating the sample size required would be to determine how many students would be needed to measure a change in the proportion who used a condom during an encounter with a female sex worker during the previous year as described above—for illustrative purposes, say $n = 200$. However, because only students who had an encounter with a female sex worker in the last year are to be considered in this indicator, it will be necessary to determine how many students would have to be interviewed in order to find the required number of respondents who had sex with a female sex worker during the prior year.

Computationally, the procedure is simple. One merely divides the required sample size calculated as described above by the estimated proportion of the target group that exhibited the required “qualifying” behavior. For example, if 40 percent of male vocational students in a given setting are thought to have had sex with a sex worker in the last year, it would be necessary to interview $n = 500$ ($= 200/.4$) students to find $n = 200$ subjects needed to measure the desired indicator. Additional illustrative computations are provided in Box 9-10.

The more difficult part is anticipating what the appropriate underlying proportion would be. Here, other surveys or anecdotal information might be consulted for guidance. As there may be considerable uncertainty concerning these parameters, **the general guidance is to err toward underestimating the proportion engaging in a given behavior, as this will ensure a sufficient sample size for the main survey effort.** For example, if it were thought that between 20 percent and 30 percent of students typically engage in sex with sex workers on an annual basis in a given setting, the 20 percent figure should be used in determining sample size requirements for target group surveys.

The sample size requirements for any given

target group survey will be the **largest** of the sample sizes calculated for the key indicators measured by the survey.

Determining the Magnitude of Change to Measure

One of the more important considerations in determining sample size requirements is the magnitude of change to be measured. The quantity $(P_2 - P_1)$ is the minimum change in a given indicator that successive target group surveys aspire to be able to measure accurately. Sample size requirements vary inversely with the magnitude of $(P_2 - P_1)$. For small values of $(P_2 - P_1)$, the required sample size may be quite large. For practical reasons, it is thus recommended that risk behavior surveys not attempt to measure changes in behavioral indicators smaller than 10-15 percentage points. Attempts to measure smaller changes will likely exceed the resources available to most such efforts.

It should be emphasized that the magnitude of change in a parameter specified in sample size determination calculations may or may not correspond to program targets with regard to the indicator in question. In some cases, a program might have to accept measuring changes of larger magnitude than what they expect to achieve in a given period of time. This is because measuring changes of smaller magnitude may not be feasible. For example, where condom use in a given setting is only 5 percent, a program might wish to measure a 5 percent increase in a 1-year period. However, they may have to be satisfied with measuring a 10 percent increase over a 2-3 year period instead. This is because the sample size required to detect a change of 5 percentage points may be larger than the available resources for survey-taking can support. In this case, the change parameter $(P_2 - P_1)$ might be set to 10 or 15 percentage points in determining sample size requirements, simply because this is all that is feasible. Even though the program target of increasing condom use by 5 percentage points within 1 year may

have been reached, it will not be possible to conclude statistically that the indicator has changed until a change of 10-15 percentage points has been realized unless, of course, additional resources can be found to support surveys with larger sample sizes. In other words, the sample size is too small to give the survey enough power to detect such a small change as statistically significant and to demonstrate that the program actually had an effect on condom use.

Illustrative calculations of sample sizes needed to locate a predetermined number of target group members with specified characteristics
Example 1

For the indicator “proportion having used a condom during the last sexual encounter with a casual partner” (non-spouses and non-sex workers), suppose that it had been determined, using the formula presented above, that a sample size of $n = 320$ respondents was needed. Data from a recent survey suggest that about 20% of men and 5% of women had engaged in such encounters in an earlier 12-month period. The number of males and females that would have to be contacted in order to obtain $n = 320$ respondents would thus be estimated as:

$$\text{Males: } n = 320 / .20 = 1,600$$

$$\text{Females: } n = 320 / .05 = 6,400$$

Example 2

For the indicator “proportion using a condom during their last sexual encounter with a sex worker” among truck drivers, it is thought that most truck drivers in a given setting (75%) use the services of sex workers during any given 12-month period. Thus, if the required sample size (again calculated as described above was to be $n = 275$), the number of truck drivers that would have to be contacted would be $n = 275 / .75 = 367$ per survey round.

Considering Statistical Power

A second key sample size consideration is that of statistical power. Unless sample sizes are sufficient to be able to detect changes of a specified size, the utility of repeated surveys as a monitoring tool is compromised. To illustrate, suppose we desire to be able to measure a change of 10 percentage points in the proportion of sex workers who always use a condom with their clients. We compare two pairs of hypothetical surveys taken 2 years apart: one with a sample size of $n = 500$ in each survey round and the other with a sample size of $n = 200$ per survey round. While both surveys might indicate the expected increase of 10 percentage points, this change may well not be statistically significant at a given level of significance based upon the surveys with sample sizes of $n = 200$. Thus, we would be forced to conclude that no meaningful change in this behavior occurred over the study period, when, in fact, there was a real increase but it was not statistically significant. To ensure sufficient power, a minimum value of $Z_{1-\beta}$ of .80 should be used; .90 would be preferable where resources permit.

Further guidance on determining sample size requirements for target group surveys is provided in Family Health International's *Guidelines for Repeated Behavioral Surveys in Populations at Risk of HIV*¹².

OTHER SAMPLE AND SURVEY DESIGN ISSUES

Retaining or Replacing Sample Sites or Clusters in Each Survey Round

One of the key design issues in repeated surveys is whether to retain the same PSUs or clusters or choose a new sample of sites in each survey round. There are two advantages to retaining the same sample of sites or clusters. The first is that background characteristics and behaviors of individuals associated with particular sites tend to be correlated over time, and this factor can

increase the statistical precision with which changes are measured. For example, sites such as brothels, bars, or truck stops may attract certain types of target group members and/or may encourage or discourage certain types of behaviors. The effect of this correlation is to reduce the standard error of survey estimates of change between the two survey rounds. Secondly, retaining the same sites reduces the sampling frame development work that needs to be done at the beginning of each survey round.

Balanced against this are several important disadvantages. Among these, the problem of resistance by site “gatekeepers” to repeated visits to the same sites and the loss of sites due to business failures loom especially large. In some settings, the sites where members of certain types of target groups congregate might change so rapidly over time that there is no choice but to construct a new sampling frame and select a new sample of sites in each survey round. Finally, retaining the same sites over an extended period of time does not allow for new sites or “pockets” of risk behavior to be reflected in the behavioral survey monitoring data.

While a compromise strategy of retaining a fixed proportion of sites between any two successive survey rounds and replacing the remaining sample of sites with a new sample might be considered, the advantages of retaining even some sample sites over time in target group surveys are debatable. For one thing, it is unclear that the correlations on characteristics and behaviors over time will be as large for the types of sites used in target group surveys as is often found when residential areas are used as clusters in household surveys. Thus, the magnitude of gains to be realized by maintaining the same clusters is uncertain. The general recommendation, therefore, is to choose a new sample of sites in each survey round.

Dealing with Duplicate Observations

Irrespective of which sampling method is used, one problem that may occur is that of duplicate observations. Duplicate observations may arise because some target group members are associated with more than one PSU. For example, sex workers may work at more than one location, or truck drivers may use more than one truck stop during the course of fieldwork for a survey. There is thus a possibility that the same target group member might be interviewed twice or possibly more.

The statistically correct way to deal with this problem is to adjust the sampling weights to account for the fact that some target group members had more than one opportunity to be included in the sample for a given survey round. However, the recordkeeping, statistical, and data processing requirements of doing so are likely to be beyond the resource capacity in most applications.

A more feasible, but less technically satisfactory solution would be to screen out potential duplicate observations by inquiring whether sample target group members had already been interviewed during the period of survey fieldwork, and not conducting interviews with respondents answering yes. If this approach were adopted, appropriate screening questions would have to be added at the beginning of the survey questionnaires used.

A third option is to do nothing. Except when the total population of a target group of interest is very small, the probability of encountering a sufficient enough number of duplicate observations in a given survey round to introduce serious bias is likely not large enough to worry about.

Dealing with Insufficient Numbers of Target Group Members at Sample Sites

In many of the sampling schemes described above, a number of decisions are driven by an expected number of target group members at each site during a specified time interval. What should be done if, during the course of fieldwork, researchers

find that the actual number of target group members is substantially lower than expected?

Two options are available. The first is to return to sample sites for additional intervals of data collection. The second is to select a supplementary sample of PSUs. Returning to sample sites for an additional interval of data collection is the less desirable option for two reasons. First, if the expected daily volume of target group members were a serious overestimate, returning to the same sites would be an inefficient way of increasing the sample size. Secondly, sampling additional cases per PSU would increase the precision of survey estimates less so than sampling additional PSUs.

What should be done in cases where all PSUs are already included in the sample, and it is thus impossible to choose more PSUs? In this situation, the only alternative is to visit sample PSUs for longer intervals than had been originally planned.

What should be done if even after repeated visits to all PSUs, it is still not possible to reach the target sample size? The answer to this question depends upon the reason why it was not possible to reach the target sample size. One possible cause is that the sampling frame was incomplete. In such a case, one option would be to update the sampling frame and choose a supplementary sample of PSUs of sufficient size to enable the target sample size to be reached. Alternatively, a lower-than-planned sample size could be accepted for the current survey round, and more resources could be put into sampling frame development in subsequent survey rounds in which larger sample sizes would be used (larger sample size in subsequent survey rounds can offset the effects of sample size deficit in earlier rounds).

In Nepal, one approach to dealing with this problem was to inquire from successfully interviewed establishment-based sex workers about any friends who worked at the same establishment but had not been present during the times that data were being collected for the target group survey. These leads were then followed up

and included in the sample as having been sampled from the “referring” establishment. Such an approach should be used cautiously, however. In the Nepal case, researchers found that a number of the leads were in fact not sex workers, but friends of the sex workers who were nominated so that the sex worker could collect the incentive offered for identifying other sex workers. In the final analysis, it may have been preferable to accept a lower than expected sample size.

A final note on the problem is that in some instances, there may simply not be enough target group members in the population. In such cases, the key issue is whether there is sufficient justification for doing surveys for the target group.

Ensuring Replicability Through Documentation

Given the difficult sampling problems posed by HIV/AIDS target group surveys, it is important that steps be taken to make the resulting data as unbiased and sampling plans as replicable as possible. Of crucial importance is that a thorough documentation of sampling plans and adopted selection criteria is prepared to enhance the replicability of data collection efforts over time. This is especially important where probability sampling methods are not used, as the credibility of estimated trends in behaviors over time depends very heavily upon whether a convincing case can be made that identical sampling and survey methods were used across repeated survey rounds. Being able to demonstrate that constant sampling procedures were used adds considerably to the credibility of such estimates.

CONCLUSION

Monitoring trends in HIV risk behaviors through periodic surveys presents some formidable sampling challenges. At the heart of these problems is the fact that many of the population subgroups or target groups that may be of interest for behavioral surveillance or monitoring are difficult to capture in conventional household surveys.

In this chapter, we have described in some detail two approaches to sampling for risk behavior surveys in key HIV target groups. The first approach extends the basic principles of cluster sampling in ways that should be both applicable and feasible for most target groups. The second approach, a more rigorous form of snowball or chain referral sampling, is recommended for use when the development of any type of meaningful sampling frame of sites where target group members congregate is infeasible. Applications of these two strategies to the key HIV target groups were presented in the chapter.

By providing guidance on more rigorous sampling methods, we hope the validity and quality of risk behavior surveillance data can be greatly improved. We acknowledge, however, that the use of these more rigorous sampling approaches in undertaking repeated risk behavior surveys is still in the testing stage. While recent experience suggests that the recommended approaches are feasible, further verification is required. Applications planned in a wide variety of settings over the next few years should provide further guidance on how relatively rigorous /sampling methods for such survey undertakings might be adapted to meet field realities.

REFERENCES

- 1 Family Health International/IMPACT and Joint United Nations Programme on HIV/AIDS (UNAIDS). Meeting the behavioural data collection needs of national HIV/AIDS/STD programmes. Proceedings from a joint IMPACT/FHI/UNAIDS workshop. Arlington (VA) and Geneva: Family Health International/IMPACT and UNAIDS; 1998.
- 2 WHO/GPA/TCO/SEF/94.1. Evaluation of a national AIDS programme: a methods package. Geneva: World Health Organization; 1994.
- 3 Kalton G. Introduction to survey sampling. Newbury Park (CA): Sage Publications; 1989.
- 4 Kalton G. Sampling rare and elusive populations. New York: United Nations, Department for Economic and Social Information and Policy Analysis; 1993.
- 5 Watters J, Biernacki P. Targeted sampling: options for the study of hidden populations. *Soc Prob* 1989;36(4):416-430.
- 6 Hogle J, Sweat M. Qualitative methods for evaluation research in HIV/AIDS prevention programming. AIDSCAP evaluation tools, Module 5. Arlington (VA): Family Health International; 1996.
- 7 Brown T, Sittitrai W, Carl G, et al. Geographic and social mapping of commercial sex: a manual of procedures. Honolulu and Bangkok: Program on Population, East-West Center, and Thai Red Cross AIDS Research Center; 2000.
- 8 Lemp GF, Hirozawa A, Givertz D, et al. Seroprevalence of HIV and risk behaviors among young homosexual and bisexual men: the San Francisco/Berkeley Young Men's Survey. *JAMA* 1994;272(6):449-454.
- 9 Turner A, Magnani R, Shuaib M. A not quite as quick but much cleaner alternative to the Expanded Programme on Immunization (EPI) cluster survey design. *Int J Epidemiol* 1996;25:198-203.
- 10 United Nations Children's Fund (UNICEF). Monitoring progress toward the goals of the world summit for children: Practical handbook for multiple-indicator surveys. New York: UNICEF; 1996.
- 11 FOCUS on Young Adults Program. Impact of the Peruvian National Family Life Education Program on reproductive health knowledge, attitudes, and behaviors. Washington (DC): FOCUS Program and Ministry of Education, Republic of Peru; in press.
- 12 Family Health International (FHI). Guidelines for repeated behavioral surveillance surveys in populations at risk of HIV. Arlington (VA): Family Health International; 2000.

TECHNICAL APPENDIX

Sample size requirements for selected combinations of P1, P2, Z_{1-α} and Z_{1-β}

P ₁	P ₂	Combinations of Z _{1-α} / Z _{1-β} (significance level / percent power)			
		95/90	95/80	90/90	90/80
.10	.20	432	312	331	228
.10	.25	216	156	165	114
.20	.30	636	460	485	336
.20	.35	299	216	229	158
.30	.40	776	558	594	408
.30	.45	352	255	270	186
.40	.50	841	607	646	444
.40	.55	375	271	288	198
.50	.60	841	607	646	444
.50	.65	367	266	282	194
.60	.70	773	558	594	408
.60	.75	329	238	253	174
.70	.80	636	460	485	336
.70	.85	261	189	200	138
.80	.90	432	312	331	228
.80	.95	163	118	125	86

Note: sample sizes shown assume a design effect of 2.0 and are based on one-tailed values of Z_{1-α} (one-sided significance test).

10

INDICATORS AND QUESTIONNAIRES FOR BEHAVIORAL SURVEYS

CHRISTINE KOLARS-SOW, TOBI SAIDEL, JAN HOGLE, JOSEPH AMON, STEPHEN MILLS

Developing and Identifying Indicators for HIV/AIDS Behavioral Surveys	152
Key Indicators in Assessing Behavioral Risk	154
Condom Use	154
Onset of Sexual Intercourse	156
Injecting Equipment Sharing Behaviors	156
Knowledge	157
Risk Perception	157
Additional Indicators	158
Determining the Time Frame of Indicators	158
Developing Questionnaires for HIV/AIDS Behavioral Surveys	159
Informed Consent and Ethical Considerations	160
Questionnaire Administration and Interview Settings	160
Assuring Quality Control Before and During Fieldwork	161
Conclusion	161
Recommended Reading	163

10

INDICATORS AND QUESTIONNAIRES FOR BEHAVIORAL SURVEYS

This chapter examines two areas critical to evaluating behavior change interventions: identifying indicators concerning HIV-related risk behavior and developing questionnaires to measure those indicators.

Since the earliest recognition of the HIV/AIDS epidemic, attempts have been made to identify and define indicators of HIV-related risk behavior that can be used to measure individual and community levels of risk, as well as to track changes in vulnerability over time. At the onset of the epidemic, a major concern was simply assessing general knowledge and awareness of the epidemic, as well as measuring general patterns of sexual behavior. Over time, and with the development of comprehensive HIV prevention and AIDS care programs, the need has been increasingly felt for more detailed indicators that can assess the subtle aspects of behavioral responses to the epidemic as well as help identify obstacles to behavior change.

As the understanding of the HIV epidemic has deepened and the sophistication and orientation of HIV-related programming has developed, it has become increasingly clear that the epidemic cannot be considered from a static

perspective in which all epidemics in all settings are considered to involve the same elements and dynamics. Rather, a more targeted approach has been developed, which considers the epidemic in three stages:

- **Low-level**—HIV prevalence has never risen above 5 percent in groups of people with known high-risk behaviors, including sex workers, injecting drug users, and men who have sex with men.
- **Concentrated**—HIV prevalence has risen above 5 percent in at least one of the groups known to have high risk behaviors, but remains below 1 percent among sexually active adults in the “general population” (represented by antenatal clinic clients).
- **Generalized**—HIV prevalence among sexually active adults in the general population has surpassed 1 percent.

In considering these stages, considerable effort has gone into better defining the critical groups whose behaviors should be tracked over time as well as the indicators used to measure these behaviors. For example, in a concentrated epidemic, priority would be assigned to tracking the behaviors of high-risk sub-populations (such as sex workers and migrant men), whereas in a generalized epidemic, more emphasis would be placed on identifying and tracking risk behaviors in the general population, along with those of high-risk sub-populations.

Consequently, the information provided in this chapter is presented in the context of these developments. In addition, the many lessons learned about how best to ask questions and which indicators provide the most relevant information for measuring change are also presented. Accordingly, this chapter conveys these lessons and proposes standardized indicators and questionnaires appropriate to the stages of the epidemic as well as to the potential sub-populations most important in the epidemic's dynamic. These recommended indicators and questionnaires have been refined over time in collaboration with a variety of national and international partners, in a global initiative to determine a minimum set of key indicators for use in evaluating national HIV/AIDS programs. This key group of indicators is presented here, as well as other indicators deemed important for collecting more detailed data concerning HIV/AIDS-related behaviors.

DEVELOPING AND IDENTIFYING INDICATORS FOR HIV/AIDS BEHAVIORAL SURVEYS

Because of the difficulty in measuring HIV incidence directly, as well as the difficulty in using HIV prevalence as a proxy for incidence, behavioral indicators have been heavily relied upon to help predict and monitor the course of the HIV

epidemic. Monitoring risk behaviors (such as the frequency of unprotected sex with non-regular partners) over the short and medium term provides a means of assessing changes in behavior that might influence the course of the epidemic. The adoption and use of a limited set of behavioral indicators that are sensitive to the dynamic aspects of the HIV epidemic over time is therefore critical to the work of program planners, managers, and evaluators faced with few other interpretable tools.

For an indicator to be useful in a national, regional or local program, it must meet the following criteria:

- An indicator must measure a concept that is relevant to program effort.
- A program must be undertaken that proposes to effect change in the concept being measured (there is no point in measuring something that is not expected to change over time).
- The indicator must be able to measure trends over time.
- It must be feasible to collect data for reporting the indicator.
- The indicator must measure only one concept at a time and be easy to interpret.

Key behavioral indicators recommended for use in HIV and AIDS programs are presented in Table 10-1. These indicators are mostly related to sexual behavior, focusing on number and types of partners and condom use. But in many parts of the world, in addition to sexual transmission, transmission through sharing needles is also responsible for a considerable proportion of new HIV infections. Therefore, indicators relating to sharing drug injecting equipment are included as well. While it may be desirable to monitor indicators concerning knowledge and attitudes (especially in the areas of stigma and discrimination), including questions related to

Key behavioral indicators in HIV/AIDS programs

TABLE 10-1

Key behavioral indicators in HIV/AIDS programs	
<p>Indicators for Adults</p> <ol style="list-style-type: none"> 1 Knowledge of HIV prevention methods 2 No incorrect beliefs about AIDS transmission 3 Number of non-regular partners in the last year 4 Condom use at last sex with a non-regular, non-commercial partner 5 Consistent condom use with non-regular, non-commercial partners 6 Commercial sex in the last year (men only) 7 Number of commercial partners in the last year (men only) 8 Condom use at last sex with a commercial partner (men only) 9 Consistent condom use with commercial partners (men only) 10 Unprotected sex with a higher-risk partner 11 Population seeking voluntary HIV tests 12 Exposure to interventions 	<ol style="list-style-type: none"> 5 Sex workers injecting drugs in the last 12 months 6 Sex workers seeking voluntary HIV tests 7 Exposure to interventions
<p>Indicators for Youth</p> <ol style="list-style-type: none"> 1 Knowledge of HIV prevention methods 2 No incorrect beliefs about AIDS transmission 3 Median age at first sex 4 Youth sexually active 5 Youth with multiple partners 6 Number of sexual partners among young people 7 Youth using a condom at last sex 8 Youth using condoms consistently 9 Youth using condom at first sex 10 Commercial sex (young men only) 11 Condom use at last commercial sex (young men only) 12 Consistent condom use during commercial sex (young men only) 13 Young people seeking voluntary HIV tests 14 Exposure to interventions 	<p>Indicators for Men Who Have Sex With Men</p> <ol style="list-style-type: none"> 1 Knowledge of HIV prevention 2 No incorrect beliefs about AIDS transmission 3 Anal sex with multiple partners in the last 6 months 4 Number of anal sex partners in the last 6 months 5 Condom use at last anal sex 6 Consistent condom use in anal sex with male partners 7 Men paying for sex with commercial male partners 8 Condom use at last anal sex with a commercial male partner 9 Consistent condom use in anal sex with commercial male partners 10 Men who have risky sex with men and women 11 Men who have sex with men seeking voluntary HIV tests 12 Exposure to interventions
<p>Indicators for Female Sex Workers</p> <ol style="list-style-type: none"> 1 Knowledge of HIV prevention methods 2 No incorrect beliefs about AIDS transmission 3 Condom use at last sex with client 4 Consistent condom use with clients in the last month 	<p>Indicators for Injecting Drug Users</p> <ol style="list-style-type: none"> 1 Knowledge of HIV prevention 2 No incorrect beliefs about AIDS transmission 3 Injecting equipment sharing at last injection 4 Injecting drug users never sharing equipment in the last month 5 Injecting drug users sharing in high equipment-sharing situations 6 Injecting drug users with access to sterile needles 7 Injecting drug users selling sex 8 Drug injectors using condoms the last time they sold sex 9 Drug injectors using condoms at last sex with a non-regular partner 10 Drug injectors using condoms at last sex with a regular partner 11 Drug injectors seeking voluntary HIV tests 12 Exposure to interventions

Details on how to measure these indicators can be found in behavioral surveillance system (BSS) guidelines, which can be found on www.fhi.org or in Mills et al., 1998.

these areas will be a matter of choice depending upon the focus of a particular program. Examples of indicators of this type are included in Table 10-1 as well.

Indicators presented in this chapter are divided according to sub-population group of interest, including adults ages 15-49, unmarried youth, female sex workers, men who have sex with men, and injecting drug users. All indicators should be measured and reported by gender, as it is widely recognized that risk behavior characteristics tend to vary greatly between men and women.

KEY INDICATORS IN ASSESSING BEHAVIORAL RISK

Indicators most relevant to programming for HIV/AIDS and sexually transmitted infections (STIs) are those related to sexual behavior, partner networking, and drug injecting. Information on partner types and multi-partner behavior (including multiple regular partners, non-regular partners, and commercial partners), the frequency of sex, partners' sexual behaviors, and the timing of multiple partnerships (concurrent or serial) are important factors to consider when attempting to understand the level of risk and vulnerability of different groups. Unfortunately, however, the choice of key indicators measuring sexual behavior can be very complex. The ideal scenario, in which information on every sexual partner and every sexual act over an extended period of time is available, is unrealistic both in terms of the capacity of human memory and the logistical limitations of quantitative surveys.

Thus, the indicators related to sexual behavior and networking presented here represent a compromise between the ideal and the practical. Indicators related to the percentage of the sub-population who are sexually active, who have more than one partner, who have one or more non-regular partners, and who have engaged in

commercial sex, provide information important to understanding sexual norms and practices within a community. The recall period for these indicators reflects practical considerations as well, defining a period long enough to allow for potential change and short enough to allow for reasonably accurate recall. This recall period varies, depending on the sub-population.

Condom Use

Following basic information on sexual partnering and networking, information on condom use is critical to appropriate STI/HIV/AIDS programming. Aside from abstinence, consistent condom use represents the only certain means of preventing the sexual transmission of HIV. Thus, tracking indicators of condom availability and use represents an important means of assessing levels of risk among individuals and communities. (See Chapter 5, "Evaluating Condom Programming," for more on this issue.) Questions concerning condom use can be posed in terms of ever use, frequency of use over varied recall periods (week, month, 6 months, year) or frequency of use during most recent sex act with particular partner types.

Each of these methods of examining condom use has strengths and weaknesses. First, while the question of ever use of condoms is very specific, it may not relate to recent changes in HIV-related behavior, and a one-time ever user of condoms is not guaranteed any assurance of reduced risk of HIV infection. In addition, the use of this indicator is becoming increasingly arcane, as availability of and familiarity with condoms increases globally. Nonetheless, it can be used as a gross measure of population behavior related to the introduction of condoms in a community where they were previously unavailable. In addition, increases in ever use of condoms without increases in their regular use may point to specific barriers to or dissatisfaction with their use.

Indicators measuring frequency of condom use over a given time period with different types of partners provide additional information regarding risk behavior. For example, identifying the percent of populations inconsistently or never using condoms can highlight vulnerable populations that may require targeted interventions. Distinct problems are associated with the use of

suggests that individuals either adopt consistent condom use or do not, and therefore, knowing the proportion of individuals who have adopted consistent condom use is an important measure for assessing the impact of prevention programs on communities and predicting future trends in the epidemic, regardless of partner change and sex act frequency.

Indicators most relevant to programming for HIV/AIDS and sexually transmitted infections (STIs) are those related to sexual behavior, partner networking, and drug injecting.

this type of indicator, however. First, categorizing condom use according to frequency is imprecise, and absolute categories such as every time (100 percent of the time) and never (0 percent of the time) may have fewer responses than such categories as almost all the time or sometimes. In an effort to provide adequate (sensitive) response categories for the reporting of condom use, four response categories are therefore suggested: every time, almost every time, sometimes, and never. The use of four categories helps to achieve an improved distribution of responses across categories.

When interpreting condom use data, it is important to also consider partner type and frequency of sex, in addition to frequency of condom use. For example, men with many non-regular partners who report inconsistent condom use are likely to be at greater risk of infection and transmission than men with only one non-regular partner who report inconsistent condom use. At the same time, however, evidence

Measuring the frequency of condom use during the most recent sex act is often used as an internal validity check to the reported frequency of condom use over a long time period. While consistent condom use may be the ultimate goal, measuring condom use for a randomly selected sexual act (most recent) may provide an indication of the general frequency of condom use across all sex acts, and be a more sensitive measure of intermediate change.

When measuring condom use, however, it is essential to measure condom use related to specific categories of partners. Research has shown that condom use varies with the perception of partner, and asking about frequency of condom use over a given time period without specifying the type of partner will increase reporting of inconsistent use. For example, a person who consistently uses condoms with commercial partners but never uses them with regular partners will report sometimes use, although the risk of

unprotected sex with these different types of partners varies greatly. Also, if some individuals are responding to last time use with a regular partner while others are responding to last time use with a non-regular partner, the understanding of the level of risk involved is obscured.

While program planners frequently wish to assess risk perception, undertaking this estimate accurately is extremely difficult.

Onset of Sexual Intercourse

Another key behavioral indicator, median age at first sexual intercourse, is often chosen to reflect changes in broad social norms. When considering use of this indicator, it is important to specify the appropriate denominator. For the indicator to accurately reflect changes in the age at first sex among a population, all members of the population must be considered. Thus, the denominator must include those who have not yet initiated sexual intercourse as well as those who have. If only sexually active persons are included in the denominator, and a significant percentage of a specific age group (for example, girls aged 15-19) have not yet initiated sexual activity, the indicator will report a skewed figure of the median age at first intercourse among those sexually active. This figure will necessarily be a lower figure than the eventual median for the age group once all members have initiated sexual intercourse. This problem can be avoided by including all

members of the population in question in the denominator.

Injecting Equipment Sharing Behaviors

In epidemics where there is a concentration of HIV infection among injecting drug users, it is important to include indicators that can measure awareness of the risk of unsafe injecting, as well as the frequency of unsafe injecting behaviors. Reduced sharing of injecting equipment and access to sterile injecting equipment are important IDU indicators. Sharing injecting equipment is both the biggest factor for HIV transmission among drug injectors and the most common focus of interventions, especially in non-industrialized countries where there is not a long history of prevention interventions among drug injectors. Measuring levels of sharing will not only serve to alert program planners about the need for interventions, but can also be used to advocate with policymakers about the need for harm reduction programs. Changes in sharing behavior will be especially valuable for tracking trends over time for programs that support needle exchange initiatives, or that work to improve easy access to safe injecting equipment.

When measuring sharing indicators, it is very important to operationalize what is meant by “sharing,” because sharing can either be active (loaning injecting equipment) or receptive (borrowing injecting equipment). Being injected by a professional injector or a dealer should also be considered as sharing.

Just as with condom use, sharing can be measured either at last injection, or in terms of frequency over time. Because injectors who are addicted tend to inject daily, the time frame for sharing indicators needs to be relatively short—no more than 1 month. Beyond that time period, recall bias becomes an obstacle to valid data. Equally important for IDU populations are those indicators that measure sexual risk-taking behaviors, because sexual partners of IDU are

at increased risk of being infected. Also, drug-taking itself may lead to increased likelihood of high-risk sexual behavior. Reducing sexual transmission should be a primary objective of IDU interventions.

Knowledge

Knowledge indicators have frequently been included in HIV/AIDS-related surveys, but while these indicators are generally useful in measuring overall awareness of HIV/AIDS information in a community, they often provide only a weak indication of risk levels within a community. An established gap exists between correct knowledge of prevention methods and the use of these prevention methods. For example, even though people know that condoms can prevent the transmission of HIV, it does not mean they correctly or consistently use condoms.

In addition, increasing levels of knowledge over time often means that knowledge indicators are not useful in interpreting risk-related change. For example, early on in the epidemic, the proportion of respondents reporting having ever heard of AIDS was often measured. Now, however, awareness is increasingly widespread and subsequently little change in this indicator can be expected over time in most populations. As a result, knowledge of the existence of AIDS is less and less frequently measured, and has been replaced in many settings by the measurement of accurate knowledge of means of transmission and prevention and attitudes concerning stigma and discrimination. These indicators have been proven to better indicate levels of risk and acceptance of people living with AIDS within a community and, thus, more useful to program evaluation and planning in the current global context of HIV/AIDS.

Another knowledge area that is often assessed is knowledge of STI symptoms because of the established link between STIs and the increased potential of HIV infection and transmission.

The higher the prevalence of unrecognized and/or untreated STIs, the higher the risk of increased transmission and infection due to the presence of STIs. For a population to seek appropriate and timely care for STIs, they must first be able to recognize that they are infected. The behavioral indicator that correlates with this knowledge area is seeking appropriate and timely treatment for STIs. While measuring this indicator is desirable, it is also problematic because the number of respondents reporting having had STI symptoms and having sought treatment is often insufficient to accurately report findings or follow trends over time.

Risk Perception

While program planners frequently wish to assess risk perception, undertaking this estimate accurately is extremely difficult. First, risk perception must be examined in the context of behavior in order to be meaningful—without an accurate idea of past and current behavior, self-perceived risk is difficult to interpret, if not meaningless. But even when assessed in the context of self-reported risk behaviors, “correct” risk perception is difficult to ascertain. For example, a survey among sex workers in Jamaica found that an equal proportion felt they were at risk and not at risk. Subsequent questions determined that of the women who reported nearly 100 percent rates of condom use, half reported that they were at high risk and therefore used condoms, while the other half reported that they were not at risk because they used condoms. In addition, risk perception depends a lot on the stage of the epidemic. In low-prevalence settings, even people with high-risk behavior may justifiably perceive that their risk is low, whereas in high-prevalence settings, people may already know they are infected (making risk perception for becoming infected irrelevant). In high-prevalence settings, despite having adopted safe sex practices at present, people may perceive that

they are at high risk of being infected because of their past behaviors. Finally, it is unclear whether the desirable outcome for this indicator is that it should be increasing or decreasing. It may increase initially, as people become more aware of the risk of HIV infection, but then may decrease again as people adopt safe-sex behaviors. Because of all these ambiguities, assessing risk perception is a complex issue that is more appropriately explored through qualitative research, including focus group discussions and in-depth interviews. (See Chapter 12, “The Role of Qualitative Data in Evaluating HIV Programs,” for more on these methodologies.)

ADDITIONAL INDICATORS

A range of contextual factors, such as discussions between regular partners about HIV/STIs, knowledge of someone infected with HIV or who has died of AIDS, regular use of alcohol or drugs, or having been tested for HIV infection, are understood to have a correlation to risk behavior change. These indicators may be included as additional indicators to allow for a multifaceted examination of determinants of and obstacles to behavior change. These factors, which relate to the context within which risk behavior decisions are made, contribute to the improved understanding of the sometimes unclear relationships between knowledge levels and change in risk behavior. Other indicators that can contribute to the understanding of HIV spread are frequent travel away from home for extended periods of time and age difference between sexual partners. These are not necessarily indicators that one expects to see a change in over time. They merely play a diagnostic role in helping explain the dynamics of the epidemic. Yet another category of indicators are those known as “overlap” indicators. An example of such an indicator would be injecting drug use

among sex workers. This type of indicator is meant to raise awareness about the synergistic effects of multiple risk behaviors, and the potential for increased spread of the virus between the IDU network and commercial sex networks. In a similar fashion, it would also be important to measure the extent of sexual risk and selling of sex by a population of injectors.

The indicators recommended here, both the primary and additional ones, are limited in their ability to act as stand-alone measures of the course of the epidemic. However, in conjunction with one another, they provide a comprehensive framework that is useful for analyzing risk within a community and comparing risk levels between communities. In the past, quantitative surveys have often been limited in generalizability and replicability because of non-standardized instruments and indicators that restrict comparisons with other data sets. In an effort to diminish the problems associated with variations among and across questionnaires, it is strongly recommended that indicators be selected and constructed in a standardized manner so that an understanding of broad trends in the epidemic can be defined and compared to other populations within the same country, across different countries, and across time.

DETERMINING THE TIME FRAME OF INDICATORS

One of the most potentially confusing aspects of indicator development is the time frame of behaviors. For example, when asking a male respondent about whether he has had sex with a sex worker, should the time frame for this behavior be the past month, the past 6 months, or the past year?

People tend to remember recent behaviors more accurately and this seems to argue for the use of shorter time frames. However, if the

behavior is not extremely frequent or common, too short a time frame will yield few respondents and make it difficult to track trends in the behavior over time with any degree of statistical confidence. Likewise, populations with very frequent risk behaviors will not remember the details of their behaviors over a long time period, such as 12 months.

The standardized set of indicators have taken these factors into account, and that is why some are over a 12-month period, while others only consider the past 6 months or the past 1 month. Consistency is critical, because changing the time frame for a behavioral indicator across survey rounds, or deviating from the standardized time frame for a given sub-population will yield dramatically differing results. While different time frames may be appropriate because of differences in local context, it should be kept in mind that comparisons of findings to other sub-populations and settings will be difficult if not impossible. Thus, time frames must be adopted that are both convenient to the respondent and analyzable to the researcher.

DEVELOPING QUESTIONNAIRES FOR HIV/AIDS BEHAVIORAL SURVEYS

Standardized questionnaires should be used for behavioral surveillance to maximize the comparability of data between survey rounds and across sub-populations and geographic regions. Small variations in wording and the order of questions can greatly affect responses to questions so that observed changes in behaviors over time may, in fact, be due to these changes as opposed to any real changes in behavior. Preparing a well-developed questionnaire that can be maintained over multiple survey rounds with minimal changes is critical, therefore.

International experience in surveying key sub-population groups has generated a wealth of

knowledge on what types of questions work and do not work when asking people about their sexual and drug-using behaviors. These questions have been brought together to form questionnaires that, in turn, are used to measure the indicators discussed in the preceding section. These standardized questionnaires have been extensively tested in international settings and are available at Family Health International’s website (www.fhi.org) as well as in published behavioral surveillance system (BSS) guidelines. The website also provides information about how and when to use these questionnaires.

Five key sub-population groups are covered by these questionnaires:

- adults 15–49 years old;
- unmarried youth;
- female sex workers;
- men who have sex with men; and
- injecting drug users.

Using standardized questionnaires has many advantages. First, questionnaire development is a difficult process, and already developed instruments such as those cited above contain formulations of questions, time references, and skip patterns that have been tested and are known to produce high-quality data. Second, because these instruments have been used in numerous settings throughout the world, their continued use will allow behavioral surveillance results to be compared internationally to determine differences in the dynamics of behavior change and the characteristics of different population groups.

It is still essential, however, to pretest and adapt survey instruments for every local setting. This involves translating the instruments into local languages and using appropriate local terminology to ensure that the original meaning of the question is not lost. It is also necessary to conduct qualitative research and involve local members of the sub-population groups who can

help interpret and adapt the questions and response categories. Back-translating the questionnaire to ensure that the translation into the local language maintains the original version's key concepts and meanings is an important additional step in ensuring the quality of the instrument.

It is also useful to develop a guide for interviewers and supervisors, which goes through the questionnaire one question at a time, explaining in full the rationale behind a question and its intended meaning. This guide can be used in training and in the field, to clarify any ambiguities or misunderstandings that may arise. An example of a supervisor/interviewer guide can be found at www.fhi.org as well as in BSS guidelines.

will derive some benefit from participating in the study. Involving the community in the planning and implementation of the study is one way to achieve this, as is the training of interviewers to ensure that respondents are informed of the purpose of the study and that their participation is requested in a factual and neutral manner. Measures taken to ensure the confidentiality of the respondent should be explained, and consent should be clearly given by the respondent before interviewing begins. In addition, the interviewer should sign the questionnaire at the time of consent in order to indicate that consent has been given. No respondent names or other identifying information should be recorded. It should be explained to respon-

Confidentiality and informed consent are important for all research subjects, but when the research involves an illegal or stigmatized activity, the importance of the privacy of the respondent is magnified.

Informed Consent and Ethical Considerations

Confidentiality and informed consent are important for all research subjects, but when the research involves an illegal or stigmatized activity such as sex work, injecting drug use, or illegal migration, the importance of the privacy of the respondent is magnified.

Thus, behavioral surveys cannot take place without the informed consent of the respondent. Special efforts must be made to ensure that potential respondents understand any risks involved in taking part in the study, and every effort must be made to ensure that the community

respondents that they have the right to refuse to take part in the study as well as to drop out of the study at any time, and should be thanked politely for their time, whether or not they choose to participate.

Questionnaire Administration and Interview Settings

In situations in which some members of the sub-population group of interest may be illiterate, data should be collected by a trained interviewer who explains questions to the respondent and records answers. It is important to use the same data collection approach with all respondents

because varying the way in which data are collected could bias results (for example, if some respondents are literate and others illiterate, an interview should be conducted with all respondents, rather than self-administered questionnaires being given to literate respondents and interviews conducted with the others). Where respondents are literate and educated (such as in student populations), respondents may record answers to questions themselves on an anonymous written questionnaire, submitting it to a data collection manager in a sealed envelope so that it cannot be distinguished from that of other respondents. It is important to ensure that the style and wording of this type of questionnaire is appropriate to self-administration, and that these questionnaires are pre-tested to ensure clarity and answerability.

When using trained interviewers to conduct a survey, it is important for them to conduct survey interviews in a setting where questions and answers cannot be overheard by others as well as to engage in a rapport-building conversation before asking survey questions. This will reduce the likelihood that respondents will give "socially desirable" answers rather than telling the truth. If a third person enters the room or is within hearing distance, the interviewer should explain that it is important to interview the respondent in privacy.

Assuring Quality Control Before and During Fieldwork

If care is taken to exercise quality control during fieldwork, two main sources of error that interfere with the ability to collect valid data will be avoided. One source comes from the people collecting data, and the other from the people from whom the data are being collected.

It is commonly said that people do not tell the truth about their sexual behaviors, and that they exaggerate, withhold information, or refuse to admit to behaviors that are culturally

unacceptable. Experience has shown that certain techniques can increase the likelihood of honest sharing of information. When interviewers are well trained to discuss sensitive behaviors with respondents and make them feel at ease, research suggests that respondents will provide truthful information. Comprehensive interviewer guidelines (such as those cited above) can contribute to the comfort levels of both interviewers and respondents when discussing sensitive topics such as sex and drug taking.

Other quality control issues relate to interactions with the groups being surveyed. Often these surveys are conducted with communities who are vulnerable and therefore reticent to open up to strangers. Working through community organizations that have relationships with the sub-populations of interest in an essential component of ensuring access to community members. In some cases, such as with men who have sex with men and injecting drug users, it is necessary to use members of the community itself (or those working with them) to do the interviewing because it cannot be expected that sufficient rapport can be built between interviewers and respondents in a short period of time. If interviewers do not come from the community, then they must be carefully chosen individuals who will not threaten respondents in any way. Concern for the privacy and confidentiality of respondents and the community must be maintained at all times and winning the trust of the community is essential to obtaining valid results.

CONCLUSION

In summary, key considerations when identifying indicators and developing questionnaires include the following:

- Indicators most relevant to STI/HIV/AIDS programming are those related to sexual

behavior, partner networking, and drug injecting. However, it has become increasingly clear that HIV cannot be considered from a static perspective in which all epidemics in all settings are considered to involve the same elements and dynamics. Therefore it is crucial to consider the stage and location of the epidemic when deciding which indicators to measure.

- When selecting indicators, consider how much a given indicator can be expected to change over time. Some knowledge indicators (or even behavioral indicators) may peak after several years and no longer indicate changes. How the information gathered through this indicator will be used in the program setting should also be considered.
- Use standardized questionnaires to increase generalizability and replicability across populations and geographic regions and pretest all questionnaires adequately and appropriately.
- Assurances of confidentiality to the respondent and community should be provided.
- All indicators should be measured and reported by gender, as it is widely recognized that risk behavior characteristics tend to vary greatly between men and women.
- Tracking indicators of condom availability and use represents an important means of assessing levels of risk among individuals and communities. When measuring condom use, it is essential that it be related to specific categories of partners, because research has shown that condom use varies with the perception of partner.
- In epidemics where there is a concentration of HIV infection among injecting drug users, it is important to include indicators that can measure awareness of the risk of unsafe injecting, as well as the frequency of unsafe injecting behaviors
- Equally important for IDU populations

are those indicators that measure sexual risk-taking behaviors, because drug-taking may be associated with increased sexual risk, and because sexual partners of IDU are at increased risk of being infected.

- Regarding timeframes for indicators, people tend to remember recent behaviors more accurately and this seems to argue for the use of shorter time frames. However, if the behavior is not extremely frequent or common, too short a time frame will yield few respondents and make it difficult to track trends in the behavior over time with any degree of statistical confidence. Likewise, populations with very frequent risk behaviors will not remember the details of their behaviors over a long time period, such as 12 months. Thus time frames must be adopted that are both convenient to the respondent and analyzable to the researcher.

RECOMMENDED READING

- Brown T, Sittitai W, Carl G, et al. Geographic and social mapping of commercial sex: a manual of procedures. Honolulu and Bangkok: Program on Population, East-West Center, and Thai Red Cross AIDS Research Center; 2000.
- Catania J, Gibson D, Marin B, et al. Response bias in assessing sexual behaviours relevant to HIV transmission. *Eval Progr Plann* 1990;13:19-29.
- Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. *Psych Bull* 1990(3);8:339-362.
- Catania JA, Moskowitz JT, Ruiz M, et al. A review of national AIDS-related behavioral surveys. *AIDS* 1996;10(Suppl A): S183-S190.
- Cleland J, Ferry B, editors. *Sexual behavior and AIDS in the developing world*. London: Taylor & Francis; 1995.
- Coates RA, Soskolne CL, Calzavara L, et al. The reliability of sexual histories in AIDS-related research: evaluation of an interview-administered questionnaire. *Can J Public Health* 1986;77(5):343-348.
- Dare OO, Cleland JG. Reliability and validity of survey data on sexual behavior. In: *AIDS impact and prevention in the developing world: demographic and social science perspectives*. *Health Transit Rev* 1994; 4(Suppl):93-110.
- Family Health International (FHI). *Guidelines for repeated behavioral surveillance surveys in populations at risk of HIV*. Arlington (VA): Family Health International; 2000.
- Family Health International/IMPACT and Joint United Nations Programme on HIV/AIDS (UNAIDS). *Meeting the behavioural data collection needs of national HIV/AIDS/STD programmes*. Proceedings from a joint IMPACT/FHI/UNAIDS workshop. Arlington (VA) and Geneva: Family Health International/IMPACT and UNAIDS; 1998.
- Figueroa JP, Brathwaite AR, Wedderburn M, et al. Is HIV/STD control in Jamaica making a difference? *AIDS* 1998;12(Suppl 2):S89-S98.
- Fishbein M, Pequegnat W. Evaluating AIDS prevention interventions using behavioral and biological outcome measures. *Sex Transm Dis* 2000;27(2):101-110.
- Hogle J, Sweat M. Qualitative methods for evaluation research in HIV/AIDS prevention programming. *AIDSCAP evaluation tools, Module 5*. Arlington, VA: Family Health International; 1996.
- Konings E, Bantebya G, Caraël M, et al. Validating population surveys for the measurement of HIV/STD prevention indicators. *AIDS* 1995;9(4):375-382.
- Lemp GF, Hirozawa A, Givertz D, et al. Seroprevalence of HIV and risk behaviors among young homosexual and bisexual men: the San Francisco/Berkeley Young Men's Survey. *JAMA* 1994 Aug;272(6):449-454.
- Mills S, Saidel T, Bennett A, et al. HIV risk behavioral surveillance: a methodology for monitoring behavioral trends. *AIDS* 1998;12(Suppl 2):S37-S46.
- Pequegnat W, Fishbein M, Celentano D, et al. NIMH/APPC workgroup on behavioral and biological outcomes in HIV/STD prevention studies: a position statement. *Sex Transm Dis* 2000;27(3):127-132.
- UNAIDS/00.17E. *National AIDS programmes: a guide to monitoring and evaluation*. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2000.
- Watters J, Biernacki P. Targeted sampling: options for the study of hidden populations. *Soc Prob* 1989;36(4):416-430.
- Zenilman JM, Weisman CS, Rompalo AM, et al. Condom use to prevent incident STDs: the validity of self-reported condom use. *Sex Transm Dis* 1995;22(1):15-21.

11

ASSESSING THE VALIDITY AND RELIABILITY OF SELF-REPORTED BEHAVIORAL DATA

MICHEL CARAËL

Improving Participation Rates	168
Methods for Checking Validity and Reliability	169
Consistency with Independent Sources	169
Internal Consistency Checks	169
Convergent Validity Checks	170
Techniques to Assess Reliability	171
Biomedical Validation	171
Conclusion	171
References	172

11

ASSESSING THE VALIDITY AND RELIABILITY OF SELF-REPORTED BEHAVIORAL DATA

The onset of the HIV epidemic has made the validity and reliability of measurements of sexual behaviors, particularly behavioral change and condom use, critical and salient. The availability of sexual behavioral data is vital to developing and monitoring appropriate prevention programs, but their effectiveness depends to a large extent on their quality.

However, validating sexual behavior data—in other words, assessing the degree to which reported behavior reflects actual behavior—is not easy. The most obvious difficulty is that direct observation or public records are virtually impossible in the context of private sexual behavior. By implication, sexual behavior data obtained only through questions represents reported, not observed behavior. Hence, there is a need for caution in using these data for policy purposes or as the basis of education programs^{1,2}.

For decades, questions on sexual activity have been avoided in most health or fertility surveys on the grounds that such questions were morally and culturally sensitive and the responses were of low reliability and validity. However, many of the problems raised about validity are not specific to sexual behaviors. Questions about child death, fertility, marital status, and income are also highly

sensitive. The extent to which individuals are willing to recall and report details of past sexual behavior may indeed greatly vary according to gender, age, social status, risk behavior and many other cultural conditions. In some communities, certain behaviors may be actively concealed because they are illegal or considered socially unacceptable, such as prostitution, homo- or bisexuality, adultery, and premarital sex. In some cultures, men may overestimate their number of sexual partners, and in most societies, women are more likely than men to underestimate their sexual activity outside marriage.

Validity is improved when the research variables to be incorporated in the questionnaire or in the interview guide are tested during in-depth interviews in which the meaning of the language used and the accuracy of recall are checked. Accuracy of recall is an important issue because

it can be influenced in a number of ways and it has been shown to vary greatly according to the emotional significance of the event. For example, a person is more likely to accurately recall an emotionally important event, such as age at first sex, than to recall details of regular visits to sex workers. Memory error is more likely to occur for frequent events than for single ones. Respondents with multiple partners may better remember their number if asked to give detailed characteristics of each partner, starting with the most recent one. The reference period should also have meaning for respondents, such as a month, a year or lifetime, rather than artificial periods, such as 3 or 6 months. In addition, a very short reference period may maximize the accuracy of responses but may not capture infrequent behaviors in a particular population.

Pilot testing of the questionnaire should also facilitate decisions about its sequence and structure. It has been shown that validity is improved when the questionnaire is structured in such a way that sensitive questions are asked only to appropriate respondents by using filters and skips, and are placed near the end of the interview.

Another issue of sexual behavior measurement error is the social desirability bias, sometimes called self-presentation bias, that can partly be addressed by careful wording of the questions, by a non-judgmental attitude of the interviewer, and by ensuring maximum privacy. However, it is also recognized that the social pressure to give desired responses may vary greatly according to the perception of the appropriate preventive behaviors. “Increased” condom use after intense prevention programs is one example.

Several cross-national standard questionnaires on sexual behavior have been developed in the past 10 years and used in various cultures that have different sensitivities to sexual activity^{3,4}. Core questions about first sex, number of sexual partners, and condom use yield answers that seem to be quite comparable across and within cultures.

Other topics, such as commercial sexual contact and risk perception, seem much more ambiguous.

IMPROVING PARTICIPATION RATES

Drawing a probability sample of potential respondents—household- or non-household-based—for surveys of sexual behavior was long considered an impossible task. However, recent experience has demonstrated that it is feasible and that participation rates in sexual behavior surveys are as high as in other surveys when several precautions are taken. These precautions also serve to increase the legitimacy of the survey. They are:

- informing the community and the selected individuals or households in advance with letters, radio messages, or community leader visits, and explaining the purpose of the survey;
- selecting interviewers in terms of age, marital status, and gender in order to better match cultural norms on sexual communication;
- training interviewers before going to the field with emphasis on rapport building, talking about sexuality, and overcoming embarrassment;
- including a provision in the survey for a large number of call-backs;
- guaranteeing anonymity and confidentiality to the respondent; and
- ensuring privacy during the interview.

In many developing countries where survey teams have used careful procedures, response rates close to 80-90 percent have been achieved⁵. These rates have been less in urban areas than in rural areas and less in high income areas than in lower. Certainly, although these high response rates are reassuring, they do not guarantee validity if participation is highly selective in terms of risk behaviors.

METHODS FOR CHECKING VALIDITY AND RELIABILITY

Several methodologies can be implemented at a minimal cost and contribute to establishing validity and reliability.

Consistency with Independent Sources

One of the objectives of most sexual behavioral surveys is to interview a representative sample of men and women. A comparison of estimates of socio-demographic characteristics derived from these surveys (for example, age/sex composition, marital status, urban/rural residence, educational level, occupation) with similar estimates obtained from independent sources of information may provide insights into their representativeness and the extent to which these findings can be extrapolated.

Observed deviations may be due to measurement error and/or sample bias. Non-response rates, including refusal rates and participation bias, are a major concern because of the stigma attached to AIDS. This non-response may simply vary according to repeated absence of families or individuals. In some instances, absence means that selected respondents refuse to participate in the survey. There is limited documented evidence that individuals with high-risk behaviors may be less likely to participate in surveys that address sexual health. However, if HIV/AIDS is explicitly the focus of the questioning, fear and denial may lead individuals with high-risk behaviors to participate less⁶.

Usually, non-response rates are not different for sexual behavior surveys than for those of other types of surveys using similar methods of sampling. However, because non-respondents in developing countries are often urban male youth and mobile people, their lack of participation may affect the validity of survey results that focus on risk behaviors. Non-household-based surveys, such as school- or workplace-based surveys, that target specific sub-population groups may lead

to increased participation if privacy and confidentiality are guaranteed, although the effects of place of interview on validity is still unknown.

When comparing socio-demographic indicators such as age/sex composition, marital status, urban/rural residence, or educational level with other data sets, other demographic indicators, such as mean or median age at first marriage, may be useful to consider. Median age at first sex and median coital frequency may be added if other data sources exist and if definitions of marriage and first sex are the same in both surveys. The same principles may, in theory, be applied to assess the aggregate validity of sexual behavior parameters, such as sexual contacts before or outside marriage, reported symptoms of sexually transmitted infections (STIs), or condom use within different partnerships.

However, the possibilities of comparison are usually quite limited because of the absence of independent sources or because of differences in timing of the survey or sampling methods. In a few instances, some data are available in family planning clinics, STI clinics, or health centers, but individuals included in household surveys are likely to be very different from those who report to clinics. Data on condom sales and distribution may be of use, at least as a plausibility check. Commercial sex workers' reports on condom use may be compared with those of clients⁷.

Internal Consistency Checks

A number of internal consistency checks may elucidate the variation of the quality of response over the recall period. These checks depend directly on the type of question asked in particular questionnaires. According to the WHO/GPA's definition of core prevention indicators, the key variables to assess should include: age at first sex; sex with a non-regular partner in the last 12 months; condom use in the last sexual intercourse with a non-regular partner; reported symptoms of STIs; and knowledge of ways to prevent HIV

transmission. Comparisons may thus include: direct versus indirect measure of these specific variables; aggregate reporting of women versus men; aggregate reporting of wives versus husbands (when possibilities exist of linking husband and wife responses). A limited number of aggregate male/female comparisons of coital frequencies and condom use have been made with data sets from five countries where the WHO/GPA protocol was used⁸. Generally, the correspondence was reasonably high, particularly over short time periods.

questions (limited to literate populations). Ideally, the comparison should be performed by repeat measures on the same sub-sample and in a short period of time.

However, in the context of sexual behavior, interpreting differences between different methods is not straightforward because of uncertainty about their relative validity. A high level of consistency between two approaches is reassuring; conversely, inconsistencies should stimulate further research on specific variables. The pattern of

Data quality from sexual behavior surveys critically depends on high standards of execution at each phase of the survey: planning, implementation, analysis, and interpretation.

Tests of plausibility should also be conducted when associations between variables are known or when different questions on the same topic are asked. For instance, coital frequency between partners usually decreases with the duration of marriage/partnership; age at first sex after age at first marriage should be uncommon; ever use of condoms versus use of condoms as a contraceptive method versus use of condoms as behavioral change due to AIDS may be compared.

Convergent Validity Checks

Convergent validity is usually used to describe the level of agreement between results provided from different methods of data collection (triangulation), such as face-to-face standard questionnaires, diaries, in-depth interviews, self-completion of questionnaires, or self-completion of sensitive

inconsistencies might also suggest which method yields more valid information⁹.

If differences in results are found between structured face-to-face questionnaires and in-depth interviews on the same variables among the same sample population, they may be explained by a variety of factors: the quality and training of interviewers; the time spent to establish rapport and trust-building between interviewer and respondent; the time spent for recall and reflection; the explanation of the meaning of the questions; the flexibility and adaptability of the questioning.

Comparison of daily diaries with retrospective reporting from interviews may provide a useful way of checking recall accuracy. However, willingness to keep sexual diaries might introduce a problem of selection bias¹⁰.

Qualitative methods usually require the “play-back” method, which means the continual interchange of ideas between investigators and informants and the verification of research hypotheses.

Techniques to Assess Reliability

Reliability usually refers to the ability of a method to give consistent results over many tests, repeated at different times. While low reliability always means low validity, high consistency of responses across several measurement may reflect a constant bias. Reliability may be directly assessed by test-retest measures, using the same instrument and method. It can also be assessed indirectly by examining internal consistency. Dare and Cleland recently reviewed five reliability studies conducted in various countries with retest subsamples of about 300 respondents at an interval of 2 to 6 weeks after the main survey⁸. Consistency varied not only according to specific variables and questions but also according to the characteristics of the study population. At the group level, however, reliability was acceptable for most of the key sexual risk behavior factors.

Biomedical Validation

Validating self-reports with biological or clinical markers of STIs is a research strategy that has been used in some studies⁴. Community surveys with appropriate probability sampling procedure and serious ethical precautions may include sampling of blood, urine, or saliva specimens for HIV and STI testing or clinical examination. Such surveys have achieved high rates of participation. Anonymity of the respondent is guaranteed but special procedures may allow researchers to link individual reporting on risk behavior to biological markers at the data processing stage. However, the major limitation is that the association between sexual partnerships, number of sexual partners, and STIs is far from straightforward. Only a limited number of validity checks may

be made using this method, such as consistent condom use versus recent STIs or no sexual activity and the presence of STIs, and even these may be confounded by other factors.

CONCLUSION

Small- or large-scale surveys have an important role to play in monitoring behavioral change over time or to evaluate program impact. However, despite a growing body of knowledge accumulated on how to conduct such surveys, validity measures are still to a great extent lacking. Obviously, data quality from sexual behavior surveys critically depends on high standards of execution at each phase of the survey: planning, implementation, analysis, and interpretation. The issue of representativeness of the study population should also be a major concern. Whether the study sample is randomly selected or of convenience, a key factor is to ensure that all reasonable efforts are made to achieve the highest level of response or representativeness, and that the effects of selection bias are documented and integrated into the analysis. This will be critically important to keep consistency between successive surveys.

Indeed, the same type of instrument, interviewer, training, fieldwork strategy, and supervision should be used in successive surveys. This requires a detailed documentation of the baseline survey, including definition of the population of interest and biases in participation. In AIDS surveys, some of the individual characteristics of most concern to the analysis tend to be those associated with unconventional lifestyles and, thus, respondents may be more difficult to reach for interview. This bias cannot be prevented by interviewing more accessible substitutes. Rather, interviewers must strive to interview every selected person. Training interviewers in nonjudgmental attitudes, careful question struc-

ture, and culturally appropriate wording are important strategies to reduce respondent self-presentation bias.

Surveys on STIs, sexual behavior, and HIV require special attention to issues of informed consent, anonymity, and confidentiality. The effect of the location of the interview (household or other settings such as clinics, workplaces, or community centers) on respondent's perception of privacy has not yet been adequately measured.

The lessons learned from previous validity and reliability checks in developing countries show that evidence is varied and that the acceptability of measurement errors may well depend

on the objectives of the surveys. For example, a constant bias over time is less important for monitoring trends than for developing an appropriate intervention or making epidemiologic forecasts. Validity checks of different kinds are needed that are compatible with time and resources pressures in order to identify errors and biases and to be certain that behavioral changes are real and not due to the unreliability of reports. Triangulation is especially needed not only to validate the accuracy of responses but also to provide a context for the interpretation of quantitative indicators.

REFERENCES

- 1 Miller HG, Turner GF, Moses LE, editors. Committee on AIDS Research and the Behavioral, Social, and Statistical Sciences, National Research Council. AIDS, the second decade. Washington (DC): National Academy Press; 1990. Appendix C, Methodological issues in AIDS surveys; p. 206-241.
- 2 Caraël M, Cleland J, Deheneffe JC, et al. Sexual behavior in developing countries: implications for HIV control. AIDS 1995(10); 9:1171-1175.
- 3 Caraël M. Sexual behavior. In: Cleland J, Ferry B, editors. Sexual behavior and AIDS in the developing world. London: Taylor & Francis; 1995.
- 4 Caraël M, Cleland J, Ingham R. Extra-marital sex: implications of survey results for STD/HIV transmission. In: AIDS prevention and impact in the developing world: demographic and social science perspectives. Health Transit Rev 1995;4(Suppl):153-173.
- 5 Ferry B, Deheneffe JC, Mamdani M, Ingham R. Characteristics of surveys and data quality. In: Cleland J, Ferry B, editors. Sexual behavior and AIDS in the developing world. London: Taylor & Francis; 1995.
- 6 Konings E, Bantebya G. Validation of the population survey protocol for the measurement of HIV/STD prevention indicators in Kampala, Uganda. Geneva: World Health Organization/Global Programme on AIDS; 1993.
- 7 Pickering H, Todd J, Dunn D, Pepin J, Wilkin A. Prostitutes and their clients: a Gambian survey. Soc Sci Med 1992;34(1):75-88.
- 8 Dare OO, Cleland JG. Reliability and validity of survey data on sexual behavior. In: AIDS impact and prevention in the developing world: demographic and social science perspectives. Health Transit Rev 1994;4(Suppl):93-110.
- 9 Konings E, Bantebya G, Caraël M, et al. Validating population surveys for the measurement of HIV/STD prevention indicators. AIDS 1995;9(4):375-382.
- 10 Hogsborg M, Aaby P. Sexual relations, condom use and perceptions of AIDS in a urban area of Guinea-Bissau with a high prevalence of HIV-2. In: Dyson T, editor. Sexual behavior and networking: anthropological and socio-cultural studies on the transmission of HIV. Liege: Ordina; 1992.

12

THE ROLE OF QUALITATIVE DATA IN EVALUATING HIV PROGRAMS

JAN HOGLE AND MICHAEL SWEAT

Defining "Qualitative"	176
Tools and Techniques for Collecting Qualitative Data	177
Participant Observation	177
Mapping	178
Individual In-depth/Key Informant Interviewing	179
Focus Group Discussions	181
Rapid Ethnographic Studies	183
Analyzing Qualitative Data	186
Using Field Notes as the Foundation	186
Turning Data Into Usable Findings	186
Using Computers to Assist in Analyzing the Content of Qualitative Data	187
Triangulation and the Role of Qualitative Research	190
Interpreting, Reporting, and Disseminating Data Results	191
Conclusion	191
References	194

12

THE ROLE OF QUALITATIVE DATA IN EVALUATING HIV PROGRAMS

The use of qualitative research methods as a component of efforts to assess and evaluate HIV prevention programming effectiveness has received much attention during the past decade. This chapter is intended for program managers who accept the idea that

qualitative evaluation research can ultimately improve the quality of programming but who know very little about the methods involved. The information here can help guide program managers as they make decisions about resource allocation and evaluation demands.

Probably the most important piece of advice to those interested in using qualitative research for evaluation purposes is to identify local social scientists skilled in qualitative methods to assist with data collection activities. Manuals cannot replace technical expertise; they can only supplement or assist. Qualitative research skills are not confined to anthropologists or sociologists. Skilled researchers may be most easily identified by contacts with local research firms, university departments, or through donor agencies who have funded past research projects. Whether the researchers are native to the country or expatriates, they need to have the necessary local experience (geographically) and language experience

to be able to function easily in short-term research environments. Conducting qualitative research is not any easier or less expensive than survey research, and the analysis effort is often more difficult and time consuming.

The chapters of this Handbook describe a variety of instruments or tools for evaluating HIV prevention programming, including standardized survey instruments for specific target groups. With qualitative evaluation research, however, it would be difficult to propose standardized tools. For example, at the beginning of an HIV epidemic and in the context of a hostile political climate in denial, what needs to be known about a target group that has not previously been reached and that is clandestine, stigmatized, and not well organized for advocacy is very different from what needs to be known in the midst of an advanced epidemic with high prevalence rates about a target group that has been receiving interventions for several years in a supportive

political climate and that is organized for advocacy. These two extremes represent a long continuum. As a result, question topic guides that are used to gather information have to be tailored specifically to the target groups involved and to the particular information needs at a given time. These guides are as varied and unique as the languages and cultures of the target groups. Qualitative research presents reality in the words of the subjects of the research, and standardization is not the objective.

This chapter begins with some definitions and then briefly describes several of the more common qualitative research techniques that have been used in the context of evaluating HIV prevention programming in the recent past. The methods or techniques are considered tools for gathering evaluative qualitative data—tools to be used by people who have technical expertise in qualitative research.

DEFINING “QUALITATIVE”

Because the word “qualitative” is an adjective, it is difficult to define without also describing the nouns the adjective frequently modifies:

- Qualitative **data** consist of words describing non-structured or semi-structured observations of reality, resulting in in-depth information of high validity about a relatively small number of cases (namely, respondents, interviewees, people) **from the point of view of the people studied.**
- Qualitative **information** provides insights into attitudes, beliefs, motives, and behaviors of target populations, including in-depth understanding about what people **think** and how they **feel**, using the **actual words** of the people being interviewed or observed.

- Qualitative **methodologies** attempt to answer the “why” questions and deal with the emotional and contextual aspects of response, adding “feel,” “texture,” and nuance to quantitative findings, again, **from the point of view of the people being studied.**
- Qualitative **research** does not attempt to generalize to larger populations because it usually involves smaller numbers of non-randomly selected respondents; however, the way that respondents are chosen could allow some degree of limited generalization.

Sometimes it is easier to define “qualitative research” by describing what it is **not**: qualitative research is any kind of research “that produces findings not arrived at by means of statistical procedures or other means of quantification”¹⁴; p.17. Some of the data may be quantified, but the analysis itself is qualitative. Or the data may be textual (qualitative) but may be coded in such a way that computerized content analysis might produce quantitative or statistical results. However, for our purposes, we use the term “qualitative” to refer to **nonmathematical analytical procedures** resulting in findings derived from data gathered by a variety of means that might include observations, interview, and reviews of documents, books, videotapes, and even data that have been quantified for other purposes, such as census data¹.

The information in this chapter focuses primarily on qualitative interviewing and observation techniques, and not on interpreting secondary data from written documents. The following descriptions of techniques (sometimes called approaches or methodologies) focus on those commonly used when designing or evaluating HIV/AIDS prevention programming.

TOOLS AND TECHNIQUES FOR COLLECTING QUALITATIVE DATA

The complementary techniques described in detail below have been used by evaluation researchers to help assess changes in sexual behavior, knowledge about HIV/AIDS and STIs, beliefs about transmission and prevention, and attitudes toward various aspects of the epidemic. The concurrent use of qualitative research to complement knowledge, attitude, behavior, and practices (KABP) surveys provides a comprehensive, in-depth understanding of the context in which specific behaviors occur. In addition, formative qualitative research conducted at the beginning of interventions has contributed to solid project design and materials development. Complementing survey data, qualitative data help to track and assess, at key times, the shifts in trends in sexual behavior that will lead in the long term to reduced HIV incidence. The protocols and instruments for assessing capacity building of non-governmental organizations (NGOs) and government institutions also incorporate qualitative data collection techniques.

Participant Observation

The major source of qualitative data is what people say in interviews or write in documents. However, major differences often exist between what people say or write and what they actually do. Direct participation in and observation of the phenomenon of interest is likely to yield more accurate data. However, in HIV/AIDS prevention, the phenomenon of interest is sexual behavior, which puts certain limits on the level of participation of the researcher or evaluator. Tracking trends in sexual behavior change is a challenging task because of the sensitive subject matter and the private nature of the settings in which such behavior occurs. There are, however, other behaviors occurring in the wider context

of sexual behavior in which a researcher can participate and observe.

Simply described, participant observation involves unstructured interviews (everyday conversation) and unstructured observation that are recorded in great detail in written field notes. Participant observation means that the researcher participates in the lives of the people under study while at the same time maintaining enough professional distance to allow adequate observation and recording of data². In evaluating HIV interventions, participant observation means immersion to some degree in the setting in which interventions occur—physical presence in the environments in which high-risk sexual behavior is likely to occur or to be initiated. Participant observation produces textual descriptions (words, paragraphs, and pages) describing the daily life and events of the study population.

Qualitative data help to track and assess the shifts in trends in sexual behavior that will lead in the long term to reduced HIV incidence.

“Classic” participant observation, as defined by anthropologists, means that the ethnographer lives and works for 6 months to a year or more in the community under study, learning the language and viewing behavior patterns over time. Realistically, such a long-term approach to evaluation research is seldom possible. Therefore, participant observation in its strict traditional

sense is not an approach commonly used in evaluating the outcome or impact of HIV prevention programming. Rapid ethnographic techniques that incorporate participant observation are often more practical and still extremely useful, particularly in formative or baseline research conducted to design intervention strategies.

Understanding the behavioral context of AIDS interventions involves focused description and systematic note-taking. To rapidly construct an accurate descriptive context, researchers must already be part of the culture, know the language, and have an established network of contacts or gatekeepers who can assist in putting together a description. Their application of the technique of participant observation at the design stage of an intervention, or even during a final evaluation,

Creating maps of geographic areas is an important complementary data collection technique frequently used at the beginning of participant observation and ethnographic studies.

thus builds on their past participatory history with the target population and geographic location in question. In short, participant observation contributes to a more accurate (qualitative) context in which one can interpret the meaning of (quantitative) indicators of program achievement.

What exactly do participant observers do? Participant observation involves attention to all the elements that are normally needed to tell a story. Those elements include the following³:

- Who is present? How can you describe them? What roles are they playing? How did they become part of the group? Who directs the group?
- What is happening? What are people doing and saying? How are they behaving? How and what are they communicating? What body language are they using?
- When does this activity occur? How is it related to other activities or events? How long does it last? What makes it the right or wrong time for this to happen?
- Where is this happening? How important is the physical surrounding? Can and does this happen elsewhere?
- Why is this happening? What precipitated this event or interaction? Are there different perspectives on what is happening? What contributes to things happening in this manner?
- How is this activity organized? How are the elements of what is happening related? What rules or norms are evident? How does this activity or group relate to other aspects of the setting?

A participant observer, then, takes notes on all of these aspects during and after time spent in a setting, and develops the notes into a coherent written record of his or her experiences participating in and observing the context under study. Participant observation, conducted in a relatively rapid fashion, is useful to evaluators in many of the locations frequented by groups targeted for AIDS prevention.

Mapping

Creating maps of geographic areas in which research is conducted is an important complementary data collection technique frequently used at the beginning of participant observation and ethnographic studies. Maps can show major action settings, social divisions of a community,

directions and distances between key sites, and natural features of the landscape⁴.

Occasionally, lists and maps of locations of key gathering points for target group members, such as brothels, bars, massage parlors, truck stops, hotels, or other locations, may exist. Sources of sexually transmitted infection (STI) diagnosis and treatment services, condom sales outlets, or community distribution points for contraceptives are additional types of locations for which data may already exist. In most settings, however, there is likely to be no geographic orientation to key points of existing or potential interaction with target group members. In those cases, implementing agencies conducting formative research before intervention design or at the beginning of implementation should create maps that identify the locations important to the intervention.

For example, in a rapid assessment of the transport industry conducted in Papua New Guinea, researchers constructed 12 maps of major ports, truck stops, and several ships' interiors. The maps provided a way to visualize the spatial arrangements where sex could be negotiated and performed. They also showed the relationship between the locations of markets, trade stores, and other current or potential outlets for condoms or venues for educational activities and the movements of the transport workers⁵.

An intervention program in Bangkok, Thailand, used mapping to identify occupational networks and their leaders, including pilot sites and numbers of workers at each site, in six districts. The mapping process required 2 months and relied on both primary and secondary data. First, aerial maps of the districts were obtained through the design office of the local district administration. The maps with blank streets and structure markings were enlarged and systematically filled out using symbols to represent places of employment, low-income residences, and popular gathering sites. The local post office staff was particularly

helpful in this phase, but ultimately, the mapping teams had to canvass the district on foot and by motorcycle taxi and other means to verify and update the map entries. The community network inventory for each district enumerated target sites and included information on number of workers per type of site⁶.

Individual In-depth/Key Informant Interviewing

Individual interviewing means that a single researcher interviews one other person, most often referred to as an "informant," which simply means "the individual who provides information." Other terms used include interviewee, respondent, source, participant, actor, consultant, or friend.

Key informants are respondents who have special knowledge, status, or access to observations denied to the researcher and who are willing to share their knowledge and skills. Because these types of informants tend to be especially observant, reflective, and articulate, they are usually consulted more than once or even regularly by the research team. Informants' abilities to describe events and actions may or may not feature analytical interpretation. In other words, some key informants may simply describe things without offering their thoughts on meaning or significance.

The process of interviewing can be highly structured, with a precoded questionnaire, or completely unstructured and open-ended. Something in between might be a one-page list of topics to be covered in an interview or a set of questions without precoded answers. Interviewers might record the interview and have a tape transcribed later, or they might take notes during the interview and then expand and formalize the notes just after the interview. Evaluative data collected through individual interviews also can be used to assess project implementers' opinions about projects' effect on communities, about the successes and challenges of implementation, about lessons learned, and about future directions

interventions could or should take. “Project implementers” can include project managers at various levels, community outreach workers, peer health educators, trainers, decision makers, community leaders, and other categories of individuals connected with an intervention. Their opinions, observations, and interpretations add a qualitative component to quantitative process data that concentrates on people trained, people educated, condoms distributed and sold, and materials produced and distributed.

Question topic guides for these kinds of individual interviews can range from the most simple set of five or six open-ended questions that might take 20 minutes to administer, to a more complex and detailed topic guide of specific implementation-related questions. A general and open-ended set of simple questions could include the following:

- How do you feel members of this community have responded to this intervention? [wording could be specifically tailored to the intervention]
- What have been the strengths of this project from your perspective?
- What have been the weaknesses of this project and how could they be addressed in future activities?
- What are the lessons to be learned from this community’s experience with the project that the implementing agency can use in future projects? How would you do things differently?
- How do you think this community would respond to a community-based condom distribution and sales program—in other words, do you think it would be a good idea to have condoms available for sale at the village level in this country? How do you think such a program should be implemented?

An end-of-project evaluation of an intervention

with rural Muslim Ugandans incorporated these five questions to collect information from key individuals who had participated as Family AIDS Workers, trainers, or coordinators in the project. Six members of the evaluation team conducted four to five interviews each during 3 weeks of collecting survey data on a sample of the population within the target area. Each interviewer made notes on the comments from respondents, then summarized the notes from interviews into one document. Then the team worked together to discuss the various responses to each question in a group analysis exercise that resulted in a single document summarizing the results. Interviewers made a special effort to preserve short direct quotes from respondents in their notes so that the exact words of individuals could complement the summary statements made about the answers to the questions. Because the topic guide was very short and specific, the interviews did not last a long time and the analysis time was minimal. The questions were open-ended, allowing respondents to interpret the meaning of the questions in their own way⁷.

In another example, as part of the final evaluation of a country program in Haiti, evaluators conducted 42 individual interviews with a variety of project implementers, including project managers, community outreach workers, peer health educators, and other fieldworkers. The interviews focused on identifying “lessons learned” in the area of behavior change communication for HIV/AIDS prevention by collecting information on experiences from the field. Researchers conducted interviews over a 2-week period in either Creole, French, or English by two communication officers. Following that period, the interviews were transcribed, translated into English, and typed into a word processor, then imported into The Ethnograph software for analysis of textual data. Experiences described by the respondents illustrating principles, trends, and innovations were coded, analyzed, and reported⁸.

Focus Group Discussions

Focus groups involve a skilled moderator or facilitator who leads an informal but structured discussion or conducts an interview with a group of 6 to 12 people. An assistant moderator takes notes, operates the tape recorder, and deals with interruptions. Sometimes there are observers. Focus groups are used to gain an in-depth, but not necessarily representative, understanding of the attitudes, beliefs, and perceptions of a specific group of people in their own language. Reports of group discussions should contain a wealth of direct quotes from the participants that illustrate the summary points presented by the researchers.

and program managers. Focus groups can offer an opportunity to probe people’s perspectives on some set of topics, rather than assuming that the researcher already knows what the appropriate categories or issues are. The group setting is particularly useful because the moderator can listen to interaction between people and observe agreement, disagreement, and clarification. Focus groups generally are not working well when the “discussion” becomes a formal series of questions by the moderator and short answers by the participants.

These group discussions can be used in a variety of ways as components of evaluation research. They are commonly used in the plan-

Focus group discussions can be used for types of research where there is a need to explore unknown aspects of experiences, feelings, and beliefs in such a way that the information is useful to researchers and program managers.

The role of the moderator is critical: He or she should attempt to maintain the group discussion among the participants, and needs to be skilled in stimulating participation, guiding discussion, and probing for detail and depth without seeming to dominate the discussion. The moderator must also be able to manage overbearing participants and draw shy participants into the conversation⁹.

Focus group discussions can be used for types of research where there is a need to explore unknown aspects of experiences, feelings, and beliefs in such a way that the information is useful to researchers

and program managers. Focus groups can offer an opportunity to probe people’s perspectives on some set of topics, rather than assuming that the researcher already knows what the appropriate categories or issues are. The group setting is particularly useful because the moderator can listen to interaction between people and observe agreement, disagreement, and clarification. Focus groups generally are not working well when the “discussion” becomes a formal series of questions by the moderator and short answers by the participants. These group discussions can be used in a variety of ways as components of evaluation research. They are commonly used in the planning stages of research and in the design of interventions. Focus group data can inform the content and language of survey questionnaires or can be used at the midpoint of a project to assess participants’ feelings about how the project is going or where things need to change. Data from focus group discussions can be very useful to assist in explaining confusing findings or further clarifying the meaning of statistical results from KABP surveys. They are useful at the end of projects to assess project implementers’ views about project successes, strengths, lessons learned, and how challenges were overcome.

Focus groups also can be used at the end of projects to assess target group members' opinions about behavior change in their communities and about response to programs. Ideally they should be used along with other techniques and sources of data for cross-validating (triangulating) information.

A key characteristic of focus group discussions is that participants are not expected to reveal personal experiences. The emphasis should be on the participants' opinions about what "people like them" are doing or might do, rather than on the participants' personal behavior¹⁰. So, for example, instead of posing a question in a group discussion with university students, such as, "How many people here have ever used a condom?" the moderator might say, "In general, how do students at this university feel about using condoms?" The focus is on the participants' opinions about how other students

Sensitive topics, such as sexuality and AIDS, can be quite difficult or relatively easy to discuss in focus groups, depending on the target group and the cultural setting.

like them feel about condom use. The discussion should reveal something about social norms at the university as well as the language students use to talk about sexual behavior and condoms.

Sensitive topics, such as sexuality and AIDS, can be quite difficult or relatively easy to discuss

in focus groups, depending on the target group and the cultural setting. In Thailand, for example, young women workers and students were more comfortable talking about sexual behavior (their peers' or their own) in individual, open-ended interviews rather than in group settings. Sex workers and military personnel, however, did not seem reluctant to discuss these sensitive issues in group settings. Other researchers in Thailand encountered no resistance at all in focus groups with males and females on the topic of male extramarital sexual behavior¹⁰. Pretesting focus group question topic guides can provide researchers with insight into how their target group members will react to certain lines of questioning.

Choosing participants can be tricky. The idea is to have a homogeneous group according to the characteristics important to the research questions. Generally age, gender, and status are key for any group. Participants should be of similar ages, the same gender, and similar status in terms of education, occupation, political status, or authority. In a focus group discussion with STI clinic staff, for example, it would be a mistake to combine counselors, nurses, and the clinic medical director in one group. If the focus group is to be with university students, it probably would not work nearly as well if the group combined sexually active individuals with students who have not yet become sexually active. A group convened to explore transport workers' attitudes about condom use would work better if men who used condoms frequently were gathered in one group and other men who never used condoms talked together in a second group. Mixing ethnic groups may or may not work well, depending on the cultural and linguistic situation. HIV serostatus may be an important characteristic, again depending on the objectives of the research.

The level of detail or instruction in the question topic guide depends on the amount of direction and information needed by the moderator. It is important to avoid recreating a survey

questionnaire in a focus group topic guide. Even though a focus group topic guide may be designed to gain an understanding of people's perceptions of their peer's sexual behavior change, the wording of the questions will not be the same as when individuals are being interviewed about their own behavior. If the moderator is familiar with the research objectives and the target population, the question guide could simply be a list of topics with reminders about when to probe and key words needed to explore particular issues. Some topic guides may be as simple as five or six or a dozen questions, while others may be much longer. Again, the length depends on the research objectives involved; pretesting will determine whether the length of time is appropriate for the participants.

Focus group data are usually recorded on audiocassette tapes, which must be transcribed word for word and sometimes translated. The assistant moderator's notes are an important complement to the transcription because they contain descriptions of nonverbal communication during the session, information on the environment in which the discussion takes place (including timing, distractions, or other descriptive information), and information about the demographic characteristics of the participants. Analyzing these data is labor intensive and demanding of the principal investigators, whose understanding of the objectives of the research drives the analysis process. Computer-assisted data analysis is helpful for organizing and retrieving textual data, but preparing the text to be used by the software and coding the data are time consuming (this issue is discussed in greater detail later in this chapter). The quality of the analysis is enhanced by the use of an "overview grid," or matrix that summarizes a set of themes broken down by group, and also by the use of a team approach to interpreting the transcripts¹¹. External sources of information on the issues involved, such as ethnographic material or

survey results, should be used to check the plausibility of the focus group findings¹².

Rapid Ethnographic Studies

"Ethnography is what those of us who are not anthropologists think anthropologists do. The popular image is captured by the vision of Margaret Mead in her tent, taking notes from the natives"¹³.

Traditional ethnography evolved over the past century as a way of describing exotic, remote, and culturally "different" people and environments. Although ethnographic research is no longer limited to "non-Western" cultures, the original process of ethnography remains that of discovering, describing, and analyzing a group's culture, way of life, or shared sense of reality within a specific geographic setting^{2,13-15}. Ethnographers' priority methodology is participant observation conducted during intensive fieldwork. The findings are then interpreted from a cultural perspective—the assumption being that all human groups evolve a culture—and written up into a book or report¹⁶.

The concept of "rapid ethnography" evolved as a result of the need to make ethnographic data available to program managers in a timely fashion^{17,18}. The methodology uses mainly qualitative (or anthropological) data-gathering techniques in a relatively short period of time (such as a few weeks) to describe a cultural setting, usually within specific demographic and geographic boundaries. In addition, an ethnographic perspective attempts to interpret the **meanings** behind observed behaviors, such as responses to disease¹⁴.

Rapid ethnographic research is particularly useful during formative evaluation stages. In Nepal, for example, a program design team drafted a plan for an HIV/AIDS prevention project focusing on at-risk behavior among commercial sex workers and their clients in the country's Terai Region, the southernmost districts

adjacent to Nepal's border with India. Because no data existed on the sex industry in Nepal's international border areas, a rapid ethnographic assessment was conducted over a 3-month period to explore the social and behavioral context of commercial sexual interaction in five border cities. The results of the study helped refine the design of the overall intervention.

The assessment used qualitative and semi-quantitative methodologies for interviews with 56 female sex workers (FSWs), 100 clients, and 80 potential AIDS educators (health workers, education professionals, NGO staff, pharmacists, and social workers). These interviews were conducted by an assessment team consisting of an expatriate anthropologist with long-term experience in Nepal, a Nepalese physician with extensive

researchers. Prostitution in Nepal varies widely by ethnicity, religion, caste, and degree of organization. Among the Badi ethnic group, prostitution is a social norm and respected profession, but among non-Badi women, persecution, transience, and isolation characterize the practice of the profession. The interviews revealed that levels of knowledge of HIV/AIDS, awareness, and access to and use of condoms differed between Badi and non-Badi sex workers, suggesting a need for different approaches to prevention.

The rapid assessment also provided better definition of the target population and helped focus the program design on a comprehensive, mutually reinforcing set of interventions (STI service improvement, condom distribution and use, targeted behavior change communication,

The rapid assessment also provided better definition of the target population and helped focus the program design on a comprehensive, mutually reinforcing set of interventions (STI service improvement, condom distribution and use, targeted behavior change communication, and outreach education).

experience in HIV/AIDS prevention programming, and male and female field interviewers.

Interviews with potential AIDS educators and clients of sex workers were relatively easy to obtain compared to interviews with the sex workers themselves. The sex industry in Nepal operated in an “underground” context; public officials underplayed the existence and size of the industry. Sex workers led a very mobile lifestyle and tended to be suspicious of

and outreach education). The research revealed that the commercial sex industry operated in remote areas as well as urban centers, and provided new insights into the organization, lifestyle, and difficulties of Nepali sex workers. The range of variation among geographic locations and ethnic groups was surprisingly extensive.

Perhaps the most important finding emerged from a mapping exercise. During interviews with transport workers, each respondent was

asked to draw on a road map of Nepal the preferred locations for commercial sex. The maps identified small rest stops along the major east-west and north-south transport routes leading from India into Nepal. Transport workers preferred remote, highway-based teashops, restaurants, and lodges as rest and entertainment sites to similar establishments in large urban locations. In Nepal, truck drivers and their assistants say they would rather rest and relax at small roadside locations where there is plenty of space to park and maintain their trucks and where police and others do not bother them as often.

Given men's preferences for remote, out-of-the-way locations and sex workers' identification with transient lifestyles, intervention planners determined that the Nepal program's geographical focus should expand beyond the major urban centers to include smaller commercial centers, transport bazaars, and small rest stops adjacent to Nepal's major highways.

As the Nepal strategy refined its focus to the nation's primary transport routes, other related factors were also considered, most importantly the initiation of new and large donor initiatives—the American Foundation for AIDS Research (amfAR) funding of 17 NGOs in Kathmandu Valley and the far eastern and western highways, and the European Community's plan to develop a public sector STI clinic targeting transport workers along the central region highway. These factors, combined with a more realistic analysis of available project financial, human, and NGO resources, refocused the program from a strategy covering the nation east-to-west, to one focused on the country's primary transport routes leading to and from India, all located in the Central Development Region—a more manageable area covering a road distance of 435 kilometers.

Targeted Intervention Research (TIR)

A specialized methodology for conducting rapid or focused ethnographic research, TIR has been used to help STI program managers understand their communities' perspectives on STIs and HIV/AIDS^{19,20}. Like other types of ethnographic research, TIR uses mainly qualitative methods, but also can include a quantitative survey of a specific population, which allows evaluators to generalize findings to some extent.

TIR is designed to be used without outside technical assistance. It begins with the formation of a technical advisory group (TAG) that includes the STI program manager, someone responsible for the provision of STI services, a local communication expert, and a local social science researcher who supervises fieldwork. Led by the STI program manager, the TAG identifies the critical programmatic questions and concerns, then designs the details of the research, refining TIR methods to suit the specific local situation. The TIR is designed to be conducted over a period of 3 months, including a 1-month start-up, 1 month of data collection, and 1 month for data analysis and write-up.

The TIR methodology attempts to build consensus among TAG members about local programmatic priorities for STI control and to obtain information about the most common illnesses affecting adults, illnesses affecting the genital area, and illnesses transmitted through sexual intercourse. These descriptions of illnesses—“explanatory models”—include information on peoples' perceptions about transmission modes, causes and symptoms along with their timing and mode of onset, severity, and appropriate treatments. Explanatory models are constructed by combining information about an illness from the descriptions of many individuals, including community members and clinic workers, thus triangulating information to increase the validity of the data.

ANALYZING QUALITATIVE DATA

“Analysis is just an old Greek word that means ‘undoing’... Analysis is no more than the undoing of something already put together, something ready to be undone... Picking out things to know and to talk about in fact is the main way in which analysis has always been done... Analysis, the undoing of things, picking things apart, is a commonplace worldly procedure of selection”²¹.

Using Field Notes as the Foundation

Qualitative data come from fieldwork in which the researcher or evaluator spends time in the setting under study—whether the setting is a program, an organization, or a community²². Such settings for HIV prevention programming might include an STI clinic, a brothel, a truck-stop, a factory, an antenatal clinic, a youth center, a women’s group, a restaurant, a hotel, a roadside cafe, a university or secondary school, a family planning clinic, a theater, or a marketplace. Within these settings, researchers might participate and observe, talk to people or with groups, take voluminous notes, record conversations or discussions, review documents and records, or even videotape activities.

Field notes are to qualitative research what a filled-in set of questionnaires is to survey research. Field notes—description and direct quotations—are the source of the data that support the conclusions and interpretations that will ultimately be presented by the researchers. Even when using a tape recorder, the interviewer must take notes of key phrases, major points made by the respondent, and key words shown in quotation marks that capture the speaker’s own language used. If, for some reason, a tape recorder cannot be used or if it malfunctions, it is even more critical to take detailed notes, especially of actual quotations as often as possible. It is perfectly acceptable to say to a respondent, “Could you repeat for me exactly what you said so I can write it down word

for word. I don’t want to lose that particular quote. Let me read back to you what I have written to make sure it is exactly what you said.” Capturing what people say in their own words is the most important contribution of qualitative research to understanding human behavior.

The raw data of the field notes must then be reduced in quantity to a readable, narrative description organized into major themes and categories and illustrated with case examples and direct quotations from respondents. A summary of the results, either alone or in combination with quantitative results, should be presented to project stakeholders, disseminated with an emphasis on lessons learned, and referenced again during the design of future interventions. The remainder of this section describes the process used by researchers to distill raw field notes into usable findings. Increasingly, computers are used to help with this task and several important considerations in their use also are discussed.

Turning Data Into Usable Findings

Analyzing qualitative evaluation data should proceed under the direction of someone trained in qualitative data management. One good definition of “analysis” is Miles and Huberman’s, which describes the process as “consisting of three concurrent flows of activity: data reduction, data display, and conclusion drawing/verification”²³; p.10.

- **Data reduction** means summarizing or coding large amounts of text into smaller amounts of text, and it occurs continuously throughout the evaluation research process. It involves selecting, focusing, simplifying, abstracting, and transforming the raw data of field notes or transcriptions into typed summaries organized around themes or patterns based on the original objectives of the evaluation research. Data reduction continues to occur until a final report is written.

- **Data display** is defined as “an organized assembly of information” that allows conclusions to be drawn and actions to be taken²³; p.11. Most frequently, qualitative data are displayed as narrative text, which tends to overload peoples’ information-processing capabilities. Matrices, graphs, networks, and charts can present information in compact forms that make the data accessible to program managers.
- **Conclusion drawing/verification** refers to the process of deciding what things mean, noting themes, regularities, patterns, and explanations. Researchers will draw conclusions throughout the entire data collection exercise, but eventually this process becomes more explicit and firm at the point when the final report is written. Conclusions must also be verified as analysis proceeds. As researchers try to explain what the data mean, they should continually examine their explanations for plausibility and validity—do their explanations make sense within the context of the study? In evaluating HIV prevention programming, it might be possible to test the validity of conclusions during presentations of preliminary research findings and interpretations to project stakeholders and/or members of target populations. Researchers could incorporate this feedback into a final version of the written report.

During the conclusion drawing phase, the principal investigator conducting qualitative research has a very different role than he or she does in survey research. As Knodel explains:

“In survey research, the coding of questionnaires is a routine matter of data processing to be relegated to assistants, and tabulations can be run by a programmer. In contrast, in focus-group research, code-mapping and overview grid construction are

essential parts of the analysis process itself. It is exactly through these time-consuming and somewhat tedious tasks that the researcher comes to understand what the data are revealing. They can only be relegated to an assistant at the cost of detracting from the quality of the analysis...”¹²; p.103.

In developing a budget for focus group discussions, therefore, a sufficient amount of time must be included for analysis by the **principal investigator**.

These three components of analysis—data reduction, data display, and conclusion drawing/verification—need thorough documentation in the final reports of evaluation research so that readers can clearly understand how researchers arrived at certain conclusions. Only by understanding just what is done during data analysis of qualitative data can researchers delineate more reproducible methods.

Capturing what people say in their own words is the most important contribution of qualitative research to understanding human behavior.

Using Computers to Assist in Analyzing the Content of Qualitative Data

The biases and subjectivity inherent in qualitative data analysis, compounded by humans’ limited ability to process large amounts of information, can be reduced somewhat by using computer software to assist in organizing, sorting, and categorizing textual data for a

content analysis. These textual data can include literature, transcripts of discussion groups, notes from interviews, official reports, or any material that can be typed as text. People often associate computerized data analysis with quantitative or statistical approaches. This makes sense, given that computers are needed to efficiently process the calculations required for many statistical procedures and for large sample sizes. However, computers are also valuable tools for analyzing qualitative data. As with statistical analysis, rigorous qualitative analysis can be time consuming. Many of the basic procedures can be streamlined and enhanced through the appropriate use of computers and software designed for content analysis. Box 12-1 provides some suggestions on choosing a software package.

It is important to recognize that using a computer to conduct a content analysis of qualitative data does not somehow render the results “quantitative” or change the basic processes of data interpretation. The computer simply assists the researcher in coding and analyzing the data. Coding data refers to the process of associating a code word with pieces of the text that represent a certain theme. This allows the analyst to search for the codes rather than the words in the text. Using a computer to assist in this way with analysis still requires skill in qualitative data analysis and some expertise in the subject area. Moreover, planning the process is crucial: The final analysis will only be as sound as the logic that grounds the research. Computers cannot interpret the findings. As with statistical analysis, it is incumbent on the researcher to carefully interpret the findings, place the analysis in context, and relate findings to theory.

Although computer-organized text may improve the reliability and validity of the analysis, there can be trade-offs in terms of the time and effort required to complete the analysis and final reports. Evaluation activities usually need to produce results quickly to meet various deadlines, so

researchers should make a realistic assessment of timing and budget resources before using computer software in qualitative data analysis. The following sections review some of the basic issues involved in using computers to analyze qualitative data.

Advantages and Disadvantages of Using Computers for Content Analysis

Using computers to assist in analyzing qualitative data is advantageous because it allows evaluators to:

- be very systematic with the data analysis;
- search data according to predefined codes and combinations of codes;
- use the coding from a second person who analyzed the data to identify biases in coding; and
- manage large amounts of text and associated codings.

Using computers to analyze qualitative data also has some disadvantages because:

- it is time consuming to set up and code the data;
- it may be necessary for evaluators to learn to use a computer and software;
- the computer hardware can malfunction;
- data can be lost due to computer failures;
- it can be time consuming to use a computer; and
- the computer and software are expensive.

Most of the hardware and software disadvantages can be easily dealt with through planning, training, and patience. Disadvantages related to cost and time must be weighed against the advantages of using computers for analysis. For many small projects, especially if the computer approach will never be used again, it is best to conduct an analysis manually.

Steps Involved in Using a Computer to Code and Analyze Data

Several computer programs are available for use in coding and analyzing data, and all are different. However, using most of the commercially available software packages to assist in conducting content analysis involves a few basic steps. These include:

- transcription (and sometimes translation);
- formatting;
- coding data; and
- interpreting data.

Once the first piece of data has been collected (one group discussion or a single interview), it can be transcribed into computer-readable format (usually ASCII) to be used by the computer program. Occasionally, it will be necessary to translate the text into another language. Translation should be undertaken carefully—normally with the text translated twice for comparison, or back translated (language to English, French, or Spanish, and back to language) to detect differences in translations. The data are then formatted to be used with the particular software program. Next, the data are systematically coded using a predetermined coding scheme. Note that the coding scheme can, and should, be modified as the coding progresses. Normally, the evaluation team starts with a basic set of issues to code and, as the process of reading the text progresses, they will add new codes to the list based on their reading. Once the data are coded, the software are used together with the coded data to produce interpretations and conclusions.

It is probably best to have two people code the data, using the same set of code words, and comparing the findings for discrepancies. When discrepancies are found, a system should be in place for resolving them—such as having a third person also read and code the text.

Controlling for Bias in the Data or in Interpreting Data

It is important that findings not be biased by the personal opinions of the researcher. Rather,

Choosing a software package

BOX 12-1

Weitzman and Miles reviewed 22 different software programs for qualitative data analysis. These programs cover a wide range of analysis needs and computer operating systems. They emphasize that computer software is a fast-moving field and that reviews of programs rapidly become out of date as programs are refined and new ones appear²⁴.

As always, the objectives of the research and the level of analysis needed should dictate the analysts’ choice of software. Evaluation researchers may find that “code-and-retrieve” programs are sufficient for the level of analysis required for project evaluation data. These software programs were developed specifically by qualitative researchers and help the analyst divide text into segments or chunks, attach codes to the chunks, and find and display all instances of coded chunks (or combinations of coded chunks)²⁵. Examples of code-and-retrieve programs include ATLAS/ti, HyperQual, Kwalitan, MAX, NUDIST, QUALPRO, and The Ethnograph.

A researcher’s final choice of a software program depends on his or her computer’s operating system (Macintosh or Windows) as well as on the analyst’s familiarity with the various programs. The analysis will proceed more quickly if the investigators are using a program they already know well.

research findings should be based on an objective interpretation of the data. However, because so much is open to interpretation, it can be difficult to analyze qualitative data objectively. Using computers to assist in analyzing qualitative data can help reduce bias and identify bias when it does occur. If a systematic plan of analysis is followed, there is less chance for bias in interpreting results. For example, an analyst who finds that a text has few coded responses to substantiate a particular interpretation of the content should wonder whether the coding is truly unbiased. Or, if a second person who codes the same data has different coding patterns (in other words, codes the same passages differently), the analyst should be especially careful in interpreting those particular findings.

- **data** triangulation—using several data sources in a study;
- **investigator** triangulation—using several different researchers or evaluators;
- **theory** triangulation—using multiple perspectives to interpret a single set of data; and
- **methodological** triangulation—using multiple methods to study a single problem or program.

While triangulation is ideal, it is also expensive and poses problems for evaluators' limited budgets, short time frames, and political realities. Nevertheless, most evaluation experts agree that triangulation greatly reduces systematic bias in the data. "Triangulation is a process by

"Triangulation is a process by which the researcher can guard against the accusation that a study's findings are simply an artifact of a single method, a single source, or a single investigator's biases."

Patton, 1990

TRIANGULATION AND THE ROLE OF QUALITATIVE RESEARCH

The concept of triangulation is taken from land surveying—a person being able to identify his or her location on a map by taking bearings on two different landmarks and determining their intersection. The two landmarks and the surveyor become the three points of a triangle. In evaluation research, triangulation can be of four different types²²:

which the researcher can guard against the accusation that a study's findings are simply an artifact of a single method, a single source, or a single investigator's biases"²²; p.470.

In assessing and measuring behavior change resulting from interventions, KABP survey data and qualitative focus group data have most commonly been used to compare and contrast perceptions of trends in behavior change as program managers face the task of evaluating their program achievements. Assessing experi-

ences from the field—"lessons learned"—can involve group interviews with peer health educators and individual interviews with project managers. These qualitative sources of information, contrasted and compared with survey data on common behavioral indicators, provide multiple methods as well as multiple interpreters as multidisciplinary teams prepare final project reports.

INTERPRETING, REPORTING, AND DISSEMINATING DATA RESULTS

Because the purpose of evaluation research is to make a difference in decision making and programmatic action, a researcher is obliged to report and present the results of his or her data reduction efforts and to share interpretations of the findings with project stakeholders. Boxes 12-2 and 12-3 provide guidance on writing qualitative research reports. Ideally, the evaluator provides stakeholders with description and analysis, verbally and informally, and includes the stakeholders' reactions as part of the data in the written report. Stakeholders expect evaluators to "confirm what they know that is supported by data, to disabuse them of misconceptions, and to illuminate important things that they didn't know but should know"²²; p.423.

The definition of the word "interpret" is "to clarify the meaning of by explaining or restating." This means that the researcher "clarifies the meaning" of the observed behavior or descriptions (data) by restating the observations made or the things that respondents have told the interviewer. The evaluation researcher uses his or her training and expertise to explain and clarify the data for people who have not participated in the data collection experience. "Interpretation" is really an art that requires sophistication, maturity, and wisdom. "Interpretation" means **explaining** what is happening, whether the events are observed first-hand or described by someone else. Interpretation is part of analysis and goes

beyond data description and data reduction.

When the evaluation researcher explains and interprets the meaning of information, he or she is doing so from a certain personal perspective and in a certain context. That perspective and its context should be identified at the beginning of the written report in order to enhance the readers' abilities to understand the research results.

CONCLUSION

During the past decade, much attention has been devoted to the use of qualitative research methods as a component of efforts to assess and evaluate the effectiveness of HIV programs. Today, there is consensus that qualitative data and research methodologies are a valuable complement to quantitative data and research techniques, and in fact, that these research approaches should be used in combination to assess the progress and effects of prevention and care interventions.

Several methodologies are commonly used in qualitative research, including participant observation, mapping, in-depth interviews, focus group discussions, and rapid ethnographic studies. These approaches are useful in all types of evaluation research, from formative evaluation that is used in planning and designing interventions, to process evaluation that tracks how well interventions are carried out, to effectiveness evaluation that assesses the impact and outcomes of interventions.

By examining issues from the point of view of the people being studied, qualitative research provides context and meaning, often answering the "why" question of behavior and barriers to behavior change. Answers to these questions permit programs to develop more effective communication messages and design better services that respond to target groups' barriers and needs.

Vignettes and the real-life experiences of individuals captured in qualitative data collection

BOX 12-2

Tips on writing qualitative research results

- 1 Keep the research focus in mind at all times, but be skeptical about it. Keep an open mind to the possibility that you are not quite “on the mark.” Use as a guiding question, “What is this really a study of?” It is definitely not sufficient to say that the purpose of focus groups was to complement survey data, and it is probably not sufficient to say that the purpose of interviews was to assess changes in sexual behavior. Objectives of evaluation research need to be stated in more specific terms.
- 2 Get rid of data as the focus of the paper becomes progressively sharper. Indicate in notes or asides what is important and interesting but can be dealt with in another article or paper. Evaluation research generates a great deal of data covering many more areas than just the core indicators of knowledge of prevention methods, numbers of partners, and condom use with nonregular partners. Some of the domains of information (such as intermediate steps on the road from knowledge to behavior change) are important to include in the evaluation research report. Other areas may be best addressed in separate papers or appendixes.
- 3 Do not get stuck because of non-existent data or elements that are unclear or cannot be interpreted. Indicate this and move on. It is always fine to identify areas that need further research.
- 4 Use concrete examples and direct quotes as often as possible. Do not include generalized statements such as, “Knowledge of HIV prevention is high among members of this target group.” Knowledge about what aspects of prevention?
- 5 Write for an audience that does not already know about the topic and who may not know much about the country and culture. Write for an international audience with the longer-term objective of sharing your findings with the international HIV/AIDS/STI prevention community.

can serve as an effective evaluation tool for informing the decisions of program managers and policymakers. Far from the often dry statistics of quantitative data, the personal details of human experience can captivate an audience so that they can better understand and grasp the effects of a program on people.

Because the raw data of qualitative research consist of words, thoughts, opinions, quotes, and observations, the process of content analysis is often a labor-intensive and time-consuming activity. Computer software can be helpful in organizing the raw data and streamlining and enhancing the analytic process, but it has limitations. It is incumbent on the researchers to carefully interpret the findings, place the analysis in context, relate the findings to theory, and clearly present the results to stakeholders. The process of analysis and reporting must be closely guided by principal investigators who are trained in qualitative data management and experts in the subject matter.

**Suggested outline for a qualitative research report
(focus group discussions, key informant interviews, rapid ethnographic study)**

BOX 12-3

Section	What to Include
Report cover	<ul style="list-style-type: none"> • title of report, date, donor logos, contract numbers
Title page	<ul style="list-style-type: none"> • title of project, contract number, country name, authors of report, implementing agency name(s), donor logos, date of report
Executive summary	<ul style="list-style-type: none"> • summarize objectives, date of data collection, target groups; describe how respondents were chosen to be interviewed, geographic location of data collection; summary of procedures followed for data collection • include bulleted summary of results and recommendations
Table of contents	<ul style="list-style-type: none"> • with page numbers
List of tables/figures	<ul style="list-style-type: none"> • with page numbers
1 Background	<ul style="list-style-type: none"> • describe role of the qualitative research in sub-project implementation, relationship to other data collection methods being used; concisely describe the context in which the research took place; include map of the relevant geographic area; describe objectives/purpose of the research
2 Methods	<ul style="list-style-type: none"> • briefly describe the details of the research (number of focus group discussions, key informant interviews and/or observations, maps produced, methods of selecting participants/respondents location(s), logistics, dates, and other details of data collection) • present demographics of participants/respondents; provide details on how respondents were chosen • describe interviewer choice and training and length of time of data collection • describe quality assurance methods used and problems encountered in conducting research • discuss potential biases; how analysis was done (translations, transcription, use of computer software)
3 Research findings	<ul style="list-style-type: none"> • present data organized around key questions or main ideas in the interview or from the topic discussion guides; combine raw data with descriptive summaries; use plenty of quotations from transcripts
4 Discussion and interpretation	<ul style="list-style-type: none"> • describe findings and their meaning in the context of that project, with highlights of unexpected findings; discuss potential problems with the data • compare findings to other research, especially quantitative studies that may be closely linked to the qualitative research
5 Recommendations	<ul style="list-style-type: none"> • provide bulleted list with recommendations for project implementation, designing educational materials, reaching target members, policy implications, redesign of instruments for future research
6 Dissemination plan	<ul style="list-style-type: none"> • describe plans for oral presentations to stakeholders, presentations at international meetings and/or publications
Appendixes	<ul style="list-style-type: none"> • include—in English and in language of administration—copies of instruments, such as moderator’s topic guides (for focus group discussions) or question topic guides (for key informant interviews) • Include screening questionnaire (if used); other instruments (e.g., survey instrument that might be used during a rapid ethnographic assessment); additional quotations

Source: Modified from Reference 16

REFERENCES

- 1 Strauss A and Corbin J. Basics of qualitative research: grounded theory procedures and techniques. Newbury Park (CA): Sage Publications; 1990.
- 2 Fetterman DM. Ethnography step by step. Vol. 17, Applied Social Research Methods Series. Newbury Park (CA): Sage Publications; 1989.
- 3 Bogdewic S. Participant observation. In: Crabtree BF, Miller WL, editors. Doing qualitative research. Vol. 3, Research Methods for Primary Care. Newbury Park (CA): Sage Publications; 1992. p. 45-69.
- 4 Pelto PJ, Pelto GH. Anthropological research: the structure of inquiry. London: Cambridge University Press; 1978.
- 5 Jenkins C. Final report: behavioral risk assessment for HIV/AIDS among workers in the transport industry, Papua New Guinea. PNG Institute of Medical Research, Goroka. Arlington (VA): Family Health International; 1994.
- 6 Supammatas S. Community mobilization working paper #1, the community network diagnosis. Faculty of Public Health, Mahidol University. Durham (NC): Family Health International/AIDSCAP; 1993.
- 7 Kagimu M, Marum E, Wabwire-Mangen E, et al. 1995 Family AIDS education and prevention through Imams (FAEPTI) project, follow-up evaluation report. Islamic Medical Association of Uganda (IMAU). Kampala: World Learning Inc.; March 1995.
- 8 Damier E, Mahler H. BCC experiences from the field in Haiti. Arlington (VA): Family Health International; 1996.
- 9 Shedlin MG, Schreiber GM. Using focus groups in drug abuse and HIV/AIDS research. Paper prepared for the National Institute on Drug Abuse (NIDA) Technical Review: qualitative methods in drug abuse and HIV research. Rockville (MD): NIDA; 1994.
- 10 VanLandingham M, Knodel J, Saengtienchai C, Pramualratana A. Aren't sexual issues supposed to be sensitive? Health Transit Rev 1994;4(1):85-90.
- 11 Jato M, van der Straten A, Kumah OM, Tsitsol L. Using focus-group discussions to explore the role of women's groups (tontines) in family-planning information dissemination in Yaoundé, Cameroon. Health Transit Rev 1994;4(1):90-95.
- 12 Knodel J. Conducting comparative focus-group research: cautionary comments from a coordinator. Health Transit Rev 1994;4(1):99-104.
- 13 Gilchrist VJ. Key informant interviews. In: Crabtree BF, Miller WL, editors. Doing qualitative research. Vol. 3, Research methods for primary care. Newbury Park (CA): Sage Publications; 1992. p. 70-89.
- 14 Schensul JJ, Schensul SL. Ethnographic evaluation of AIDS prevention programs: better data for better programs. In: Leviton LC, Hegedus AM, Kubrin, editors. Evaluating AIDS prevention: contributions of multiple disciplines. San Francisco: Jossey-Bass; 1990. p. 51-62.
- 15 Vidich AJ, Lyman SM. Qualitative methods: their history in sociology and anthropology. In: Denzin NC, Lincoln YS, editors. Handbook of qualitative research. Thousand Oaks (CA): Sage Publications; 1994. p. 23-59.
- 16 Wolcott HF. Writing up qualitative research. Thousand Oaks (CA): Sage Publications; 1990.
- 17 Bentley ME, Pelto GH, Straus WL, et al. Rapid ethnographic assessment: applications in a diarrhea management program. Soc Sci Med 1988; 27(1):107-116.
- 18 Scrimshaw SCM, Carballo M, Ramos L, Blair BA. The AIDS rapid anthropological assessment procedures: a tool for health education planning and evaluation. Health Educ Q 1991;18(1):111-123.
- 19 Helitzer-Allen DL, Allen HA. The manual for targeted intervention research (TIR) on sexually transmitted illnesses with community members. Arlington (VA): Family Health International/AIDSCAP; 1994.
- 20 Helitzer-Allen DL, Allen H, Field ML, Dallabetta G. Targeted intervention research on sexually transmitted illnesses. Pract Anthropol 1996;18(3):20-23.
- 21 Rose E. The first book of the werald: the world undone. Boulder (CO): The Waiting Room Press; 1991. Cited in: Seidel J, Friese S, Leonard DC. The Ethnograph v4.0: a user's guide. Amherst (MA): Qualis Research Associates; 1995.
- 22 Patton MQ, editor. Qualitative evaluation and research methods. 2nd edition. Newbury Park (CA): Sage Publications; 1990.
- 23 Miles MB, Huberman AM. Qualitative data analysis: an expanded sourcebook. 2nd edition. Thousand Oaks (CA): Sage Publications; 1994.
- 24 Weitzman EA, Miles MB. Computer programs for qualitative data analysis. Thousand Oaks (CA): Sage Publications; 1994.
- 25 Miles MB, Weitzman EA. Choosing computer programs for qualitative data analysis. In: Miles MB, Huberman AM. Qualitative data analysis: an expanded sourcebook. 2nd edition. Thousand Oaks (CA): Sage Publications; 1994. p. 311-317.

13

EFFECTIVE DISSEMINATION OF DATA COLLECTION RESULTS

TIM BROWN AND JOHN STOVER

Audiences for Survey Results	198
Possible Forms of Dissemination	198
Tailoring Dissemination to Specific Audiences	200
Conclusion	202

13

EFFECTIVE DISSEMINATION OF DATA COLLECTION RESULTS

An evaluation or surveillance system can become a meaningless exercise in data collection unless the findings motivate key stakeholders, including policymakers, program managers, and the surveyed communities, to take action to reduce HIV transmission. Thus, broad

dissemination of the findings in formats that promote responses is an essential component of any data collection system. However, dissemination is often treated as an afterthought, receiving only limited attention after the results have been analyzed and formal reports generated. Some people even consider their dissemination work complete with the publication of a detailed final technical report. However, if dissemination is to produce necessary actions on the part of the key stakeholders, it will involve much, much more. Indeed, effective dissemination is an ongoing process that begins when a survey or evaluation project is formulated—namely, at the proposal and development stage—and continues throughout the life of the project.

The elements of this process include:

- Building consensus among various stakeholders about the communities to be sur-

veyed, the data to be collected and disseminated, and the forms of dissemination.

This helps to build a sense of ownership of the findings and ensures that their presentation is appropriate and relevant for the various target audiences.

- Developing a complete dissemination strategy at the time that the project is planned. This strategy should include dissemination of some key findings as soon as possible after data collection is completed to sustain interest and speed the implementation of prevention activities.
- Preparing target audiences to understand the meaning, limitations, and interpretation of results well in advance of their actual release to these audiences.
- Developing separate dissemination materials for each target audience that explain the findings in clear and simple language they

can understand and that avoid discipline-specific language that has no meaning to that specific target audience.

- Actively following up with target audiences to answer questions, clarify meanings and interpretations, and make informed recommendations on appropriate policies, programs, or actions as suggested by the data.

AUDIENCES FOR SURVEY RESULTS

The persons to whom the results should be disseminated will vary from country to country depending upon local conditions, influential groups, and social and political structures. Possible recipients might include:

- policymakers, including senior officials in the office of the President, the Ministry of Health, other Ministries and Parliamentarians;
- program managers in government agencies, non-governmental organizations (NGOs), and community-based organizations (CBOs);
- members and leaders of the communities surveyed;
- mass media;
- donors and other funding agencies;
- the general public;
- universities and research institutions;
- private employers;
- labor unions;
- religious organizations;
- military;
- professional organizations; and
- provincial and district officials.

All other things being equal, the broadest dissemination of results is preferred. Each of the potential target audiences listed above has a role in reducing HIV transmission and can take action based on the findings. However, it must

also be remembered that there are often local concerns and sensitivities about the open discussion of the behaviors that transmit HIV. These sensitivities are often strongest among key government officials or high-ranking religious leaders, whose alienation might seriously impede prevention efforts. In designing dissemination activities and materials, careful attention should be paid to these concerns, especially those of key gatekeepers. Whenever possible, the data should be disseminated in forms that respect the concerns of the target audience, involve them in the dissemination process, and present the results in language that will not be found objectionable. Indirect avenues for dissemination, such as explaining the significance of the findings to those closest to influential policymakers and religious leaders in hopes that they may assist in reaching their colleagues, can also be explored.

Because members of the communities under evaluation are actively involved in data collection and have the ability to respond to the findings by changing behavior, those who carry out surveys have a special obligation to ensure that the leaders and other members of these communities are informed of the findings. Their active participation throughout the entire process of design, implementation, and dissemination will strengthen the system and help to ensure its reliability and validity. The remainder of this chapter provides concrete suggestions for implementers on how to effectively disseminate the results of their surveys so as to ensure that key stakeholders are involved and informed.

POSSIBLE FORMS OF DISSEMINATION

Those implementing surveys or any evaluation project cannot assume that they have finished their work when they produce a final report with detailed statistical analysis. Though a

technical report is necessary, most people do not know the meaning of “p-value” or “confidence interval” and such language is likely to confuse the issues in their minds rather than clarify them. To reach all potential target audiences with the appropriate messages, a number of forms of dissemination will usually be required, including:

- **A detailed report with complete statistical analysis**—This report serves as the technical foundation for preparing other dissemination materials and may be appropriate for distribution to program managers and staff of organizations actively working with the communities surveyed. Even within such a report, project staff should highlight and summarize significant results in clear, non-technical language at regular intervals.
- **Briefing materials for the press or other mass media**—These materials should avoid all technical language. Each press release or briefing should focus on only one or two key findings and their implications so as not to confuse the reader.

reaching policymakers and the public, and staff should plan the content accordingly.

- **Short one- and two-page policy briefs**—Because policymakers and leaders of the surveyed communities are positioned to have a major impact on prevention measures, materials specifically targeted at them are essential. Each policy brief should focus on one or two key aspects of the findings, discuss implications for their own activities, and provide recommendations regarding actions they might take to influence prevention activities positively. Separate briefs will often be necessary for different groups of policymakers or community leaders so the materials can be made directly relevant to them.
- **Group dissemination meetings or presentations**—Larger meetings or presentations can offer an opportunity to present the findings to members of the surveyed communities, large groups of policymakers or program managers, or the general public. The full process of dissemination may involve such meetings before, during, and after the release of the data. Meetings before

To reach all potential target audiences with the appropriate messages, a number of forms of dissemination will usually be required.

Whenever possible, staff should provide a written summary of remarks when discussing findings with broadcast media. This will help to reduce misquoting and misinterpretation. It is important to remember that the press is also an indirect way of

can be used to explain the data that are being collected, discuss issues of its interpretation, and prepare people to accept it. Meetings during and after offer an opportunity to present the findings, answer questions, clarify any

misunderstandings, and discuss the implications for people in light of their respective responsibilities and the activities they can influence.

- **Individualized meetings**—Individual personalized meetings with influential policymakers, community leaders from the surveyed groups, or interested NGOs and government program managers offer them an opportunity to obtain and process the results and explore the implications for their own work in a non-threatening environment. Because such meetings are time and preparation intensive, staff should pay careful attention to who can act effectively on the information when they are identifying the recipients of such individualized attention.

TAILORING DISSEMINATION TO SPECIFIC AUDIENCES

When material is presented to the mass media, policymakers, and the public, it must be understandable to the target audience and presented in a form that allows them to interpret figures accurately. A number of suggestions can be made for improving dissemination materials:

- Avoid use of discipline-specific language, such as statistical terms (“chi-squared” or “statistically significant”) or social science terminology (“non-regular partner” or “anogenital intercourse”). These terms are not well understood by the target audience or are likely to be misinterpreted.

The importance of dissemination should not be underestimated.

With effective dissemination, survey and other evaluation data can become an important component of advocating for expanded action and motivating a broader societal response to HIV prevention.

In preparing materials for each target audience, the basic goal of dissemination to each possible audience is to give them the information they need to encourage, target, or carry out meaningful prevention activities. Phrase the results presented in action-oriented or policy-oriented terms. In the materials explain the relevance and importance of this information to the target audience in their specific context.

- In material prepared for policymakers and community leaders, include recommendations for action in their spheres of influence. For example, materials for the Ministry of Education might emphasize the need for school-based sexual health education at early secondary level based on the number of young people reporting an early age at first intercourse. To maintain maximum credibility, any recommendations

given in policy briefs should follow from the data findings themselves.

- In material prepared for other audiences, discuss the implications of the findings in the recipients’ own situations. For example, a presentation about high levels of risk among factory workers might emphasize the need to increase condom use, or a presentation on high levels of male premarital sex might discuss implications for HIV exposure of young women about to marry.
- Always present data with appropriate age, gender, and other breakdowns to help people understand the meaning of the results. Because levels of risk behavior change from one population to the next, vary between men and women, or change as people go through life, it is important that presentations take these factors into account in displaying the findings. Far too often, such results have been presented with no sub-population, gender, or age breakdowns. This makes it difficult for people to see important patterns, such as high levels of risk among adolescents.
- Highlight items on which the target audience can act in the presentations of the data. Because surveys often look for trends, some people mistakenly assume the lack of any clear trend is not interesting, but this is frequently not the case. For example, a low but stable level of condom use among casual sexual partners can be a catalyst for mobilizing efforts at condom promotion in the surveyed community.
- Present data regarding marginalized communities, such as sex worker or men who have sex with men, carefully and with sensitivity to the concerns of these communities. Plan carefully and present the data in a way that does not increase stigmatization or discrimination.

- Present information that shows the audience why they should care about AIDS. A presentation for the Ministry of Finance or Planning should show that it will be very difficult to achieve economic development goals without addressing AIDS. For health specialists, the presentation should show how AIDS affects the achievement of goals for child survival and life expectancy. For labor unions, the presentation should show how union members and their families are affected by AIDS. Relating AIDS to the issues of most concern to the audiences is an important component of an effective presentation.
- Show audiences that something can be done to improve the situation. Otherwise, a powerful presentation can create a feeling of hopelessness. The last part of a presentation or document should describe effective actions that can be taken now. These may be actions that those in the audience can implement themselves, or actions by others that they should support.
- Use projections to illustrate how past trends may be changed and how AIDS affects the ability to achieve future goals. It can be useful to include epidemiological projections of the number of people infected, number of AIDS cases and AIDS deaths, and projections of the social and economic impacts of AIDS, such as effects on health care costs, the number of orphans, economic growth, women’s lives, and rural development. Such projections can be used to show the full range of impacts that AIDS can have on families, communities, and nations.

CONCLUSION

The importance of dissemination should not be underestimated. With effective dissemination, survey and other evaluation data can become an important component of advocating for expanded action and motivating a broader societal response to HIV prevention. In summary:

- Dissemination is a process that begins when an evaluation is first conceptualized and designed and continues through the life of the project. The objectives of this process are not only to disseminate the findings widely, but to see that they are understood and acted upon.
- In general, results should be disseminated as broadly as possible, but in ways that respect the concerns and sensitivities of the target audiences and key stakeholders. Possible target audiences include policymakers, leaders and members of the communities under surveillance, program managers in government, NGOs and CBOs, mass media, and the general public.
- Many forms of dissemination, including technical reports, policy briefs, mass media briefings, and group and individual meetings, will typically be required to reach all target audiences who can act upon the findings.
- Materials should be prepared in simple language that can be understood by the target audiences. Targeted materials should discuss the implications of the findings for the specific audience and present ideas for appropriate action by them.
- Staff should pay careful attention to the way information regarding stigmatized communities is disseminated so as to not contribute to their further marginalization.
- An effective strategy for disseminating findings to key stakeholders is an important factor in expanding the societal response to HIV prevention.

IV

ASSESSING PROGRAM IMPACT

Understanding HIV Epidemics and the Response to Them: Linking Data on Behavior Change and HIV Decline	205
Translating Survey Data into Program Impact: The AVERT Model	219
Guidelines for Assessing the Economic and Financial Costs of HIV/AIDS Prevention and Care Programs	235
Guidelines for Performing Cost-Effectiveness Analysis of HIV/AIDS Prevention and Care Programs	251

14

UNDERSTANDING HIV EPIDEMICS AND THE RESPONSE TO THEM: LINKING DATA ON BEHAVIOR CHANGE AND HIV DECLINE

STEPHEN MILLS, THOMAS REHLE, BERNHARD SCHWARTLÄNDER

Monitoring HIV Trends	209
From Prevalence to Incidence: Focusing ANC Sentinel Surveillance on Younger Age Groups	211
Linking Behavioral Data Collection and HIV Serosurveillance	211
Country Examples of Behavioral Change and HIV/STI Decline	212
Strengthening the Link Between Behavior Change and Epidemic Decline	214
Conclusion	216
References	218

14

UNDERSTANDING HIV EPIDEMICS AND THE RESPONSE TO THEM: LINKING DATA ON BEHAVIOR CHANGE AND HIV DECLINE

In simple scientific terms, the growth of an HIV epidemic is determined by the following factors: average probability of HIV transmission per exposure to an infectious partner, the number of exposures of susceptible persons to infectious partners per unit of time,

and the duration of the infectious period¹. These factors produce the reproductive rate or the number of new infections (secondary cases) generated by each case, expressed in the following equation: $R_0 = \beta cD$, where:

R_0 = Basic reproductive rate

β = Average probability of HIV transmission per exposure to an infectious partner

c = Number of exposures of susceptible persons to infectious partners per unit of time

D = Duration of infectious period

This simple formula of R_0 becomes more sophisticated once the variability or heterogeneity in sexual activity (the variance of the probability distribution of new partners per unit of time) within a defined community is taken into account². A high variance created by a core group of highly sexually active individuals may ensure that R_0 exceeds 1.0 in value even when

the mean rate of acquisition of new sexual partners in the wider population is small. The heterogeneity of sexual behavior, therefore, plays a very important role not only in determining the course of the HIV epidemic but also in selecting an adequate prevention and control strategy.

A schematic illustration of the relationship between incidence, prevalence, and mortality is shown in Figure 14-1³. The basic reproductive number (R_0) is the number of new infections caused on average by one infectious individual in an entirely susceptible population. The reproductive number at time t (R_t) changes as the epidemic progresses. At the beginning of the epidemic's growth phase, HIV incidence and HIV prevalence are likely to grow exponentially in the population at risk. As the epidemic grows, the number of people who are susceptible to HIV infection will decrease in the population at risk. At the same time, the proportion of contacts of those infectious with other members

of the population who have already been infected will increase. This effect reduces the reproductive rate of the infection and will slow the growth of incidence. In this scenario, HIV incidence will eventually decline ($R_t < 1$) while HIV prevalence continues to grow.

In the transition phase between the epidemic growth phase and the endemic steady state, HIV incidence stabilizes ($R_t = 1$). HIV-associated mortality, however, becomes greater than the incidence of new infections and, as a consequence, HIV prevalence will decline. Once the in-flow of new

HIV infections and out-flow of HIV-associated deaths reaches an equilibrium in the population, prevalence will remain stable at an endemic level.

In reality, however, the above parameters are difficult, if not impossible, to prescribe and quantify in populations. This is because the population dynamics of HIV transmission depend to a great extent on a number of factors, including the prevalence and intensity of multiple sexual partnerships, temporal patterns of sexual exchange rates (concurrent or serial), the presence of overlapping or closed sexual networks, and

the prevalence of HIV and other sexually transmitted infections (STIs). Also, these factors are neither uniform within a given population nor stable over time.

Other factors influencing this “natural course” of an HIV epidemic include the response of the individual and the population as a whole to the epidemic. Changes in behavior may be induced by effective prevention strategies or by the heightened awareness of the problem when increasing numbers of members in the population develop the disease. On the other hand, reversals in initial reductions in risk behavior may occur when those initially targeted “age out,” and new generations enter into their sexually active life span.

Therefore, measurements of HIV spread and trends over time cannot be interpreted without knowledge of levels and changes in risk behavior. Changes in behavior should result in changes in HIV incidence and prevalence and, given the difficulties in measuring risk behavior as discussed in previous chapters, effectiveness of programs aimed at behavior change can ultimately only be assessed when compared with trends in HIV incidence and prevalence.

A consensus is emerging among evaluation experts that prevention programs need to investigate trends in infection alongside trends in behavior that may lead to that infection⁴. To this end, HIV data have to be collected in conjunction with behavioral, socioeconomic, and socio-demographic data. The combined analysis of these data sets will provide the necessary context and range of information for an interpretation and explanation of the epidemiological data collected by sentinel serosurveillance surveys.

When reductions in HIV prevalence and HIV incidence are observed at the population level, however, the following key questions arise:

- Are the observed changes a product of behavior change?
- Are the observed changes a result of prevention interventions?

While some country surveillance systems provide convincing evidence that population-level behavioral change results in a decline or continuing low prevalence of HIV infection, a perfect correlation between the two is unlikely and, in fact, impossible to prove. As described in the general model above, other factors that are unrelated to behavioral change may be working to reduce HIV prevalence. These may include:

- saturation of the epidemic among susceptible individuals;
- increasing AIDS-associated mortality, especially in mature epidemics;
- reductions in symptomatic and asymptomatic STIs through improved treatment; and
- population differentials related to in- and out-migration patterns.

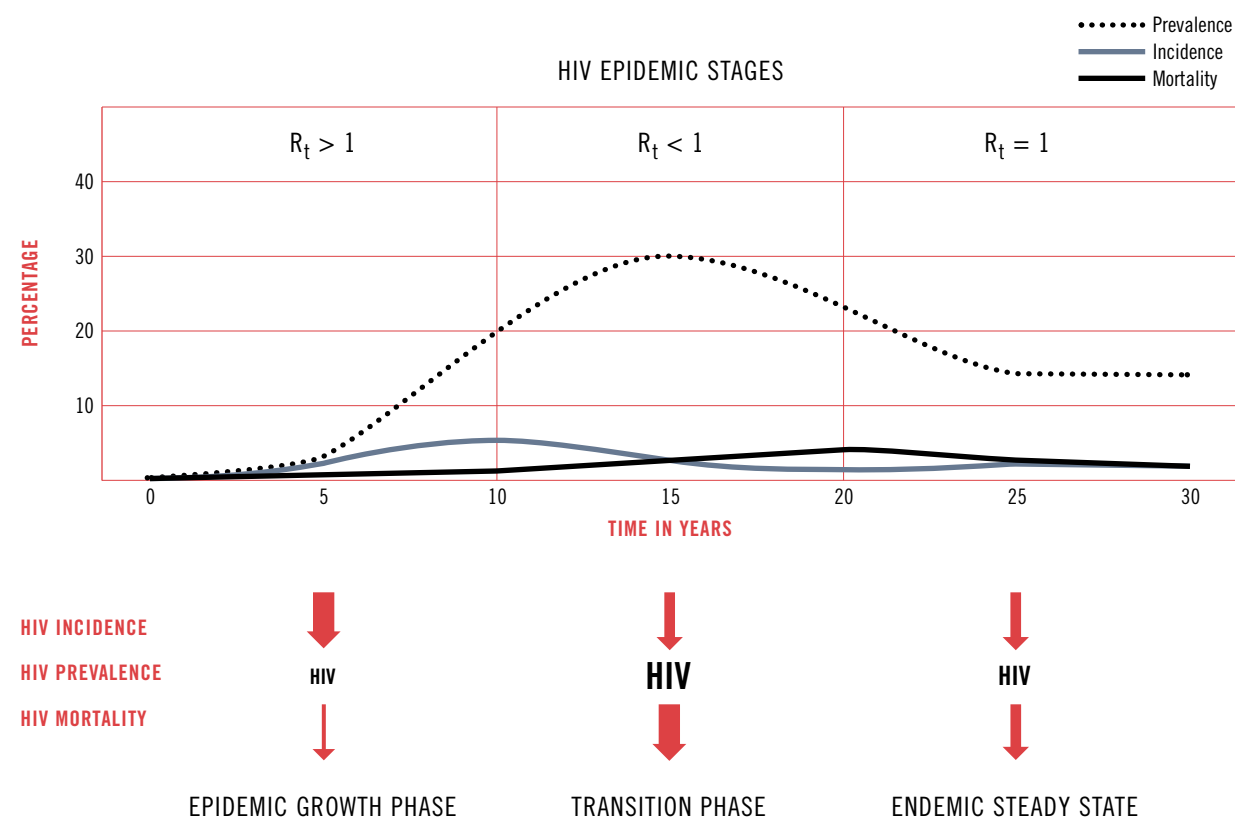
Before assigning credit to behavioral change, all of these factors should be investigated with available data to determine whether they may have a leading role in HIV epidemic decline.

The difficulties in establishing evidence of a relationship between behavior change and epidemic decline is further compounded if we add the claim that the documented behavioral change is caused by intervention effects. This may also not necessarily be the case. For example, behavioral change may be caused by the increasingly visible presence of people with HIV/AIDS over time, which helps to personalize risk. This may happen independent from any programmed intervention.

MONITORING HIV TRENDS

Ideally, HIV spread and changes over time are

FIGURE 14-1 Relationship between HIV incidence, prevalence, and mortality



Source: Modified from Reference 3

tracked through measuring the number and distribution of new infections in a population. True incidence data, however, can only be obtained through large-scale cohort studies. Such studies have many drawbacks, including cost, ethical considerations, and biases due to selection and the fact that those included in a cohort will inevitably have more exposure to HIV programs or intervention efforts.

In the absence of true HIV incidence data, a system of sentinel surveillance for monitoring HIV infection among selected populations has been recommended⁵. HIV sentinel surveillance uses unlinked and anonymous HIV testing of residual blood specimens left over from samples collected for clinical purposes. Using this approach, it is possible to perform HIV tests without informed

Comparisons between ANC sentinel surveillance and general population serosurveys, however, have shown that data from pregnant women may differ significantly from the general population data and the relationship can go in different directions at different stages of the epidemic, and for different age groups^{6,7}. The considerable variation in the findings suggests that extrapolations from ANC data should be made with caution.

Because of the large denominator used (general population of women), even a small percentage difference in HIV prevalence between pregnant women and women in the general population could potentially result in an over- or underestimation of the total number of women infected.

ANC sentinel data also are subject to **selection biases** related to convenience sampling (sites are

In the absence of true HIV incidence data, a system of sentinel surveillance for monitoring HIV infection among selected populations has been recommended.

HIV sentinel surveillance uses unlinked and anonymous HIV testing of residual blood specimens left over from samples collected for clinical purposes.

consent and minimize the possibility of participation bias associated with voluntary testing.

Antenatal clinic (ANC) attenders have been selected as a suitable “sentinel” group because they are thought to represent most closely the HIV prevalence of the general sexually active population. ANC data are the primary source of data for monitoring HIV prevalence trends in most developing countries, especially in sub-Saharan Africa.

not randomly chosen), usage and coverage of ANC services, differentials in risk behaviors and contraceptive use, differences in fertility between HIV positive and HIV negative women, and other socio-demographic factors (such as age distribution of antenatal clinic attenders, level of education, and socio-economic status). Moreover, insufficient data exist on the relative importance of these factors in different settings, and even less is known about how these factors may vary over time.

Lower fertility rates have been found among women with HIV-1 infection⁸. It has been shown that differential fertility between HIV-infected and HIV-uninfected women can have substantial effects on ANC prevalence estimates and the total number of prevalent infections estimated from it. Especially in mature epidemics, the effect due to differential fertility is expected to be most apparent in older ANC age groups. In fact, population-based studies have shown that ANC women may underrepresent HIV levels in the general female population in such settings⁷.

If factors leading to selection biases remain the same over time, then serial data from ANC sentinel sites will provide a solid basis for analysis of HIV trends in that population. However, the selection biases referred to above may change over time. In that case, trends recorded over time in ANC sentinel populations may differ from those actually occurring in the general population.

Population-based studies carried out periodically in the catchment areas of ANC sentinel sites can help to evaluate these sources of bias in different country settings. Periodic surveys of this type are needed to compare data on HIV prevalence trends in the general population with those obtained from routine ANC sentinel surveillance systems. This approach allows the necessary “calibrations” of results obtained from pregnant women. In addition, these surveys also provide information on HIV levels in the male population.

From Prevalence to Incidence: Focusing ANC Sentinel Surveillance on Younger Age Groups

In mature epidemics, the majority of new HIV infections are now occurring in young people⁹. By concentrating resources on younger age groups in ANC surveillance efforts, it will be possible to obtain more information on relatively recent infections. Bias due to differential mortality and fertility will be of less concern in young antenatal clinic attenders. It has therefore been recommended that, while continuing to

collect data in all age groups, oversampling should be attempted in the age range of 15-24 years¹⁰. Because sample sizes must be large enough to allow a more detailed age-stratification, key sentinel sites with a high volume have to be identified. This could mean that countries may have to downsize the number of sentinel sites in the interest of concentrating resources where they will be most useful.

LINKING BEHAVIORAL DATA COLLECTION AND HIV SEROSURVEILLANCE

To minimize biases, not jeopardize the validity of the serological data, and gather less biased behavioral data on the population as a whole, it is recommended that blood samples and risk behavior interviews be obtained from different individuals⁴. To establish a better association between behavior and HIV prevalence in the community, however, the data on HIV serostatus among ANC women and behavior in the general population should be drawn from the **same source population**. This means carefully defining the population from which a key sentinel surveillance site (such as a large urban antenatal clinic) draws its attendees and collecting behavioral data from a random selection of households in the same catchment area. If this is done as part of national or regional behavioral surveys, it may mean deliberately oversampling in the catchment populations of key sites (that is, the population that is served by the particular site in question).

To link the behavioral data with the HIV prevalence data, it is therefore recommended that a minimum set of socio-demographic questions be asked of all antenatal clinic attenders at sentinel sites. These data would include age, parity, last birth interval, level of schooling, occupation, and length of time living in the area (as an indicator of migration). These parameters can then

be compared with those collected in the population-based behavioral surveys, allowing any systematic differences between the two groups to be identified and adjusted for in the analysis.

Country Examples of Behavioral Change and HIV/STI Decline

The following section describes and discusses the experience of Uganda and Thailand—two countries where enough reliable behavioral and biologic data have been generated through national surveillance systems to strongly indicate the relationship between large-scale population behavioral risk reduction and declines in HIV and STIs.

Uganda

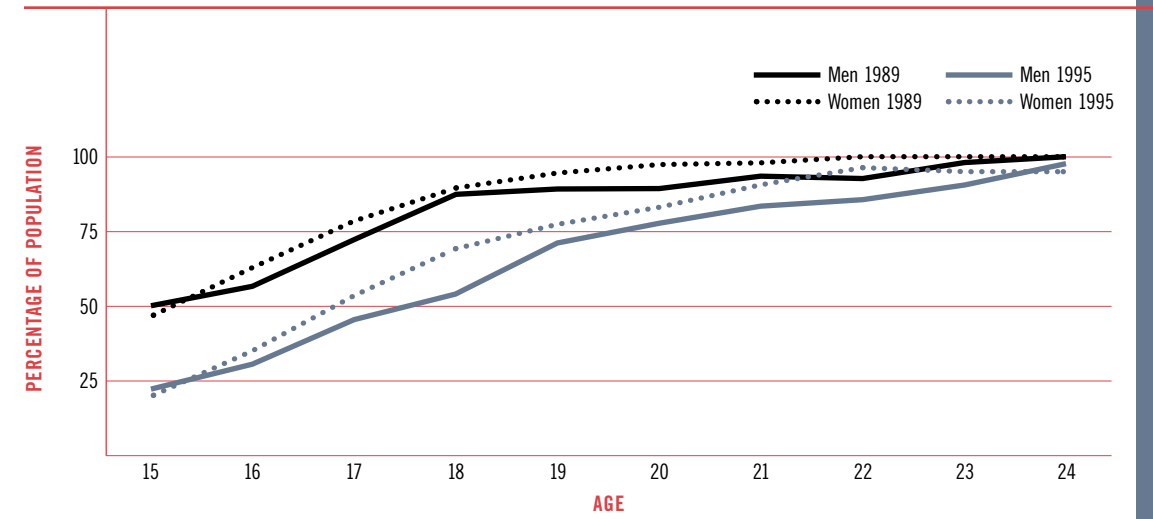
A rapidly disseminated epidemic occurred in Uganda during the 1980s. By 1992, HIV prevalence among female attenders of selected ANC clinics in Kampala had reached about 30 percent. Other urban sites in the country reached similar levels.

In 1993, HIV prevalence began a slow decline among the ANC clinic population in Kampala, and by 1996 it had reached 15 percent, half of the prevalence only 4 years before. Furthermore, HIV prevalence among women attending an antenatal clinic in Mulago showed significant declines in all age groups except persons aged 38 years and above. The overall prevalence rate decreased from 28.1 percent in 1989-90 to 16.2 percent in 1993. As shown in Figure 14-2 for the example of pregnant women in Nsambya, the declines were most pronounced in the youngest age groups¹¹.

What caused this decline in prevalence? Although several factors may have worked together to produce this decline, the strongest lies in the documented evidence of extensive behavioral change at the population level in Uganda. A representative adult survey in the country in 1989 indicated substantial risk among both men and women: 38 percent of men and 19 percent of women reported at least one non-regular sex partner in the past year. Similarly, a majority of both male and female youth were sexually active (69 percent and 74 percent, respectively).

Percentage of population with sexual experience, by age (15-24 years), in 1989 and 1995, Uganda

FIGURE 14-3



Source: References 12 and 13

During the years following this national survey, Uganda initiated a massive HIV prevention program consisting of a variety of strategies. Simultaneously, people with AIDS began visibly appearing in society, buttressing interventions with human evidence, which helped to increase perceived risk and move individuals toward behavioral change.

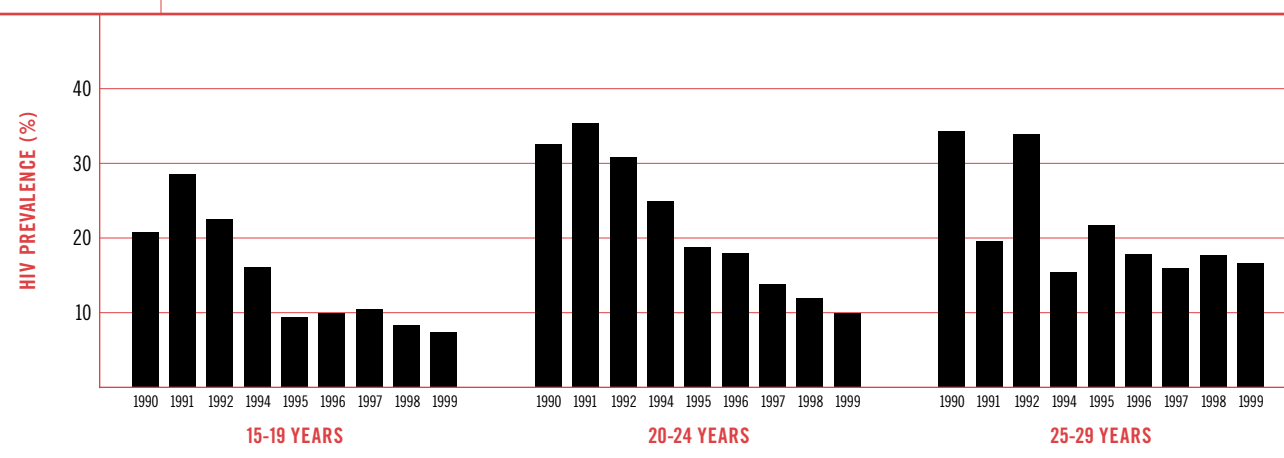
A follow-up national survey in 1995 of the same universe of adults conducted by the Uganda Ministry of Health indicated that behavioral risks had significantly declined^{12,13}. Both adult men and women had cut their risk by about one-half: only 15 percent of men and 6 percent of women reporting having a non-regular sex partner in the past year. Youth populations had similarly reduced their risk, with a greater percentage of them delaying onset of intercourse: In 1995, fewer than 25 percent of 15-year-old boys and girls reported ever having had sex, down from about 50 percent in 1989 (Figure 14-3).

Condom use increased during the same period: In urban areas, 61 percent of men and 48 percent of women reported condom use in non-regular partnerships.

Thailand

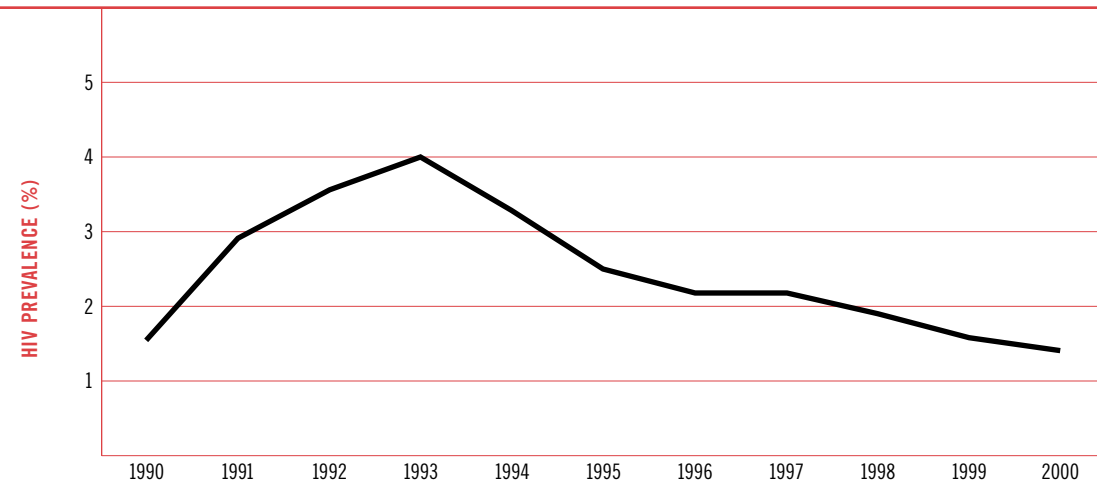
The first case of HIV in Thailand was discovered in 1984. By 1989, prevalence among sex workers in the northern city of Chiang Mai was 44 percent and among injecting drug users almost 40 percent. Over the next few years, an extensive national surveillance system documented prevalence increases among both high-risk groups as well as among the general population. HIV infection among brothel-based sex workers climbed steadily until 1995 when over one-third were HIV-positive. In that same year, 12 percent of “indirect” sex workers found in bars, restaurants, and lounges tested HIV-positive. During the same time period, infection levels in the general population also climbed. By 1993, HIV levels in 21-year-old male military recruits

FIGURE 14-2 HIV prevalence by age group, Nsambya, Uganda



Source: Reference 11

FIGURE 14-4 Trend in HIV prevalence in 21-year-old military conscripts in Thailand



Source: Reference 14

throughout the country had reached levels of almost 4 percent, with levels of 10 percent and more in some parts of the country. A year later, over 2 percent of women in ANC clinics tested HIV positive.

In 1994, hopeful signs started to emerge of decreasing new infections in the country. That year, the overall prevalence among army conscripts decreased to about 3 percent, and the reductions were even greater in the North, where the prevalence had been highest (Figure 14-4). Furthermore, new cases of five other STIs seen in government hospitals and STI clinics decreased by over 80 percent between 1989 and 1996 (Figure 14-5).

What evidence do we have that these major reductions in STIs, including HIV, were caused by behavioral change? The available behavioral data show that in Thailand as in Uganda, these reductions were preceded by significant behavioral change. In Thailand, this took the form of reduced patronage of sex workers by

males as well as increased condom use in commercial sex relationships (Figure 14-6). This temporal linkage of behavioral change followed by biologic reductions is an important factor in strengthening the causal relationship.

For example, only 24 percent of national army recruits reported visiting a sex worker in 1995, down from 57 percent in 1991¹⁶. Condom use questions attached to the national HIV surveillance survey indicated that between 1989 and 1993, reported condom use increased from 14 percent to 94 percent of commercial sex acts. These behavioral changes most definitely slowed down HIV transmission, but it was not until 1994 that they were reflected in decreased HIV prevalence rates.

Strengthening the Link Between Behavior Change and Epidemic Decline

To detect and build evidence for behavioral change leading to HIV epidemic decline, methodologies must be implemented to provide

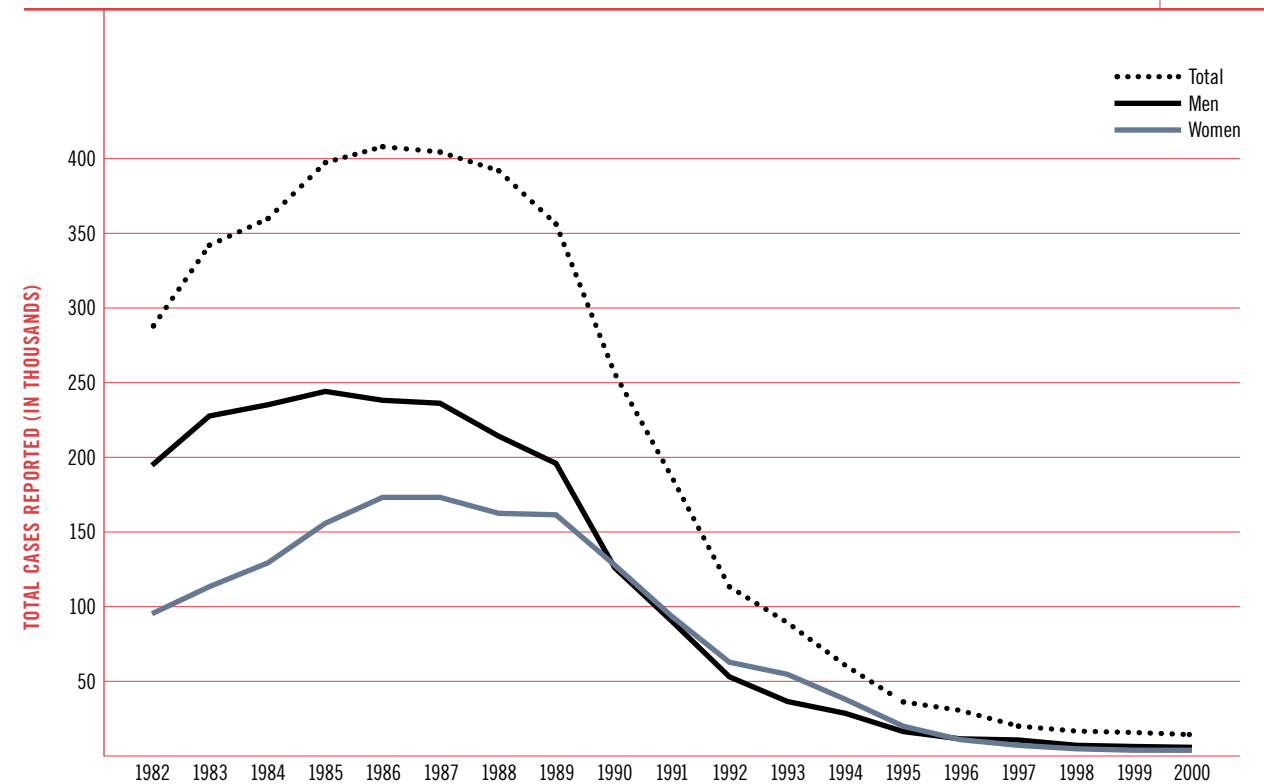
the necessary data. In their most recent guidelines for “Second Generation Surveillance Systems,” UNAIDS and WHO have stressed a combination of regular and systematic HIV surveillance combined with STI and behavioral surveillance¹⁰. Establishing these monitoring systems will allow analyses similar to those highlighted in Uganda and Thailand, which illustrate significant population-level behavioral change preceding HIV prevalence decline.

The state of Tamil Nadu, India, has integrated HIV and behavioral surveillance so that both behavioral and HIV trends are monitored in key groups. The HIV surveillance system, supported

by the Government’s Tamil Nadu State AIDS Control Society, is monitoring HIV rates in both high-risk and general population groups to assess epidemic spread. These groups include symptomatic or asymptomatic STI patients, truck drivers (through to 1997 only), tuberculosis (TB) patients, and ANC clinic attenders. Conversely, the Tamil Nadu behavioral surveillance system, sponsored by the USAID-funded AIDS Prevention and Control Project (APAC) of Voluntary Health Services (VHS), has been monitoring risk behaviors since 1996 among selected population sub-groups whose risk reduction is key to reducing HIV levels in both

Number of STIs reported from government clinics, by gender, Thailand

FIGURE 14-5



Source: Reference 15

high-risk and general population groups. These groups include female sex workers, truckers, male and female factory workers, and female and male students (through 1997 only).

HIV surveillance has shown steadily rising HIV prevalence rates in all these groups. By 1998, HIV had climbed to 14.7 percent among STI clinic attenders, 7.7 percent among TB patients, 9.4 percent among truckers (in 1997), and 0.95 percent among ANC clinic attenders.

The results from the sub-groups selected for behavioral surveys have indicated substantial behavioral risk reduction. Both truckers and male factory workers reported fewer non-regular sex partners, including commercial partners, between 1996 and 1998, and both of these groups, together with the female sex workers, reported higher levels of condom use in commercial sex during the same period.

While these changes have definitely slowed the spread of the epidemic, they have not yet resulted in a decline of the overall HIV prevalence level. It is the task of the future waves of systematic data collection, both HIV and behavioral, to detect when this critical downward trend in the epidemic will occur.

CONCLUSION

The theoretical considerations and country examples above have shown the importance of combining serosurveillance data with data from behavioral studies to understand patterns and trends of HIV in populations. Furthermore, as also discussed in the guidelines for second generation surveillance of HIV infection, it is essential to analyze data—both on HIV status and

behavior—by age. A focus on young people allows the collection of data that more closely, and with less bias, correspond to recent developments in the epidemic. Thus, the impact of program efforts on HIV transmission can be detected much earlier in these groups than in interventions that include the whole population.

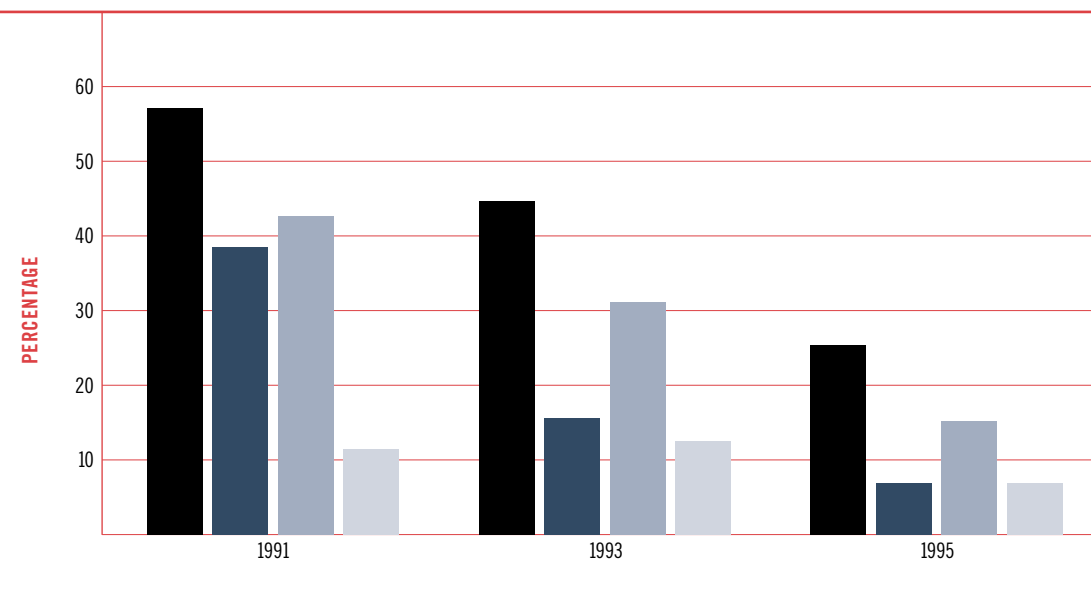
Models and thought have also focused on the question of how much behavior change is necessary to bring the epidemic to a halt¹⁷. Clearly, if risk behavior is extremely rare, the chances of acquiring or transmitting HIV infection will be very low and an epidemic will not develop. However, this concept holds only true on a population level, and by no means can guarantee that an individual will be safe when exposing him or herself by engaging in risk behavior. HIV is present in virtually all countries and regions of the world, and prevention thus needs to aim at all risks that may occur.

While the classic concept of core groups for transmission of HIV and other sexually transmitted infections has a lesser meaning in most of the high level and generalized epidemics in sub-Saharan Africa, in other regions of the world this concept still has important implications for understanding and preventing the infection. There, risk behavior, and consequently HIV infection, is usually concentrated in defined sub-populations, such as sex workers and their clients, men having sex with men or injecting drug users. While risk levels in the population overall may be low (on average), which prevents the epidemic from taking off at high levels, HIV is often concentrated within the populations at highest risk. Thus the simplistic model described in the beginning of this chapter has to be revisited when applied to populations at large. In such situations it is more appropriate to speak about several epidemics that may well occur in parallel in different populations in a country. In fact, genetic analysis of

HIV specimens isolated from different populations in Thailand have shown that epidemics in drug users seem rather distinct from those in homosexual men or female sex workers and their clients. While the spread of HIV in low-risk populations is primarily triggered through sporadic infections through contacts between members of low- and high-risk groups, the epidemics within these high-risk populations are characterized by a rapid spread of HIV in a relatively short period of time and higher HIV prevalence levels.

The determinants of HIV transmission described at the beginning of this chapter establish the theoretical basis for current HIV/AIDS prevention efforts. The challenge for program designers lies in trying to identify the most effective ways to decrease HIV transmission by influencing these determinants and to translate this theoretical concept into feasible interventions in the field.

FIGURE 14-6 Behavioral change and HIV/STI decline among 21-year-old men in northern Thailand



Source: Reference 16

■ Visited sex worker last year ■ Lifetime history of STIs
 ■ Did not use condom on last visit ■ HIV-positive

REFERENCES

- 1 May RM, Anderson RM. Transmission dynamics of HIV infection. *Nature* 1987;326(6109):137-142.
- 2 Anderson RM. Some aspects of sexual behaviour and the potential demographic impact of AIDS in developing countries. *Soc Sci Med* 1992;34(3):271-280.
- 3 Joint United Nations Programme on HIV/AIDS (UNAIDS). Trends in HIV incidence and prevalence: natural course of the epidemic or results of behavioural change? Geneva: UNAIDS; 1999. Best practice collection 99.12E.
- 4 Family Health International/IMPACT and Joint United Nations Programme on HIV/AIDS (UNAIDS). Meeting the behavioural data collection needs of national HIV/AIDS/STD programmes. Proceedings from a joint IMPACT/FHI/UNAIDS workshop. Arlington (VA) and Geneva: Family Health International/IMPACT and UNAIDS; 1998.
- 5 WHO/GPA/DIR/88.8. Sentinel surveillance for HIV infection: a method to monitor HIV infection trends in population groups. Geneva: World Health Organization; 1988.
- 6 Fontanet AL, Messele T, Dejene A, et al. Age- and sex-specific HIV-1 prevalence in the urban community setting of Addis Ababa, Ethiopia. *AIDS* 1998;12(3):315-322.
- 7 Fylkesnes K, Ndhlovu Z, Kasumba K, et al. Studying dynamics of the HIV epidemic: population-based data compared with sentinel surveillance in Zambia. *AIDS* 1998;12(10):1227-1234.
- 8 Gray R, Wawer M, Serwadda D, et al. Population-based study of fertility in women with HIV-1 infection in Uganda. *Lancet* 1998;351(9096):98-103.
- 9 Taha TE, Dallabetta GA, Hoover DR, et al. Trends of HIV-1 and sexually transmitted diseases among pregnant and postpartum women in urban Malawi. *AIDS* 1998;12(2):197-203.
- 10 WHO/CDS/CSR/EDC/2000.5 and UNAIDS/00.03E. Guidelines for second generation HIV surveillance. Geneva: World Health Organization and Joint United Nations Programme on HIV/AIDS; 2000.
- 11 STD/AIDS Programme, Ministry of Health. HIV/AIDS surveillance report. Kampala (Uganda): Ministry of Health; June 2000.
- 12 Uganda Demographic and Health Survey. Calverton (MD): Macro International Inc.; 1996.
- 13 Asiimwe-Okiror G, Opio AA, Musinguzi J, et al. Change in sexual behaviour and decline in HIV infection among young pregnant women in urban Uganda. *AIDS* 1997;11(14):1757-1763.
- 14 Armed Forces Research Institute of Medical Sciences, Thailand; 2000.
- 15 Venereal Disease Division, Ministry of Public Health, Bangkok, Thailand; 2000.
- 16 Nelson K, Celentano D, Eiumtrakol S, et al. Changes in sexual behaviour and a decline in HIV infection among young men in Thailand. *N Engl J Med* 1996;335(5):297-303.
- 17 Morris M, Dean L. Effect of sexual behavior change on long-term human immunodeficiency virus prevalence among homosexual men. *Am J Epi* 1994;140(3):217-232.

15

TRANSLATING SURVEY DATA INTO PROGRAM IMPACT: THE AVERT MODEL

THOMAS REHLE, TOBI SAIDEL, PAUL BOUEY, MICHAEL SWEAT, SUSAN HASSIG, AND DAVID SOKAL

Description of AVERT	222
What's Behind the Model?	222
Validation of AVERT Estimates	226
Practical Example From the Field	227
Points for Consideration	230
Conclusion	232
Acknowledgments	232
References	233

15

TRANSLATING SURVEY DATA INTO PROGRAM IMPACT: THE AVERT MODEL

The ultimate goal of any HIV prevention program is to reduce the number of new infections. Evaluating the impact of HIV/STI interventions on reducing HIV transmission is, therefore, an essential part of overall prevention and control efforts.

However, this type of program assessment remains an elusive goal for most AIDS control programs because of the prohibitive costs and methodological difficulties associated with field-based program impact evaluation. As a consequence, the important question of the extent to which program effects may lead to reductions in HIV transmission is usually not answered by intervention programs.

In response to these challenges, Family Health International (FHI) has developed the AVERT computer model, which may offer a useful option to address that question¹. The AVERT model can be used to estimate the impact of prevention interventions, such as those that focus on increasing use of condoms, improving treatment of sexually transmitted infections (STIs), or changing sexual behaviors, on the reduction of primary HIV transmissions through sexual intercourse over a given time period.

Various types of computer models of the AIDS epidemic have been devised for various purposes. Many, if not most of them, are relatively complex and some can only be used as research tools by their creators. There is a clear need for less complex models that are accessible to epidemiologists and other public health specialists who have limited time and resources. The AVERT model answers this need in a manner consistent with the availability of local program data, which means that data requirements are considerably less extensive than those for more sophisticated simulation models such as SimulAIDS or iwgAIDS².

While AVERT is much easier to use than other models, its static nature is not appropriate for modeling long-term scenarios. AVERT cannot be used to explore some of the more complex questions that dynamic simulation models aim to address in the context of hypothesis testing.

This was a conscious strategy to keep the model accessible to non-experts in modeling and to accommodate the types of data available to most intervention programs.

This chapter explains the approach used in AVERT and describes the required input variables. Estimates generated by the model are validated against seroconversion data from a large cohort study and practical examples of its application are presented.

DESCRIPTION OF AVERT

AVERT is a relatively simple, static, mathematical model with an interface developed especially for users with little experience in modeling. The model requires relatively few input parameters, all of which may be readily available to programs, and it focuses on population dyads, namely hetero/homosexual couples in union or female sex workers and their clients. Using changes in four major intervention parameters that affect transmission rates, the model calculates how many fewer infections there would be in a defined time period. The parameters are:

- number of partners;
- number of sex acts per partner;
- condom use; and
- STI levels.

Model estimates are usually based on a 1-year time frame.

Developing this relatively simple and easy-to-use model involved an important trade-off in that some of the complexities of the epidemic are not incorporated. For example, transmission probabilities vary over the course of an infection, HIV and other STIs are closely associated in a population, condom use patterns can vary among different dyad groups, and sexual behaviors can exhibit different patterns of mixing.

AVERT cannot account for these types of heterogeneity. However, the dynamic character of variables such as patterns of sexual activity, partner change, or condom usage plays less of a role when the modeled period of time is relatively short (1-2 years).

Version 1.0 of AVERT is programmed for use in a DOS environment with an IBM compatible microcomputer. The model is public domain software that may be freely copied. AVERT was developed under FHI's AIDSCAP project, with the support of the United States Agency for International Development (USAID). Copies of the AVERT software and the user's manual have now been widely disseminated to country programs, researchers, and academic institutions.

WHAT'S BEHIND THE MODEL?

The mathematical foundation underlying the AVERT model is a derivation of a probability formula presented by Weinstein and colleagues³. The structure of Weinstein's model is based on the risk of an individual becoming infected through sexual acts with a partner who has been randomly drawn from a population with a given prevalence of HIV infection. AVERT takes this model one step further and multiplies this probability by the number of susceptible individuals in an "at-risk" population. This provides an estimate of new infections. The equation is invoked twice in order to provide estimates for each of the participating sex partner populations. For example, under a given set of conditions, an estimate is calculated for a target population of commercial sex workers and another is calculated for their client partners.

The basic model in AVERT incorporates seven different variables:

- HIV prevalence among sexual partners— p
- average number of sexual partners— m

- average number of sexual acts with a given partner— n
- proportion of sexual encounters in which condoms are used— f
- efficacy of condoms— e
- prevalence of STIs in population— w_i
- HIV transmissibility— r_{gi}

The cumulative probability equation is:

$$P = 1 - \left\{ p \sum_{i=1}^4 w_i [1 - r_{gi} (1 - fe)]^n + (1 - p) \right\}^m$$

To implement the model, a user carries out a first sequence of calculations that yields a probability of risk to the target population (population A) resulting from their having sex with members of the partner population (population B):

$$P_{A \leftarrow B} = 1 - \left\{ p_B \sum_{i=1}^4 w_i [1 - r_{A \leftarrow Bi} (1 - fe)]^{n_A} + (1 - p_B) \right\}^{m_A}$$

The risk to A derives from several behavioral characteristics of A, and from HIV prevalence of and HIV transmissibility from B. The result is multiplied by the number of susceptible individuals in A to yield an estimate of new infections within A. A second set of calculations is then carried out following an identical procedure, but measuring risk to B on the basis of B behavioral characteristics and HIV prevalence of and transmissibility from A.

$$P_{B \leftarrow A} = 1 - \left\{ p_A \sum_{i=1}^4 w_i [1 - r_{B \leftarrow Ai} (1 - fe)]^{n_B} + (1 - p_A) \right\}^{m_B}$$

This result is multiplied by the number of susceptible individuals in B, generating an estimate of new infections within this group.

In summary, once cumulative probabilities are calculated for each study population ($P_{A \leftarrow B}$ and $P_{B \leftarrow A}$), those values are multiplied by

the corresponding HIV-negative population. These procedures produce estimates of new HIV infections within each group, and the total comprises the estimate for the target and partner populations combined.

AVERT must be run separately for each type of partner population with which members of the target population might be sexually active. For example, if the target population is male factory workers who have sex with both their regular partners and with female sex workers, the model would have to be run separately to get estimates for each group. It is up to the user to select which type of partner populations is relevant to the intervention.

When HIV prevalence and the size of the study population are thought to remain the same between, for example, a pre- and post-intervention scenario, the AVERT program allows the user to compare the two scenarios simultaneously. On the other hand, if the HIV prevalence or the population size vary between the scenarios, then separate runs with the model are necessary to estimate the number of new HIV infections for each scenario. The cumulative HIV incidence rates projected by each run with the model are then the appropriate figures to use when comparing different scenarios.

The following paragraphs provide additional detail on each of the seven variables included in the AVERT model.

HIV prevalence (p)—The HIV prevalence variable incorporates two separate values. When measuring risk of the target population (A), the HIV prevalence of their partner population (B) is used. When measuring risk to the partners (B), the HIV prevalence of the target population (A) is used and calculations are treated independently. HIV prevalence works as a constant variable in all circumstances. This is unlike dynamic simulation models, in which HIV prevalence varies over time within a single scenario. Thus, AVERT should not be used to

estimate new HIV cases for time periods of more than one year at a time because new infections are not taken into account in the probability function.

Average number of sexual partners (m)—One value is given for the target population (A) and a second for the partner population (B). These numbers can vary considerably based on the types of populations under scrutiny. The size of the target and partner population is held constant for all calculations when two scenarios are compared simultaneously. While the size of the target population is specified by the user, the size of the partner population is calculated by the program. This is necessary to ensure that the number of sexual acts by both groups are equal.

Average number of sexual acts with each individual (n)—A number is required for the target population (A)—the overall average number of times each target (A) member has sex with each of her/his partners (B)—and for the partner population (B)—the overall average number of times each partner (B) member has sex with each of his/her target population (A) contacts. (It is obvious that the number of sex acts per partner must be identical for any two individuals who have sex with each other.)

Condom use (f)—This variable represents the proportion of sex acts that are protected through use of a condom. Data on different self-reported behaviors might be used to estimate a value for this parameter, depending on what is available from behavioral surveys. For example, if the appropriate data were available, one could base this proportion on the actual number of sex acts with specific types of partners in which a condom is used. In the absence of these data, one might start with the proportion of people in the study population who report using condoms 100 percent of the time with the partner population in question, and then augment that number with a portion of those who report using condoms some of the time. In either instance,

this variable is assumed to be spread uniformly across the population.

Condom efficacy (e)—The value assigned to this variable reflects the success rate of condoms in preventing the transmission of HIV during a single sexual exposure. Based on related research, the default value found in AVERT is 95 percent and is used as a constant in all calculations^{4,5}.

Prevalence of STIs (w_i)—The STI prevalence variable is separated into four different categories (represented by the subscript i): STIs that cause genital ulcer disease (GUD), non-ulcerative STIs (Non-GUD), a combination of ulcerative and non-ulcerative STIs (GUD + Non-GUD), and no-STIs. STI prevalence estimates are used as a proxy for the proportion of HIV-associated sex acts occurring in the presence of an STI during the modeled period of time. Only one set of values is entered into the AVERT model and applied to the entire study population. If the STI prevalence levels are thought to be different for the target (A) and partner population (B), it is recommended that the higher of the two values be used as the input parameter for the model. This approach allows the model to define reasonably well the number of sex encounters when either partner has an STI. One should keep in mind that the STI cofactor operates in a two-way fashion: through increased infectiousness of HIV-positive individuals and through increased susceptibility for HIV in HIV-negative individuals.

Data entry is restricted to two values: GUD and non-GUD prevalence. These values are multiplied by each other to yield the prevalence for the combination category (GUD + non-GUD), and that value is subtracted from each of the entries to provide the adjusted GUD and non-GUD prevalences. All three of these values are added together and then subtracted from 1 to obtain the no-STI proportion. Each set of calculations—risk to target (A) and risk to partners (B)—is stratified into these four STI groups for the

central component of the equation associated with the summation sign.

Results from these calculations are summarized and the final steps in the formula completed.

HIV transmissibility (r_{gi})—Values for HIV infectivity depend on the specific gender of the study populations. Three basic figures are included as defaults in AVERT: a value for male-to-female, a value for female-to-male, and a value for male-to-male (each of which is represented by subscript g). In a scenario in which the study target population includes female sex workers and their partners are male miners, the initial series of calculations measuring risk to the females uses a male-to-female transmission rate. The second series employs the female-to-male transmission rate to estimate the risk to the miners.

Our transmission rates represent a “best estimate” extracted from published literature⁶⁻¹⁶, adjusted to the mathematical foundation underlying the AVERT model and restricted to specified types of behavior. For example, the female-male

combinations assume vaginal intercourse only, and the male-male combination treats anal intercourse generically, without regard to patterns of receptive or insertive behaviors. These transmission rates also account for the fact that STIs have been shown to enhance the transmissibility of HIV¹⁷⁻²⁶. Ulcerative and non-ulcerative STIs have varying influences, and those effects are incorporated into the model (represented by the subscript i). If one or both types of STIs are identified for the study populations, corresponding transmission rates are invoked for the appropriate stratum calculations (see above). Consequently, each of the basic rates without an associated STI also has three additional variants: one in the presence of ulcerative STIs, one for non-ulcerative STIs, and one for a combination of STIs (this value is the same as that used for ulcerative STIs). A total of 12 transmission rates are therefore available in the program (Table 15-1).

HIV-1 transmission probabilities represent a combination of per contact infectivity and

Probabilities of HIV-1 transmission per sexual exposure as used in the AVERT model

TABLE 15-1

HIV transmission probability with no STI						
Male-Female	0.002					
Female-Male	0.001					
Male-Male	0.01					
Enhanced HIV-1 transmission probability in the presence of a STI cofactor						
	GUD		NON-GUD		GUD & NON-GUD	
	MULTIPLIER	RATE	MULTIPLIER	RATE	MULTIPLIER	RATE
Male-Female	30	0.06	10	0.02	30	0.06
Female-Male	60	0.06	10	0.01	60	0.06
Male-Male	30	0.30	10	0.10	30	0.30

GUD=genital ulcer disease

susceptibility, and the probability of HIV transmission is assumed to remain constant over individuals and over time. The selected values represent time-weighted average figures subsuming the different stages of HIV infection. We also assume that exposure to HIV is distributed independently of the presence of a STI. While these choices do not reflect all of the nuances inherent in the HIV transmission dynamics, we decided that a parsimonious model was most appropriate to our purpose.

The model's approach for estimating person to person transmission is based on HIV-1 transmission probabilities per sexual exposure. Unfortunately, only limited data exist on the per-exposure cofactor effect of different types of STIs on HIV transmission⁸. Most of the published literature on transmission probabilities report cumulative risk estimates, such as risk ratios or odds ratios. Cumulative risk, however, is not easily translated into an increased risk per sex act because it is usually the result of an unknown number of sex acts, of which only a fraction may have occurred in the presence of an STI. Depending on available specific research findings on HIV infectiousness and susceptibility to HIV associated with genital tract infections, in an updated version of the model, additional stratification of the STI cofactor effect estimates may be considered in order to better distinguish between the STI effect on the infectiousness of HIV-infected individuals and the STI effect on the susceptibility of HIV-uninfected individuals in the modeled populations.

VALIDATION OF AVERT ESTIMATES

The accuracy of quantitative estimates on the annual incidence of HIV infection generated by AVERT is heavily dependent on the validity of the underlying transmission probabilities per sexual exposure used in the model. AVERT

developers, therefore, considered it important to compare the model's estimates with the number of seroconversions observed in a real-life situation. We used the data set of a recently completed randomized controlled trial evaluating the effect of a commonly used spermicide on HIV transmission among female sex workers in two cities in Cameroon between March 1995 and December 1996 to validate the AVERT model²⁷. This study provided detailed data on the number of sexual acts per year with clients and non-clients, corresponding levels of sex acts protected by condoms, and estimates on the prevalence of STIs during the study period (Table 15-2). The HIV seroprevalence among the partners of female sex workers was estimated from 1994 data on male blood donors from the towns Douala, Yaounde, and Ebolowa (HIV positive = 11.5%, n=7,148)²⁸ and from 1996 data on male military collected from 11 army bases (HIV positive = 14.6%, n=1,052)²⁹.

To perform an objective validation of the model, the investigators in the Cameroon trial were asked to provide specific input data without revealing the results of the study. The results of the Cameroon study were deliberately withheld from the authors of this chapter until after they had completed work on the estimates with the AVERT model.

The analysis showed that AVERT estimates of the total number of HIV infections and the annual incidence rate matched quite well with the actual results of the cohort study: 73 estimated new infections compared to 78 actual new infections (Table 15-2). Because the number of infections observed in the study population could not be separated in terms of how many infections resulted from clients, as opposed to non-clients, we did separate runs for the different types of partners with whom the study subjects had sexual intercourse during the year of observation. Interestingly, the analysis suggests that almost 60 percent of the total

infections in the cohort of female sex workers were the result of sexual activity with non-clients, most likely a reflection of the different levels of condom use (reported number of unprotected sex acts with non-clients was 2.9 times higher than with clients).

PRACTICAL EXAMPLE FROM THE FIELD

The following example illustrates the various uses of the AVERT model in estimating the probable number of new HIV infections averted as a result of the actual as well as projected

intervention effects of a prevention program among mine workers and women at high risk in a South African mining community.

We have used AVERT to gain a better understanding of the impact of one of the first pilot programs delivering targeted periodic presumptive STI treatment in the developing world. Such treatment has been proposed as an option for reducing STIs in groups at high risk of infection (particularly high-risk women, who often experience no STI symptoms and may not seek treatment otherwise). The program offered free monthly examinations and presumptive STI treatment and counseling, combined with

TABLE 15-2 Estimated and observed HIV seroconversions in a female sex worker (FSW) cohort, Cameroon

Parameters used for AVERT

1	TARGET POPULATION	1,170 female sex workers, HIV seronegative at enrollment
2	PARTNER POPULATION	clients + non-clients, estimated HIV level: 13.0%
3	AVERAGE NUMBER OF SEX ACTS PER YEAR IN TARGET POPULATION AND REPORTED CONDOM USE	
	SEXUAL ACTIVITY	% OF SEX ACTS PROTECTED BY CONDOMS
	sex acts with clients = 287,820	96%
	sex acts with non-clients = 167,076	80%
	sex acts with clients + non-clients = 454,896	90%
4	ESTIMATED PREVALENCE OF GUD AND NON-GUD DURING STUDY PERIOD	
	genital ulcer lesions: 7.5%	
	inflammatory lesions: 15.0 %	

AVERT estimates

- Probable new HIV infections among FSW: 73 (estimated for 1,170 women years)
- Annual cumulative incidence of HIV: 6.24%

Study results

- Observed new HIV infections: 78 (adjusted for 1,170 women years)
- Observed seroconversion rate: 6.67%

community-based peer education on STI/HIV prevention, to women who trade in sex and others at high risk of STIs in a South African mining community where migrant employees live far away from their families for much of the year. All of the women who used the services were treated with a single-dose antibiotic for the most prevalent STIs in the area. Survey results showed that this approach was effective in reducing STIs, with substantial decreases in

STI prevalence among the women using the service and their miner partners after just nine months of intervention.

To analyze this pilot program in South Africa, we constructed scenarios based on reported behaviors and STI test results (Table 15-3). We assumed that the 400 women who used the STI treatment and counseling services regularly had had sexual contact with 4,000 miners living in the nearby hostels. HIV prevalence levels were

estimated at 50 percent for the women and 20 percent for the men. Because of the substantial in- and out-migration in the study populations, it was assumed that HIV prevalence rates would remain at this level in the next two years.

After nine months of intervention, investigators estimated that the overall prevalence of ulcerative STIs had dropped from 10 percent to 7 percent and non-ulcerative STI rates had fallen from 25 percent to 17 percent. We observed a 20-percent reduction in the reported number of clients from the miner population, and we estimated that the proportion of sex acts protected by condoms increased from 13 percent to 29 percent. Modeling these scenarios, AVERT estimated that the intervention had averted a total of 237 new HIV infections based on a 1-year time frame: 41 among the women and 196 among the miners.

We also used AVERT to estimate the impact of intervention effects (again calculated for a 1-year time frame) in a scenario in which the following project goals were achieved in the near future (Scenario 3 in Table 15-3): the reduced

level of commercial sex acts maintained, 50 percent condom use during commercial sex acts, and an 80 percent reduction in STI infection rates from baseline levels. The results of Scenario 3 presented in Table 15-3 show that the estimated annual cumulative incidence of HIV would decline from 52 percent to 12 percent among the women and from 13 percent to 2 percent among their miner clientele.

We used the potential intervention effects described in Scenario 3 as inputs to generate impact estimates of different intervention components. Table 15-4 illustrates that the estimated combined intervention effect (419 averted HIV infections) is not just the sum of the effects of the various single intervention components. The analysis does, however, confirm that targeted presumptive STI treatment was clearly the most effective single intervention component in the selected mining community.

TABLE 15-3 Modeling the impact of an intervention in South Africa

	Intervention Target Population FEMALE SEX WORKERS	Partner Population CLIENTS AMONG MINERS	
Population Size	400	4,000	
Estimated HIV Level	50%	20%	
	Scenario 1 PRE-INTERVENTION	Scenario 2 POST-INTERVENTION	Scenario 3 PROJECT GOALS ACHIEVED
Average annual partners: women	40	32	32
Average annual contacts per partner: women	10	10	10
Average annual partners: miners	4	3.2	3.2
Average annual contacts per partner: miners	10	10	10
Estimated prevalence of ulcerative STIs	10%	7%	2%
Estimated prevalence of non-ulcerative STIs	25%	17%	5%
Condom use	13%	29%	50%
AVERT estimates CALCULATED FOR A 1-YEAR TIME FRAME			
Estimated HIV infections: women	103	62	24
Estimated HIV infections: miners	405	209	65
Estimated HIV infection: total	508	271	89
Projected cumulative incidence: women	51.50%	31%	12%
Projected cumulative incidence: miners	12.66%	6.53%	2.03%

Estimated impact of different intervention components (using Scenario 3 from Table 15-3)

Partner reduction	STI reduction	Condom use	Probable infections averted
✓	✓	✓	419
•	✓	✓	398
✓	✓	•	367
•	✓	•	334
✓	•	✓	235
•	•	✓	172
✓	•	•	92

✓ Intervention effect as specified in Scenario 3 in Table 15-3
• Pre-intervention Scenario (Scenario 1 in Table 15-3)

TABLE 15-4

POINTS FOR CONSIDERATION

The close result in comparing the Cameroon cohort data with the AVERT estimates provides some encouraging validation of the model. However, this result should not be taken as an indication of the model's accuracy in all circumstances.

The structure of the AVERT model reflects several constraints that can influence the validity of the model's estimates. Our intent to keep the model user-friendly and accessible for available project data requires a certain level of conceptual and procedural simplicity. We attempted to achieve this goal while minimizing the loss of model strength. When considering whether to employ AVERT, potential users should keep the following points in mind:

AVERT's static nature—The most prominent limitation in AVERT is its static nature. This model is unable to incorporate change within the time frame considered and does not adjust to the dynamic character of variables and circumstances that are known to exist. Two sets of circumstances can be compared to each other, but only in static terms. This leads to an inability to: 1) account for individuals who became newly infected during the 1-year time frame and who are not removed from the pool of susceptibles, and 2) address the issue of secondary infections that result from those newly infected persons.

AVERT's utility as a program planning tool—Because the selected per sex act transmission probabilities represent time-weighted average figures, our model also may underestimate HIV transmission in settings with an explosive early epidemic and presumably a large proportion of people with a recently acquired infection (and hence increased infectiousness) at that time^{6,30}.

Nonetheless, AVERT can be a useful tool in the design or planning phase of a prevention program because it allows implementing organizations or individuals to estimate the potential impact of a planned intervention program. If project designers have a specific change in HIV incidence rates in mind, the model's outputs can be used to estimate the magnitude of behavior change and improved STI case management that need to be achieved. Conversely, if a particular level of behavior change and STI treatment coverage is envisioned, the designers can see, by incorporating the theoretical numbers into the model, the amount of change in HIV incidence that could potentially result.

In intervention settings with low HIV prevalence, the model may produce relatively small number of HIV infections averted. This may frustrate program managers who have high expectations for the impact of their interventions. A comparison of the estimated HIV incidence rates between the pre- and post-intervention scenarios, however, may provide more insight in the relative success of the prevention program.

AVERT's data requirements: availability and validity—As with any modeling exercise, the biggest challenge involved with AVERT is accurately estimating the input data needed to run the model. The user must find available data from a variety of sources and then derive the most plausible and realistic estimates. Sometimes data are unavailable and sometimes they exist but are unpublished. In other instances, data from different sources may conflict with one another or run counter to conventional wisdom. In addition, the validity of published data may be questionable when methodological constraints exist.

However, the data requirements for the AVERT model are relatively minor in comparison with some of the other promoted models².

The main purpose of these more complex simulation models is that they can be used by a multidisciplinary team to conduct complex sensitivity analyses, operations research, and hypothesis testing. In practical terms however, their complexity has greatly limited their use in evaluating specific intervention programs. In contrast, AVERT is similar to Epimodel³¹ in its ease of use and specificity to the task at hand. With only a brief introduction, a computer-literate user with some background in epidemiology can learn to operate AVERT within an hour or so.

health and/or other partners. STI and HIV data may come from national or regional surveillance systems, or from isolated studies conducted to gather prevalence and/or incidence data for various risk groups.

Although these data may not correspond directly to the populations being modeled, the user must assess the extent to which the data can be used to represent the populations they are interested in. In some instances users may need to adjust available data upward or downward to better represent the populations that users are modeling.

Because the validity of the model's estimates greatly depends on the quality of input data, the accuracy of these inputs will also determine whether AVERT may substantially over- or underestimate program impacts.

AVERT's data requirements: matching input data to interventions—AVERT is designed to estimate the number of sexually transmitted HIV-1 infections averted due to different types of interventions, assuming that the observed changes are entirely attributable to the modeled interventions. If the project in question was a behavioral intervention, then chances are that behavioral data needed to run the model may have been collected as part of the project. If the intervention was focused on reducing STIs, then users might have good STI data but no behavioral data. In either case, it is incumbent upon the user to find the best data to run the model. Behavioral data may come from surveys conducted with specific target groups, or from behavioral surveillance surveys conducted by the ministry of

Because the validity of the model's estimates greatly depends on the quality of input data, the accuracy of these inputs will also determine whether AVERT may substantially over- or underestimate program impacts. Given the uncertainty of the input data, we recommend running the model with different likely values. For example, users could include the upper and lower values of the 95 percent confidence intervals from survey data. Such an approach allows users to conduct a sensitivity analysis and they may find that it is useful to present results as a range of model outputs.

Targeted periodic presumptive STI treatment was probably the most effective single intervention component in the studied South African mining community. While this conclusion is

relatively valid for the setting considered, one should keep in mind that AVERT does not take into account the decrease in STI rates that might result from condom use and/or partner reduction alone. The model uses STI rates as an input variable for calculating probable HIV transmissions but does not quantify the attributable effects of condom use, partner reduction, and STI treatment on STI transmission and observed changes in STI prevalence.

AVERT's time frame—AVERT estimates are usually based on a 1-year time frame. Because HIV infections through sexual intercourse are the result of accumulated risk exposures, preventing an infection in one year will not ensure that a person remains uninfected in subsequent years. Model estimates of HIV infections averted should be interpreted cautiously, especially in populations with high-risk behaviors where the observed behavior changes suggest that the interventions may only postpone infections rather than prevent them indefinitely.

CONCLUSION

In interpreting how an intervention will work, models cannot provide simple, exact quantitative predictions with absolute certainty. Models are, by definition, tools for analysis under conditions of uncertainty. They help to identify potentially effective interventions, translate the outcomes of these interventions into impact, and suggest designs for future interventions. With a small number of accessible input variables, AVERT can provide plausible and defensible impact estimates of intervention effects on the reduction of HIV transmission. This straightforward model provides an additional analytical tool for epidemiologists, decision makers, and planners in setting appropriate program priorities. In addition,

AVERT may enable program managers to carry out cost-effectiveness analyses of intervention programs that were tailored for specific target populations in various epidemiologic settings.

ACKNOWLEDGMENTS

The authors thank Richard Steen for testing the model in South Africa and Leopold Zekeng for making data from Cameroon available for validation.

REFERENCES

- 1 Rehle T, Saidel T, Hassig S, et al. AVERT: a user-friendly model to estimate the impact of HIV/sexually transmitted disease prevention interventions on HIV transmission. *AIDS* 1998;12(Suppl 2):S27-S35.
- 2 Bernstein RS, Sokal DC, Seitz ST, et al. Simulating the control of a heterosexual HIV epidemic in a severely affected east African city: computational modeling of single vs. combined intervention strategies with the iwgAIDS and simulAIDS models. *Interfaces* 1998;28:101-126.
- 3 Weinstein MC, Graham JD, Siegel JE, Fineberg HV. Cost-effectiveness analysis of AIDS prevention programs: concepts, complications, and illustrations. In: Turner CF, Miller HG, Moses LE, editors. *Confronting AIDS: sexual behavior and intravenous drug use*. Washington (DC): National Academy Press; 1989.
- 4 Cates W, Stewart FH, Trussell J. Commentary: the quest for women's prophylactic methods—hopes vs. science. *Am J Public Health* 1992;82(11):1479-1482.
- 5 Feldblum P, Morrison C, Roddy R, Cates W Jr. The effectiveness of barrier methods of contraception in preventing the spread of HIV. *AIDS* 1995;9(Suppl A):S85-S93.
- 6 Mastro TD, deVincenzi I. Probabilities of sexual HIV-1 transmission. *AIDS* 1996;10(Suppl A):S75-S82.
- 7 Padian NS, Shiboski SC, Glass SO, Vittinghoff E. Heterosexual transmission of human immunodeficiency virus (HIV) in northern California: results from a ten-year study. *Am J Epidemiol* 1997;146(4):350-357.
- 8 Hayes RJ, Schulz KF, Plummer FA. The cofactor effect of genital ulcers on the per-exposure risk of HIV transmission in sub-Saharan Africa. *J Trop Med Hyg* 1995;(1)98:1-8.
- 9 Duerr A, Xia Z, Nagachinta T, et al. Probability of male-to-female HIV transmission among married couples in Chiang Mai, Thailand. From: Xth International Conference on AIDS. Yokohama, Japan; 1994. Abstract no.: 105C.
- 10 Wiley JA, Herschkorn SJ, Padian NS. Heterogeneity in the probability of HIV transmission per sexual contact: the case of male-to-female transmission in penile-vaginal intercourse. *Stat Med* 1989;8(1):93-102.
- 11 Satten GA, Mastro TD, Longini IM Jr. Modelling the female-to-male per-act HIV transmission probability in an emerging epidemic in Asia. *Stat Med* 1994;13(19-20):2097-2106.
- 12 Mastro TD, Satten GA, Nopkesorn T, et al. Probability of female-to-male transmission of HIV-1 in Thailand. *Lancet* 1994;343(8891):204-207.
- 13 Cameron DW, Simonsen JN, D'Costa LJ, et al. Female to male transmission of human immunodeficiency virus type 1: risk factors for seroconversion in men. *Lancet* 1989;2(8660):403-407.
- 14 Downs AM, deVincenzi I. Probability of heterosexual transmission of HIV: relationship to the number of unprotected sexual contacts. *J Acquir Immune Defic Syndr Hum Retrovirol* 1996;11(4):388-395.
- 15 DeGruttola V, Seage GR III, Mayer KH, Horburgh CR Jr. Infectiousness of HIV between male homosexual partners. *J Clin Epidemiol* 1989;42(9):849-856.
- 16 Pinkerton SD, Holtgrave DR, Valdiserri RO. Cost-effectiveness of HIV-prevention skills training for men who have sex with men. *AIDS* 1997;11(3):347-357.
- 17 Laga M, Monoka A, Kivuvu M, et al. Non-ulcerative sexually transmitted diseases as risk factors for HIV-1 transmission in women: results from a cohort study. *AIDS* 1993;7(1):95-102.
- 18 Kreiss J, Willerford, Hensel M, et al. Association between cervical inflammation and cervical shedding of human immunodeficiency virus DNA. *J Infect Dis* 1994;170(6):1597-1601.
- 19 Dickerson MC, Johnston J, Delea TE, et al. The causal role for genital ulcer disease as a risk factor for transmission of human immunodeficiency virus. An application of the Bradford Hill Criteria. *Sex Transm Dis* 1996;23(5):429-440.

REFERENCES

- 20 Plourde PJ, Pepin J, Agoki E, et al. Human immunodeficiency virus type-1 seroconversion in women with genital ulcers. *J Infect Dis* 1994;170(2):313-317.
- 21 Kassler WJ, Zenilman JM, Erickson B, et al. Seroconversion in patients attending sexually transmitted disease clinics. *AIDS* 1994;8(3):351-355.
- 22 Deschamps M-M, Pape JW, Hafner A, Johnson WD. Heterosexual transmission of HIV in Haiti. *Ann Intern Med* 1996;125(4):324-330.
- 23 Mbizvo MR, Machezano R, McFarland W, et al. HIV seroincidence and correlates of seroconversion in a cohort of male factory workers in Harare, Zimbabwe. *AIDS* 1996;10(8):895-901.
- 24 Mehendale SM, Rodrigues JJ, Brookmeyer RS, et al. Incidence and predictors of human immunodeficiency virus type 1 seroconversion in patients attending sexually transmitted disease clinics in India. *J Infect Dis* 1995; 172(6):1486-1491.
- 25 Celentano DD, Nelson KE, Suprasert S, et al. Risk factors for HIV-1 seroconversion among young men in Thailand. *JAMA* 1996;275(2):122-127.
- 26 Plummer FA, Simonsen JN, Cameron DW, et al. Cofactors in male-female sexual transmission of human immunodeficiency virus type 1. *J Infect Dis* 1991;163(2):223-239.
- 27 Roddy RE, Zekeng L, Ryan KA, et al. A controlled trial of nonoxynol-9 film to reduce male-to-female transmission of sexually transmitted diseases. *N Engl J Med* 1998; 339(8):504-510.
- 28 Tapko JB, Kouinche A, Guiffo L, Kaptue L. Evolution of HIV seroprevalence among blood donors in Cameroon from 1990 to 1994. From: XIth International Conference on AIDS. Vancouver, British Columbia; 1996. Abstract no.: Pub-C-1135.
- 29 Family Health International/AIDSCAP Project. AIDSCAP project final report 1997. Arlington (VA): Family Health International; 1997.
- 30 Vernazza PL, Eron JJ, Fiscus SA, Cohen MS. Sexual transmission of HIV: infectiousness and prevention. *AIDS* 1999;13(2):155-166.
- 31 Chin J, Lwanga S. Estimation and projection of adult AIDS cases: a simple epidemiological model. *Bull World Health Organ* 1991;69(4):399-406.

16

GUIDELINES FOR ASSESSING THE ECONOMIC AND FINANCIAL COSTS OF HIV/AIDS PREVENTION AND CARE PROGRAMS

STEVEN FORSYTHE

Defining the Question	237
Planning and Budgeting	238
Sustainability	238
Economic Evaluation	238
Collecting Data on Costs	239
Macroeconomic Indicators	239
General Principles	240
Recurrent Costs	241
Capital Costs	243
Technical Assistance Costs	245
Categorizing Costs	245
Economic versus Financial Costs	245
Defining Costs	246
Preparing Technical Reports	248
Cost Analysis	248
Cost Projections	248
Cost Recovery	249
Conclusion	250
References	250

16

GUIDELINES FOR ASSESSING THE ECONOMIC AND FINANCIAL COSTS OF HIV/AIDS PREVENTION AND CARE PROGRAMS

This chapter provides guidelines for assessing the economic and financial costs of HIV/AIDS interventions. In turn, policymakers will use this information to assure that fully informed decisions are made regarding the cost-effective use of limited resources for HIV/AIDS care and prevention.

These guidelines are intended to allow the reader to understand the basic concepts of cost analysis and to actually perform an assessment. The guidelines will walk the reader through the process of:

- defining the question that needs to be addressed;
- collecting cost information;
- categorizing costs in a way that will focus the analysis on the relevant questions; and
- writing a technical report that lays out the cost of an HIV/AIDS intervention, projects future costs, and identifies possible means of cost recovery.

These guidelines are intended to provide a rapid overview of the issues and approaches that need to be considered in performing a cost

analysis. More detailed guidelines have been developed for assessing the costs of health¹, family planning² and HIV/AIDS³ programs.

DEFINING THE QUESTION

Accurate and full cost information is crucial for evaluating the current level of resource use and assessing opportunities for future expenditures on the control of the AIDS pandemic. However, cost analyses are not designed to answer all questions with a single number. In fact, the approach selected for performing a cost analysis is highly dependent on the type of questions being asked. For example, an analysis designed to evaluate all resources required to carry on an intervention will likely be performed very differently from an economic evaluation designed to identify the

incremental resources required to add HIV/AIDS services to an existing clinic.

Most cost analyses are designed to address at least one of the following types of questions:

- **Planning and budgeting**—What are the current and future financial requirements of each project?
- **Sustainability**—What is required to sustain each project?
- **Economic evaluation**—What is the cost per unit of output or outcome produced from each project?

Planning and Budgeting

One possible objective in performing a cost analysis is to plan and budget for current and future resource requirements. A well-performed cost analysis will permit decision makers to perform more accurate planning and develop new budgets. It will also allow project planners to project the financial requirements necessary to continue, expand, or replicate an ongoing project.

Sustainability

Program planners are frequently asked to ensure that their interventions can be sustained beyond the current funding cycle. One key to sustaining interventions is to estimate and plan for the costs of programs over time¹. A sustainability plan can include data about the cost of services and can identify feasible opportunities for recovering costs. For example, a cost analysis of a sexually transmitted infections (STI) clinic can help program planners to identify the proportion of project costs that can be recovered by charging various prices for services.

Financial sustainability does not necessarily require that clients pay for the full cost of the services they are receiving. A sustainability plan might include the development of financial resources from any combination of sources, including clients, the government, the private

sector, and international donors. For example, a country that wants to expand its blood banking services to ensure that the entire blood supply is screened for HIV, could perform a cost analysis to accurately ensure the start-up and recurrent costs that would be required to sustain such an endeavor. In addition to financial sustainability, interventions need political commitment and institutional capacity to sustain their operations. From the perspective of the consumer (for example, the user of a socially-marketed brand of condom), there also needs to be a sustainability of consumption to assure that uptake of services is maintained.

Economic Evaluation

A cost analysis is a critical tool for performing various types of economic evaluations of HIV/AIDS interventions, such as cost-effectiveness, cost-utility, and cost-benefit analyses. The analysis collects estimates as to the annual cost of the interventions in terms of economic requirements. The economic assessment also provides information so that changes in project costs over time can be projected.

In addition to evaluating economic inputs, it is also necessary to collect outputs that correspond with specific activities. Outputs can include items such as the number of condoms distributed, the number of patients treated for STIs, or the number of pamphlets distributed.

Cost-effectiveness analysis (CEA), covered in greater detail in the next chapter, is one technique to evaluate interventions by combining project costs and outcomes. When used for health care planning, CEA can help policymakers and program managers examine alternative ways of achieving a given objective and enable them to select the method that uses available resources most effectively and efficiently.

CEA is used primarily as a decisionmaking tool to help policymakers and program managers select a future course of action. Although it

often is used to evaluate completed programs, the ultimate use of this research method should be to help decision makers decide what to do in the future. The next chapter will provide a detailed methodology for conducting cost-effectiveness analysis of HIV intervention programs.

Cost analyses are also necessary for performing a cost-benefit analysis. Cost-benefit analysis take cost-effectiveness analysis one step further by putting both the numerator (cost) and the denominator (benefits) in monetary terms.

COLLECTING DATA ON COSTS

This section provides practical guidelines that are intended to help project planners systematically collect and analyze project costs. Methods for estimating and reporting project costs are presented throughout the rest of this section. Once these project costs are collected, a technical report can be written that incorporates the findings and conclusions derived from the cost analysis. The later section in this chapter, “Preparing Technical Reports,” provides an overview of technical report writing using cost data.

In the longer term, economists should make collecting local costs a routine matter by incorporating quarterly cost reports into the requirements of project agreements. By doing this, the process will become more efficient and sustainable.

Macroeconomic Indicators

The first step in performing a cost analysis involves collecting and presenting macroeconomic data. This information will be used in the cost analysis to project future costs and to annualize capital costs. These macroeconomic indicators should be explicitly presented in the technical report.

Inflation

The current consumer price index should be presented in the technical report, along with

available predictions concerning the future inflation rate. If future inflation projections are not available, the average rate over the previous five years should be used as the best estimate of the future inflation rate.

A cost analysis is a critical tool for performing various types of economic evaluations of HIV/AIDS interventions, such as cost-effectiveness, cost-utility, and cost-benefit analyses.

Exchange Rates

Official and black market exchange rates should be included in the technical report, as well as the dates and source of the data. Official exchange rates may be artificial (not reflecting the true value of foreign exchange) and significantly different from black market exchange rates. Therefore, black market exchange rates are occasionally used in performing an economic cost analysis. Their use will provide better cost estimates when converting costs for analysis and comparison purposes.

Discount Rates

The nominal and real interest rates (the nominal interest rate minus the inflation rate) should be provided in the technical report. The real interest rate for a long-term treasury note should be used to annualize capital costs because it will reflect the current annual return at which money can be invested in a secure financial note.

Possible sources for a discount rate include the country's Ministry of Health, Ministry of Finance, local banks, and the World Bank. A range of interest rates should be collected for the purpose of performing a sensitivity analysis of the results.

Property Values

Information should also be presented in the cost analysis that summarizes the average cost of land and building space. This will only be necessary if the intervention is located within a building for which there is no estimated replacement value.

General Principles

In order to ensure consistency, reliability, and comparability of all data, it is necessary to collect cost data based on several general principles.

Collecting Full Cost Data

It is necessary that all resources, including resources provided in-kind, that are directly or indirectly used by the project be recorded and assigned a value by the economist. There are a number of reasons for assigning a value to in-kind resources. First, measuring the full cost permits the economist and project planners to identify the full resource requirements of the effort and provides the information needed to plan for long-range financing. For example, condoms and other in-kind inputs represent limited resources that may not be provided for free in the future. Finally, for evaluation purposes, it is necessary to assess all resources used in carrying out activities. An economic appraisal of two different intervention programs or two different activities within a program will be meaningless if based only on partial cost information.

The monetary value of all in-kind resources (material or human) should be estimated based on the opportunity costs of those resources. An opportunity cost represents the amount that one would have to sacrifice in order to obtain the

particular good or service, regardless of whether or not it is actually paid for.

Partially or fully subsidized rental space, administrative services, materials, or equipment should be valued at the price at which they could be purchased on the open market. For example, condoms that are received at no cost from USAID have been valued at \$0.045 per piece (not including shipping and storage). Resources such as volunteer labor should also be valued in the cost analysis. The value of volunteer labor should be based on the current wage rate for similar activities.

Including and Excluding Taxes

Taxes, duties, and other transfers from the project to the government are financial costs to the project and therefore should be included in the cost finding forms. However, it is important to separate the purchase price of items from the tax for that asset on the cost finding sheets. In performing an economic cost analysis, taxes should be excluded because they represent a transfer of funds rather than a true economic cost.

Allocating Joint Costs

Joint costs represent resources that are used to carry out more than one activity. Most interventions involve the use of some resources that are shared across multiple activities or services. The following are examples of joint costs that have to be allocated across activities or services:

- The salary of a receptionist who greets patients at a health center and directs patients to an STI clinic or various other health services.
- The cost of electricity at an STI clinic that performs both STI treatment and outreach activities.
- The equipment in a blood bank that is used to run the equipment for identifying both STIs and HIV.

In the first example, it would be inappropriate to fully assign the receptionist's salary to the STI clinic. Similarly, it would be inappropriate to fully exclude the receptionist's salary from the cost of providing STI services. Therefore it is necessary to allocate a portion of the receptionist's salary to the STI clinic through some allocation technique.

Determining how to allocate joint costs across projects or activities can sometimes be difficult. Below are some possible methods of allocating joint costs. The method used for allocating costs should be carefully described and justified within the technical report. In most cases, there is no definitive rule concerning methodologies for allocating joint costs, and it is therefore necessary to use the most logical approach available.

- One method to allocate labor is to observe resource utilization and to allocate costs accordingly. For example, the analyst can observe the accountant in a hospital to determine how much time is actually spent on the accounts of the STI clinic that is being costed. While this method is the most accurate way of allocating joint costs, it is also one of the most difficult and time-consuming.
- Indirect costs may be allocated based on the amount of space used (area of rooms used for the intervention) by the project being analyzed. For example, in order to allocate the use of building space, it may be possible to divide the building into areas that are predominantly used for the project being surveyed and to allocate costs based on the amount of space used. An STI clinic that comprises 10 percent of a hospital might therefore be assigned 10 percent of the costs of utilities and rent.
- Indirect costs can also be allocated based on the number of clients seen. For example, consider a voluntary counseling and

testing (VCT) clinic that has 100 clients per month that is located within a health center, which has a total of 500 clients per month. In this case, 20 percent of the receptionist's salary could be allocated to VCT services, because 20 percent of the clients are attending the health center for VCT services.

- Indirect costs may also be allocated in the same proportion as the direct costs. Thus, it would be necessary to add up the costs of all direct costs and determine what proportion has been assigned to the project being studied. This percentage would then be used to allocate indirect costs. For example, if the budget of an STI clinic comprised 30 percent of a hospital's operating budget, it may be possible to allocate 30 percent of the hospital's indirect costs to the STI clinic.
- Finally, it is possible to simply divide costs evenly among ongoing projects. This is the least accurate method of allocating costs and should only be used when no other alternative is feasible.

Documenting All Information

All sources of information for cost estimates and specific calculations should be well documented in the technical report.

Recurrent Costs

Recurrent costs value resources that are expected to be consumed (or replaced) within a year. The following instructions indicate how individual recurrent costs should be collected and recorded in the cost finding sheets.

Salaries and Benefits

First, identify all personnel who are directly or indirectly connected in any way with the program. Include clerical and maintenance staff, volunteers, and consultants.

Second, calculate the amount of time each person spends on the program during the study period. Because most employees work on more than one project, it will be necessary to allocate time to the various projects based on interviews and through observation. If any employee (such as maintenance staff, administrators, or accountants) works indirectly on the project, their labor time should be allocated based on one of the methods described under “Allocating Joint Costs” in this section.

Third, determine the compensation paid to each worker. The figures may be listed as annual rates; however, in some cases, they will be noted as hourly, daily, weekly, or monthly rates.

Sixth, identify the value of any income taxes paid. This should be included in the financial analysis but may be excluded from the evaluation of economic costs.

Finally, sum all personnel costs to obtain the total personnel costs.

Materials and Supplies

First, list all of the relevant materials and supplies directly or indirectly used in the program during the study period. Second, determine the amount of materials and supplies used during this period. One way to do this is to conduct an inventory of all relevant materials and supplies at the beginning of the time period and at the end and subtract

The distinction between recurrent and capital costs is life expectancy.

Items that are consumed or replaced within a year are categorized as recurrent costs, and items with a life expectancy of more than a year are categorized as capital costs.

Fourth, multiply the amount of time each person spends on the program by the compensation figures to compute the cost for that person.

Fifth, add the cost of fringe benefits if these are not included in personnel costs. Employer contributions to medical plans, insurance plans, social security, and other benefits are sometimes computed as a percentage of total salary. Some fringe benefits have to be computed separately. Consultants are usually paid an hourly or daily rate, or sometimes a flat honorarium. They do not usually receive fringe benefits.

the ending from the beginning inventories to determine the amount used. Third, determine the cost of each item used. Finally, multiply the quantity of each item used by its cost and sum the total.

Travel and Associated Expenses

This category includes all transportation and per diem costs attributable to the intervention, whether for transporting people or supplies. Travel and transportation often make up a significant proportion of the costs of outreach services, supervision, and training. They may

also account for a significant portion of the costs of imported supplies, materials, and equipment.

Operation and Maintenance

Buildings—The ongoing costs of running and maintaining a facility will include all building maintenance functions (such as repairs, house-keeping) and all utility payments (such as telephone, electricity, water, gas).

If operation and maintenance costs are not available, these costs can be calculated as a percent of the building’s value. This percentage will vary, however, depending on the quality of construction materials, the age of the building, and the nature of the services that are delivered within the facility. If the facility is heavily used, then maintenance and operation costs will be higher. In general, operation and maintenance costs range from 2.5 to 4 percent of the building’s investment cost, although figures of 8 to 9 percent may be applicable when facilities are in poor condition, heavily used, or both.

Equipment—The estimated percentage of total costs for required operation and maintenance of equipment increases with the equipment’s complexity. When possible, the cost of operating and maintaining equipment should be itemized. In general, office furniture and general office maintenance has been estimated to be approximately 5 percent of investment cost per year; office equipment and audiovisual aids are approximately 10 percent; and laboratory and technical equipment vary between 10 percent and 20 percent of investment costs. These estimates, based on World Health Organization (WHO) and World Bank projects, will vary depending on circumstances and intensity of use.

Vehicles—To determine the cost of vehicle operation and maintenance, it is necessary to identify the actual expenditures for gasoline, oil, lubrication services, and other expenses. If the vehicle is being used for more than one project,

the costs should be allocated based on estimates of the number of kilometers traveled for each project. Studies by the World Bank have recommended the following calculations for determining operating and maintenance costs:

- Gas = # km traveled/month X 12 months X mileage (litres/100 km) X unit cost/liter of gasoline
- Oil = 15% times annual expenditures on gasoline
- Spare parts and labor = 24% times purchase price of vehicle
- Insurance = Varies greatly by country; check with ministry of health or USAID

Studies suggest that the operating and maintenance costs of motorcycles are 40 percent of the purchase price and mopeds are 20-30 percent of the purchase price.

Capital Costs

The distinction between recurrent and capital costs is life expectancy. Items that are consumed or replaced within a year are categorized as recurrent costs, and items with a life expectancy of more than a year are categorized as capital costs. The cost of a capital item is spread over its expected life. The analyst must calculate the portion of the cost that should be annually attributed to the project, as follows:

- Identify all capital cost items that are expected to be used during the study period.
- Determine the current market value of each asset, including the cost of transporting the asset to the site. This information can be obtained by contacting manufacturers or reviewing catalogs.
- Compute the annualized cost of each item (using the estimated life expectancy of the resource, the prevailing interest rates, and annualization tables).

- Determine the proportion of the item to be charged to the program.
- Sum the total.

Buildings

The cost of space is frequently the largest capital cost and should be carefully calculated to be certain that a standard approach is being used. This calculation will involve making cost estimates of the building. In calculating building costs, it should first be determined whether the space is rented or owned.

Rented space—The rental cost, including any rental fees that are incurred, should be obtained. Because the rental cost can be used as a proxy for the annualized cost of a building that is owned, it is not necessary to annualize the rental cost. Rent should be classified as “rented/leased capital.”

Owned space—For a building that is owned, the replacement value of the building should be estimated. If a cost cannot be obtained for the specific building, it should be possible to develop an estimate based on the size of the building and the cost per square meter of comparable buildings that are for sale in the community. Estimates for number of square meters and for the unit cost per square meter should be made in consultation with architects in the Ministry of Health or the Ministry of Public Works. Unit cost per square meter may also be verified with local construction firms and with bilateral and multilateral organizations that have recently completed or are in the process of financing projects with similar civil works components. It is important to note that the analysis is concerned with the economic cost of the building, so even if an occupant neither pays rent nor owns the building, a cost still needs to be assigned.

Equipment

The current replacement value of equipment should be calculated. This can be obtained with

information about the type of equipment, the manufacturer, and the model.

Furniture

Furniture costs should be itemized wherever possible. However, if no other source of cost information is available, the capital cost of furniture may be broadly estimated at 8 percent of the construction cost of the building.

Vehicles

The cost of vehicles should be estimated on the basis of whether or not the items will be purchased locally or imported. Quotes (including import duties) should be solicited from potential suppliers and verified by the Ministry of Health. All import duties and taxes should be included in the analysis under a separate line item.

Annualizing Capital Costs

Value of Assets—The value of assets will be estimated by determining the current market value. The use of current market value will facilitate the data collection process because it is not necessary to determine historical prices.

Life Expectancy—The analyst should identify the life expectancy of an existing asset by determining the **remaining** life of the existing item. For highly technical items, such as medical equipment, the life expectancy is frequently limited by technological obsolescence. This should be carefully considered in determining an appropriate life expectancy.

There is no clear-cut rule for establishing the useful life of assets. In general, it is recommended that Table 16-1 be used as a guide.

Annualization Factor—To compute the annualized cost of a capital item, an annualization table is necessary. As an example, assume a vehicle has recently been purchased for the exclusive use of a project. The replacement value of the vehicle is \$10,000, with a 7-year life expectancy and a 5 percent real interest rate. The annual cost can

be found in the table by finding the interest rate column ($r = 5\%$), locating the annualization figure for the expected remaining lifetime of 7 years (this is 0.173), then multiplying the current replacement value of the vehicle by this annualization factor: $\$10,000 \times 0.173 = \$1,730$.

Technical Assistance Costs

In order to place the cost of technical assistance in a comparable format with the capital and recurrent costs incurred in the field, the technical assistance cost should be annualized over the expected length of the intervention. Thus, if a donor’s in-house costs were \$100,000 over 3 years and the real annual interest rate were 15 percent, the annual cost would be $\$100,000 \times 0.438 = \$43,800$.

CATEGORIZING COSTS

This section defines the categories of costs that will be applied in the cost finding forms. The first step involves deciding if economic or financial costs will be calculated. The second step involves actually categorizing costs.

Economic versus Financial Costs

Before beginning data analysis, it is necessary to determine if the cost analysis is going to focus on economic or financial costs. Financial costs are those narrowly defined as having been paid for by the project. This could include such things as salaries and materials, but not donated items. These costs are particularly relevant when trying to determine the payments required to keep an intervention operational. Implicit in financial costs is the assumption that items that are donated (by the government, international donors, volunteers, or others) will continue to be donated into the indefinite future.

An economic cost analysis involves a broader evaluation of resources used, regardless of who

actually pays for those resources. Economic costs include financial costs, but also values resources that are obtained from other sources. An analysis of economic costs would include items such as donated condoms or volunteer labor. This information is particularly useful when evaluating the opportunity cost of an intervention or when comparing interventions.

One pitfall in performing a cost analysis occurs when two interventions are compared using a financial cost analysis. Imagine an inefficient and expensive intervention that is highly subsidized by either the government or an international donor. In contrast, a second intervention produces significant “value for money,” but does not receive any outside support. An economic analysis would correctly demonstrate that the latter intervention was the less costly intervention. However, a financial analysis might conclude that the first intervention was less costly, simply because the costs of

Guide to estimated life expectancy of assets

TABLE 16-1

Resource	Life Expectancy
Building	30 years
Furniture	15 years
Vehicle	7 years
Laboratory Equipment	
• Mid- and big-ticket equipment	10 years
• Small equipment (less than \$1,500)	3 years
Other equipment	
• Office (including computers)	5 years
• Audiovisual (including TV/VCR)	5 years

carrying out the intervention were being incurred by the government or the donor.

Defining Costs

As already indicated, it is necessary to determine whether the analysis is trying to address issues of budgeting, planning, sustainability, or economic evaluation. This step of defining a question is critical because it will assist the evaluator in defining what costs should be included in the analysis.

Each resource used to carry out an intervention should be identified and categorized. Even though resources may be subsequently removed from the main focus of the cost analysis, it is still important to identify each of these resources.

Once a resource has been identified, it is necessary to ask the following questions about each item.

- What is the functional classification of this resource?
- Is this a start-up cost or a recurrent cost?
- Is the cost fixed or variable?
- Is it a direct or an indirect cost?
- Who paid for this resource?

Functional Classification of Costs

A prerequisite to the standard collection of cost data is establishing proper functional categories for each project. A critical step in creating a useful costing system is to define clear program categories that allow project evaluation by project components or activities. Cost and output data collection by activities will allow cost-effectiveness analysis and comparisons between alternative programs and activities within a program. This will also facilitate planning, budgeting, and monitoring tasks by meaningful program components.

All costs should be assigned to a specific programmatic activity (such as lab support, condom distribution, data collection, evaluation, training, production and distribution of material, STI diagnosis and treatment, or health education). When categorized in this manner, the activity

will have its own measure of costs and its own measure of output.

Start-up versus Recurrent Costs

Costs can be categorized as either start-up or recurrent costs. Start-up costs involve the purchase of goods or services during the initial stage of an intervention. This can include capital items (resources with a life expectancy of more than one year), research, consultancies, and the design of information, education, and communication (IEC) materials. Recurrent costs are those required for the normal operation of the intervention.

The distinction between start-up and recurrent costs is important because:

- Some donors limit their contribution to start-up costs and expect the host country to be responsible for the recurrent costs.
- In many countries, there is one budgeting and accounting process for start-up costs and another for capital costs.

Fixed versus Variable Costs

Costs that remain unchanged despite small fluctuations in output are fixed costs. Thus, a blood bank that takes one additional blood sample would probably not incur additional staff or equipment costs. Items such as a building, salaries of permanent staff and laboratory equipment are generally considered to be fixed costs.

Variable costs, in contrast, fluctuate directly with changes in the related total activity or volume. Thus, an STI clinic would probably incur variable costs for such items as swabs and medications because the need for these items increases directly with the number of patients treated. Variable costs might include items such as medications or materials.

This distinction between fixed and variable costs is important because the cost of expanding a project in the short run will depend on the variable costs but not the fixed costs. The

distinction, however, is also somewhat subjective. For a clinic that is currently working at 100 percent capacity, the addition of one additional patient might require that another doctor be hired, or that another item of equipment be purchased. To sufficiently address the costs of expanding a program, the analyst should also look at other relevant factors, such as space constraints, trained labor availability, and sufficiency of demand.

Direct versus Indirect Costs

In performing a cost analysis, it is a common mistake to omit costs that are only indirectly attributable to an intervention. However, it is critical that both direct and indirect costs be included in the cost analysis. Direct costs represent resources that are directly attributable to the service being provided. For example,

water and electricity bill of that same hospital needs to be considered as a component to the cost of providing STI services.

Indirect labor costs may include the salaries for accountants, administrators, secretaries, and maintenance employees. Indirect capital costs may include buildings, office equipment, and furniture. The approaches for allocating a portion of indirect costs to a particular intervention have already been discussed in this chapter under “Allocating Joint Costs.”

Funding Source

Project activities may be funded by international donors or through other local sources, such as local institutions or the private sector. Resources provided in-kind will be identified under “donated goods and services.”

In performing a cost analysis, it is a common mistake to omit costs that are only indirectly attributable to an intervention. However, it is critical that both direct and indirect costs be included in the cost analysis.

the direct costs of an intervention with sex workers may consist of personnel directly involved in the intervention (including volunteer time), condoms, and transportation directly attributable to the intervention.

Indirect costs represent resources that are only indirectly attributable to the intervention. For example, the salary of a security guard at a hospital that provides STI services is an indirect cost that must be included in an analysis of the cost of providing STI services. Similarly, the

In-house costs incurred by donors for providing technical assistance to the project need to be collected separately. These costs should be collected and disaggregated into two time periods: 1) from project approval to the initiation of the intervention, and 2) from project initiation until the conclusion of the technical assistance. In this way it is possible for donors to separate their initial investment from the resources that are required for actually maintaining the project over time.

Items that are currently provided to the field at no cost to either the donor or the in-country counterparts (for example, voluntary labor and free rent) should be categorized as donated resources. Because an analysis will collect the full cost of all resources, the monetary value of these “free” resources will be estimated (an explanation on how to value donated resources was provided in the section on “Collecting Full Cost Data”).

PREPARING TECHNICAL REPORTS

A technical report generally includes at least three parts. The first, the cost analysis, concerns the actual assessment of costs. This includes an initial assessment of all resources used to carry out the intervention, along with a more narrow assessment of specific costs that are relevant for addressing the particular questions that need to be addressed.

The second part involves developing cost projections. This part will describe the resources required to assure that an intervention is sustained into the future. It will take into consideration changes in the price of resources as well as possible contingencies (such as salary increases).

The final part should identify possible opportunities for cost recovery. This might include an assessment of how much clients might be willing to pay for services. However, such an assessment should also evaluate opportunities for sustaining interventions through resources from the public and private sector and donors. Furthermore, this part of the technical report might be used to evaluate ways in which the cost of services might be reduced.

Cost Analysis

The first part of the report will include a clear statement of the questions that are being addressed in the analysis. This part should include justification for the approach taken, as

well as an explanation of all assumptions. The report should ideally provide the cost finding forms that were used to collect data.

The analysis should begin with an estimate of the cost of all resources used in implementing the project. This should be followed by a description of costs that were excluded in addressing specific issues. For example, an analysis of the costs to add voluntary counselling and testing to existing health centers will exclude the cost of indirect costs that would have been paid for regardless of the introduction of the intervention.

Cost Projections

The technical report should include estimates of changes in costs five years into the future. While cost projections are likely to have a potential for error, such estimates will provide some indication as to the potential future growth and financial needs of the project.

In order to make cost projections, it is necessary to make assumptions about the future.

The assumptions should be **clearly and carefully** stated in the technical report. The assumptions concerning cost projections will depend on the expected future growth of the project.

The technical report should identify alternative future scenarios and the expected cost under each scenario. In addition, it will be necessary to project future outputs under the various assumptions. Recommendations for possible scenarios under targeted sexual interventions are presented below. Cost projections should be made for all scenarios and the most likely growth path should be identified in the technical report.

One scenario should assume that the existing project continues to reach the current target audience with essentially no growth. Cost projections should include expected price changes for project inputs (beyond the national inflation rate), the usual recurrent costs, and the replacement of expired capital items.

The second scenario will involve an increase in the target population without an expansion in the regions covered. Cost projections will increase due to an increase in the size of the target audience (for example, reaching more commercial sex workers in a city), but not an increase in the size of the region involved. Thus, for a condom distribution program, such a scenario would involve increased spending for variable costs, such as condoms and materials, but only moderate or no growth for fixed costs, such as permanent staff.

Finally, the third scenario would involve an increase in the target audience within the existing region and an expansion into other regions. To obtain data for this scenario, it is necessary that project personnel realistically determine into which regions the project could reasonably expand. This scenario will probably involve significant increases in both fixed and variable costs.

Price Changes

All cost projections should be presented in constant currency. If information is available indicating that certain cost items will increase in value at rates different from the national inflation rate, the costs should be adjusted appropriately to reflect the expected real price increase.

For example, assume that an intervention is being assessed that used 1,000 pesos worth of fuel in the current year. In the next year, it is expected that the national inflation rate will rise at a rate of 10 percent but fuel is expected to increase by 35 percent. Fuel prices should then be adjusted by the expected price increase less the inflation rate (35 percent minus 10 percent). In this case, fuel costs should then be adjusted as follows:

- a) Current project cost of fuel (1999): 1,000 Pesos*
- b) Expected fuel price increase (2000): 35%

- c) Expected national inflation rate (2000): 10%
 - d) Adjusted fuel cost (2000): 1,250 Pesos*
- * Prices in 1999 Dominican pesos

Contingencies

When estimating resource needs for planning and budgeting purposes, an allowance should be made for physical contingencies. Physical contingencies represent the possible increase in base costs as a result of changes in quantity and/or design of the project. For example, operating costs may vary considerably depending upon the level of use in the project. Therefore, cost estimates should reflect anticipated increases in use and project scale over time. Similarly, construction costs will significantly vary according to geography, terrain, and labor availability factors. Acceptable ranges for physical contingencies will vary among and within projects. They are most appropriately expressed as a percentage of base costs. They should be calculated, presented separately, and then aggregated to provide an estimate of the total cost of project components.

Cost Recovery

The technical report should also include an analysis of opportunities for cost recovery. This may include an analysis of the possible cost recovery that could be feasible from clients of the service. For example, a condom distribution program should evaluate opportunities for charging for condoms, and include an assessment of what is affordable for consumers. For an STI clinic, the technical report should include an evaluation of what the patients are willing and able to pay for STI services.

This section of the report should also include an evaluation of additional options for sustaining the program. For example, the report should

include an overview of resources that could be reasonably expected from the government, donors, and the private sector over the following five years.

CONCLUSION

A cost analysis is a useful and necessary tool for evaluating interventions. It is particularly useful for budgeting, sustainability planning, and economic evaluations. Regardless of whether or not the project is being implemented by a governmental or a non-governmental organization, it remains vital to track costs and to project the value of resources that will need to be purchased in the future. Cost analysis is also a necessary input when performing a cost-effectiveness, cost-utility, or cost-benefit analysis.

The most critical step in performing a cost analysis involves clearly defining the questions to be addressed. This will assist both in terms of identifying the costs that should be included and determining the ways in which

costs should be categorized. As illustrated in these guidelines, there are a variety of ways in which costs can be categorized.

Throughout the process of identifying costs, it is important to include all resources that are used for carrying out an intervention. This should include resources that are only indirectly used, as well as the use of resources that are provided in-kind, such as volunteer time or donated condoms. The technical report outlining the findings of the cost analysis should clearly address the original economic questions that were being asked and provide feasible solutions.

REFERENCES

- 1 Creese A, Parker D, editors. Cost analysis in primary health care: a training manual for programme managers. Geneva: World Health Organization; 1994.
- 2 Janowitz B, Bratt JH. Methods for costing family planning services. New York: United Nations Population Fund and Family Health International; 1994.
- 3 Kumaranayake L, Pepperall J, Goodman H, Mills A. Costing guidelines for HIV/AIDS prevention strategies. London: London School of Hygiene and Tropical Medicine; 1998.

17

GUIDELINES FOR PERFORMING COST-EFFECTIVENESS ANALYSIS OF HIV/AIDS PREVENTION AND CARE PROGRAMS

MICHAEL SWEAT

Definitions and Types of Analyses	254
Cost Analysis	254
Effectiveness Analysis	255
Effectiveness Modeling	256
Cost-effectiveness Analysis	257
Assessing the Analysis Strategy	259
Guidelines for Conducting a Cost-effectiveness Analysis	260
Framing the Problem	260
Identifying the Options to be Compared	261
Identifying the Outcome Measures	261
Identifying Intervention and Outcome Costs	261
Constructing a Decision Tree	262
Conducting a Sensitivity Analysis	263
Presenting the Results	263
Conclusion	264
References	264

17

GUIDELINES FOR PERFORMING COST-EFFECTIVENESS ANALYSIS OF HIV/AIDS PREVENTION AND CARE PROGRAMS

This chapter provides guidelines for conducting various types of effectiveness analysis. Here we build upon the analysis of costs presented in the Chapter 16, “Guidelines for Assessing the Economic and Financial Costs of HIV/AIDS Prevention and Care Programs,”

by looking at how cost data can be combined with measures of effectiveness to generate powerful estimates of the value of an intervention. It is important to remember, however, that the cost-effectiveness of an intervention is only one aspect of a successful intervention program. Some projects have societal goals that are not easily captured through cost-effectiveness analysis alone, and many projects have long-term benefits that are not readily assessed through cost-effectiveness analysis (such as programs with youth).

The first section of this chapter provides a general overview of the types of available effectiveness analysis, and walks the reader through a process of identifying what data are available, and what data are needed for different types of analyses. The types of analysis reviewed include:

- cost analysis;
- effectiveness analysis;
- effectiveness modeling; and
- cost-effectiveness analysis.

The second section of this chapter provides the reader with general guidelines to achieve the following basic steps in developing a cost-effectiveness analysis¹:

- framing the problem;
- identifying options to be compared;
- identifying the outcome measure(s);
- identifying intervention and outcome costs (reviewed in the previous chapter);
- constructing a decision tree;
- conducting a sensitivity analysis; and
- presenting the results.

DEFINITIONS AND TYPES OF ANALYSES

There are various approaches to examining the effectiveness of an HIV intervention, and the choice of analytic method depends on both the goals of the analysis and the data that are available. This section reviews various types of possible analyses. It is designed to provide the reader with an overview of the types of analyses that can be conducted, the strengths and limitations of the approaches, and the data needs for each type of analysis. We begin with cost analysis, as this is a required component of cost-effectiveness analysis, and because it has its own unique utility.

Cost Analysis

A cost analysis entails compiling and standardizing the costs of an intervention in order to assess the overall cost, annualized cost, and cost per client. Analyses are often broken out by distinct components of the intervention, by target groups, and by geographic areas of implementation. Costs are also frequently broken out into distinct components of the overall costs, such as recurrent (rent or labor), commodities (condoms or test-kits), and start-up expenses (office furniture or computers).

Cost analyses can take different perspectives, such as the cost to the donor, the non-governmental organization (NGO), or the client. Client costs often include assessments of the willingness to pay for the service, or natural experiments such as the impact of user fees on service use. Another issue to consider in this sort of analysis is the efficiency of the intervention. For example, does the size of the staff reasonably match the flow of clients through the service? It may be necessary to have a large staff to accommodate busy days, and thus the intervention program can be quite inefficient if staff are idle most of time and very busy at other times. Likewise, the staff may always be busy, but the quality of the service may be poor due to long

wait periods resulting in high opportunity costs to the clients (or lost time that could have been spent productively on other activities).

Data required for a cost analysis are described in the previous chapter but, at a minimum, include the line-item costs for the intervention, independent of research costs. It is also helpful to have data on the relative costs of providing an intervention to specific sub-populations, as well as information on seasonal trends in expenditures. Data can come from project budgets, from worksheets completed by project managers, and from interviews with project and donor staff.

It is also important to standardize cost data so that they can be compared across geographic settings and over time. For example, it is usually necessary to convert costs to a common currency metric (such as the dollar) and to discount the expenditures to a standard value of the currency in time to account for inflation and deflation, and variations in currency conversion rates. Moreover, it is also important to try and account for the purchasing power of the currency in the local setting. This is achieved by calculating a cost of living index based upon how much it costs at a given time to purchase a predetermined set of commodities and services in the local setting. The “market-basket” selected for this exercise needs to be a set of common items that are frequently used in all intervention settings. The Purchasing Power Parity (PPP) is a standard scale of the value of currencies, and is available at the World Bank website [www.worldbank.org/html/prddr/trans/m&a96/art7.htm]. The PPP is an estimate of the value of a “market-basket” of good and services based on the relative value of currency across countries over time. Although cost analysis is often conducted many years after an intervention is completed, data collection must be initiated at the time of the intervention.

One problem in compiling the cost data is that once projects disband, it can be difficult to

find the original project managers to identify actual intervention costs. In addition, budgets are notorious for not reflecting actual costs. If a client perspective is taken, data are needed on the amount of time, wages, transport and childcare costs, and opportunities lost. Although these costs can be assessed through interviews with clients, the literature does show that people tend to overestimate them in most surveys. Finally, start-up costs should be annuitized over the life of the project; for commodities, an annuity function is normally applied to the items to determine the annual cost. Simple formulas exist to do this. A good source of information on these procedures and associated formulas is Haddix and colleagues¹.

Effectiveness Analysis

Effectiveness analysis entails examining the impact of an intervention on the behavioral, biologic, social, or policy outcomes that the intervention was designed to affect. These analyses are often considered “outcome evaluations.” However, they include both effectiveness analysis (impact on these measures in the context of true to life service provision, such as an actual intervention in the field), and efficacy analysis (impact on these measures in an intervention being implemented under ideal or “gold standard” conditions). Efficacy analysis typically includes a direct comparison of intervention approaches made, such as would be seen in a multi-arm research study. Health economists frequently discuss the need to understand the “counterfactual” argument for this sort of analysis. “Counterfactual” is a term used by health economists to describe what the finding is being compared to.

Issues that frequently arise in effectiveness and efficacy analysis are the ability to attribute cause and effect, and the ability to compare the intervention effect to other interventions. As discussed in Chapter 1, “Conceptual Approach and

Framework for Monitoring and Evaluation,” it is often difficult in effectiveness studies to rigorously attribute the outcomes of interest to the intervention activities unless controls and randomization are included in the study design, which is rarely done in real-life evaluations. If trials occur without controls and randomization of study subjects, it is possible that changes in outcomes identified are due to “secular changes,” which are basically unmeasured societal level changes that occur in a group of study subjects. For example, AIDS mortality has an impact on the behaviors of people in a society because they witness the impact of HIV risk, and thus, behavior tends to change over time naturally in a society even without significant interventions. Without a control or comparison group in an evaluation, these changes appear to be intervention effects. Some interventions are especially difficult to

Issues that frequently arise in effectiveness and efficacy analysis are the ability to attribute cause and effect, and the ability to compare the intervention effect to other interventions.

evaluate for effectiveness due to the diffuse nature of the intervention, such as with mass media. The target populations are exposed to so many influences and are so dispersed that it is difficult to attribute causation to the intervention. Social and policy interventions can also be difficult to evaluate for similar reasons, as it is difficult to know which of many factors affected a change.

Data needed to conduct an effectiveness analysis include risk behavior, demographic information, intensity, and time of service use. Data can be collected from surveys of individuals who received the intervention and from biologic markers such as sexually transmitted infection (STI) rates. However, it is important that data for such evaluations be collected before and after the intervention. Process data, such as the number of condoms sold, can also be used to infer outcomes of interest, such as the usage rate of condoms. Problems that often arise include having only cross-sectional data on the study population at baseline and follow up, a lack of baseline data, lack of uniformity of measures assessed across intervention types, biased sampling, and collection of only process data. Additionally, the many factors that contribute to quality research (for example, validity of measures, self-report bias, or training of interviewers) are often lacking in evaluation studies. It is also often difficult to assess the rigor of the evaluation data retrospectively because the people who implemented the studies are often not accessible and records on study methodology are not available.

Effectiveness Modeling

Effectiveness modeling involves using mathematical models to estimate the likely level of disease outcomes that will result from the presence of different behavioral and environmental factors. Modeling is valuable for policy purposes because it harmonizes the outcomes of different interventions (an STI intervention may be concerned with the outcome of STI cases treated, while a media intervention may focus on outcomes relating to number of partners), and presents the intervention impact in terms of the disease outcome. Such models require data from behavioral surveys as well as assumptions gleaned from literature on the epidemiology of the disease of interest and the environment. For estimating HIV infections averted, there are some good basic models that

can make estimations of the number of people who will become infected with HIV over a set time frame. The AVERT model, available from Family Health International, is one example (Chapter 15, “Translating Survey Data into Program Impact: the AVERT Model,” provides more detail on this model)². Data needed for the models normally include the following:

- HIV prevalence of the target population;
- HIV prevalence of the sexual partners of the target population (usually from population-based surveillance studies);
- average number of sexual partners for the target population (from survey data);
- average number of sexual acts per contact (from survey data);
- condom use rate (from survey data);
- effectiveness of condoms, (from survey data or assumptions); and
- the infectivity of HIV, or the probability of a transmission of HIV occurring in a single sexual contact between an infected and uninfected person; this is affected by STI rates, so data on these are also helpful (from literature and STI surveillance).

The time frame for each parameter needs to be predetermined, but is normally set to one year. By estimating the HIV rates from baseline data and again at follow up, it is possible to take the difference between the two values to then calculate the intervention impact on HIV incidence (for example, the number of HIV infections averted). In controlled studies, the impact of the intervention can be controlled for differences across study arms. More detailed analysis can be conducted with spreadsheets. This model uses probability formulas that convert behavioral measures into estimates of HIV-1 incidence and it has been shown to have a high degree of predictive validity in high HIV-1 prevalence settings³.

Difficulties frequently encountered in modeling exercises include a lack of needed data, the need to annualize behavioral parameters (for example, convert monthly rates of condom use to annual estimates), unknown STI rates, uncertainty in HIV and STI rates among sexual contacts, and lack of consensus on the infectivity of HIV. Studies have shown that when sexual behavior data are collected carefully, there is a high degree of validity in their measurement^{4,5}. To help deal with the uncertainty in many of these parameters, a sensitivity analysis can be conducted using different combinations of the values and examining the impact on the outcome. By running the model with different likely values (extreme values, for example) it is possible to present a range of outcomes that are most likely, and therefore convey the level of uncertainty in the model. One good technique is to use 95 percent confidence intervals from survey data for extreme values in sensitivity analysis. Sensitivity analysis can become very complex, because multi-variate interactions are not easy to identify, and they often occur between parameters. Special decision analysis software packages, such as At-Risk, are available for effectiveness modeling, which can be used with spreadsheet analyses⁶.

Cost-effectiveness Analysis

Cost-effectiveness analysis involves coupling the intervention cost per client with effectiveness measures (both described above)⁷. Various outcomes can be calculated, depending on the availability of data. For example, recent cost-effectiveness analyses have been conducted of the impact of nevirapine on HIV transmission from mother to child, and on the impact of enhanced STI services on reducing HIV transmission^{8,9}. Below is a summary of the most commonly used outcomes:

- **Cost per unit of behavior or change**— This includes outcomes such as the cost

per condom used or the cost per partner reduced. This approach is not very common in the literature.

- **Cost per HIV infection averted**—This is a frequently used approach.

Effectiveness modeling is valuable for policy purposes because it harmonizes the outcomes of different interventions, and presents the intervention impact in terms of the disease outcome.

- **Cost per QALY or DALY saved**— Society may place a greater value on averting HIV cases among persons with certain characteristics (especially with regard to the quality of life they will experience) and among persons who will have a longer life should infection be averted. Thus, techniques exist to weight the impact of the infection averted and take into account the quality of life and the number of years of saved life that will result from an intervention. The two most common approaches is to convert the HIV infection averted to Quality Adjusted Life Years (QALY) or Disability Adjusted Life Years (DALY)¹⁰. To convert the number of HIV infections averted into QALY or DALY requires that the age of infection of the target population be identified. Discrete stages of infection are

then identified, each with an associated time duration and weight for the quality of life. The weighted number of years of life saved from the intervention are calculated using the age specific distribution of the target population. There is a growing consensus in the field on how to do this for HIV. However, most applications have been based on U.S. data and information on the natural history of HIV

in developing countries is not well understood due, in particular, to the lack of long-term cohort studies that include quality of life measures.

- **Cost utility analysis**—This approach uses cost-effectiveness measures described above (mostly QALY and DALY) and takes into account the treatment costs of HIV at different stages. Simple formulas can be used to calculate the cost-utility ratio, which is a

common measure used in health economics¹⁰. The benefit of the cost-utility ratio is that it places the cost-effectiveness of the intervention in relation to the cost-saving from treatment that result. This allows for direct comparisons to other health interventions.

assessment tool can be used to determine the type of analysis that is possible with different types of data. In some cases, there are options for the outcome variable that is examined. The minimum level of data required is highlighted with asterisks. As described above, four primary types of analysis are possible: (1) cost analysis, (2) effectiveness analysis, (3) modeling of effectiveness, and (4) cost-effectiveness analysis. Cost-effectiveness analysis can be broken out

Assessing the Analysis Strategy

Table 17-1 summarizes the various procedures that are possible for effectiveness analysis. This

TABLE 17-1 Assessment tool for effectiveness analysis

	Type of Analysis	Required Data and Procedures		Type of Analysis	Required Data and Procedures	
Cost analysis	Cost analysis	<ol style="list-style-type: none"> 1 Budgets and amendments to budgets** 2 Ability to contact project managers 3 Cost data breakout by intervention components 4 Cost data breakout by sub-populations 5 Cost data breakout over time** 6 Data on the quality of services or workload of staff 7 Data on the cost to the client 	Cost-effectiveness Analysis	Unit of Behavior Change	<ol style="list-style-type: none"> 1 Budgets and amendments to budgets** 2 Ability to contact project managers 3 Cost data breakout by intervention components 4 Cost data breakout by sub-populations 5 Cost data breakout over time** 6 Data on the quality of services or workload of staff 7 Data on the cost to the client 8 HIV prevalence of target population** 9 HIV prevalence of sex partners** 10 STI prevalence of target and partner population** 11 Number of sexual partners** 12 Number of sex acts per partner** 13 Condom use** 14 Condom effectiveness (breakage rate can be estimated) 15 Above data broken out by target sub-populations 	
Effectiveness Analysis	Effectiveness Analysis	<ol style="list-style-type: none"> 1 Behavioral data** 2 Demographic data 3 HIV prevalence or incidence 4 STI prevalence or incidence 5 Pre-intervention data** 6 Post intervention data** 7 Cross-sectional data** 8 Cohort data 9 Intensity of service use by average clients 	Cost Per:		HIV Infection Averted	<ol style="list-style-type: none"> 1 Both cost and effectiveness data** 2 Pre- and post-intervention survey data**
Effectiveness Modeling	Effectiveness Modeling	<ol style="list-style-type: none"> 1 HIV prevalence of target population** 2 HIV prevalence of sex partner** 3 STI prevalence of target and partner population** 4 Number of sexual partners** 5 Number of sex acts per partner** 6 Condom use** 7 Condom effectiveness (breakage rate can be estimated) 8 Above data broken out by target sub-populations 			QALY/DALY	<ol style="list-style-type: none"> 1 Average age of the target population** 2 Natural HIV/AIDS history studies conducted in the country
					Cost-Utility	<ol style="list-style-type: none"> 1 Lifetime treatment costs for HIV and AIDS**

** Minimum Data Required

into four additional areas, including: (1) cost per unit of behavior change, (2) cost per HIV infection averted, (3) cost per QALY or DALY saved, and (4) cost-utility analysis, which incorporates the treatment costs averted from the intervention. The next section looks more closely at cost-effectiveness analysis and reviews the steps involved in estimating the cost-effectiveness of an HIV intervention.

GUIDELINES FOR CONDUCTING A COST-EFFECTIVENESS ANALYSIS

In this section, the steps normally taken to conduct a cost-effectiveness analysis are reviewed. These follow the recommendations of Haddix and colleagues¹, with special attention given to HIV interventions.

- Is it more cost-effective to target older or younger adolescents in an HIV peer education intervention?
- Are peer education programs more cost-effective than HIV voluntary counseling and testing among factory workers in Thailand?

The perspective of the analysis relates to the question, “Who is responsible for the costs and consequences of the program being evaluated?” For example, analysts may be interested in examining the perspective of the donor (such as USAID or the national government) who pays for the intervention. Other perspectives that might be examined include the entire society, the government, the implementing agency, and

It is important that the comparisons for the analysis be carefully defined.

In many cases, an HIV intervention will be compared to no intervention.

Alternatives include comparisons between two different interventions, or comparisons between two or more target populations.

Framing the Problem

It is crucially important to frame the problem carefully before initiating a cost-effectiveness analysis. This involves specifying the study question, which in turn helps to define the other key elements of the analysis: perspective, time frame, and analytic horizon. Example of study questions include:

- What is the most cost-effective behavioral intervention for in-school sexually active adolescents in Western Kenya?

the individual who receives the intervention. The choice of perspective helps to determine the types of costs that are captured. For example, a donor may incur the costs of the intervention, but not the costs of treatment for HIV infections in a given setting.

The time frame for the analysis should then be determined. In evaluating HIV interventions, a 1-year time frame is typically used. Thus, costs are generated that reflect the annual cost per client, and the annual number of HIV infections

averted. It may be useful to also expand the time frame to include more long-term impacts, although it is often difficult to access data on long-term impacts of HIV interventions. In situations in which the time frame is a number of years, it is necessary to discount the costs to the present year value. There are techniques and formulas for discounting, and one simple approach is to use the following formula: $(1 + r)^{-t}$, where “r” is the discount rate (usually set to 3% to 5%), and “t” is the number of years from the current year.

The analytic horizon should next be determined. This is the amount of time over which the outcome is examined. For example, an HIV intervention may operate for only 2 years, but the benefits of the intervention may be realized over a lifetime for the recipients of the intervention. Thus, the analytic timeframe should probably be 3 years in this instance. For HIV interventions, care should be taken to assure that adequate data are available regarding outcomes over time. In most cases, it is probably better to set the time horizon at a shorter interval to generate a more conservative analysis. If long-term estimates are made, the results may have poor validity, and will be difficult to defend and use for policy purposes.

Identifying the Options to be Compared

Once the study question has been determined, it is important that the comparisons for the analysis be carefully defined. In many cases, an HIV intervention will be compared to no intervention. Alternatives include comparisons between two different interventions, or comparisons between two or more target populations. Comparisons can also be made between the health benefits of an HIV intervention versus a non-HIV health intervention. Various references are available to assist in making such comparisons¹¹. It is important that data for the analysis defined at this step be available. The more complex the comparisons defined,

the more difficult it will be to collect the requisite data for the analysis. Additionally, for purposes of analysis, it is typically necessary to have, at a minimum, pre-intervention and post-intervention data on the effectiveness of the intervention. That means that if two intervention approaches are being compared, four data sets are needed—data for before and after the intervention for each intervention approach.

Identifying the Outcome Measures

This is perhaps one of the most difficult steps in a cost-effectiveness analysis for HIV interventions. Some of the options for outcomes were reviewed earlier in this chapter. Table 17-1 is useful in determining which outcome measure is appropriate and what data are required to use each outcome. Again, some of the most frequently used outcome measures for HIV intervention cost-effectiveness analysis include the cost per HIV infection averted, the cost per quality adjusted life year saved, and the cost-utility ratio. The cost to the client may also be of interest in these analyses. Additionally, it may be useful to examine medical and social costs and benefits of the intervention, although these are often difficult to measure.

Identifying Intervention and Outcome Costs

Next, the costs of the intervention and outcomes need to be estimated. These are reviewed in Chapter 16, “Guidelines for Assessing the Economic and Financial Costs of HIV/AIDS Prevention Programs.” Care should be taken to harmonize the costs estimates with the comparisons made, and the analytic horizon of the analysis. With HIV intervention studies, the costs of the intervention are typically conveyed in terms of the cost per client to receive the intervention.

Constructing a Decision Tree

Decision trees are a graphic way to present the data of the cost-effectiveness analysis. They can be quite complex or relatively simple. Figure 17-1 shows a simple decision tree for a hypothetical analysis. In this case, two HIV interventions are being analyzed, Peer Education and HIV Voluntary Counseling and Testing (VCT). The peer education intervention costs \$22,000 to reach 1,000 clients, and the VCT intervention costs \$29,000. Using actual HIV incidence studies, or the AVERT model, it is

found that peer education averted 22 infections (45 minus 23), while the VCT intervention averted 35 HIV infections (45 minus 10) over a 1-year period. Thus, this simple decision tree shows that the cost per HIV infection averted for Peer Education is \$1,000 (\$22,000 ÷ 22 infections averted). For VCT, the cost per HIV infection averted is \$829 (\$29,000 ÷ 35). So while the peer education intervention was less expensive, it was also less effective, and in sum, it turned out to be less cost-effective. However, it is important to realize that

interpreting such results also needs to take into account social and policy considerations. It may be that in this setting there are serious social outcomes of learning that one is infected with HIV and thus, there is a desire to avoid HIV testing. Sophisticated decision trees allow for the inclusion of attitudes toward the outcomes, and inclusion of social and policy considerations.

The process for conducting sensitivity analysis is to run the analysis multiple times with varying parameter values. This simple form of sensitivity analysis (known as a one-way sensitivity) examines how changes in individual variables affect the outcome of interest. More sophisticated sensitivity analyses can be conducted that examine how sets of input parameters act together to affect the outcome. These typically are conducted with specialized software, such as At-Risk⁶.

Conducting a Sensitivity Analysis

The final step in the cost-effectiveness analysis is to conduct a sensitivity analysis of the decision tree and its associated analysis models in situations where outcomes are modeled. Sensitivity analysis takes into consideration any uncertainty that occurs in the data used. In all scientific studies there is some level of uncertainty in data that are collected and used for analysis (the speed of an atom, the size of a tumor, the chance of a volcano eruption). One way to capture and analyze the effect of such uncertainty in data is to first model the system mathematically, and to then make systematic changes in the parameters used in the mathematical analysis to see how they affect the outcome. By varying uncertain values over a reasonable range, it is possible to examine changes in variables in the system and see how stable the system is when values are changed. Selecting a reasonable range of values to represent the uncertainty is an important qualitative process that must be conducted carefully.

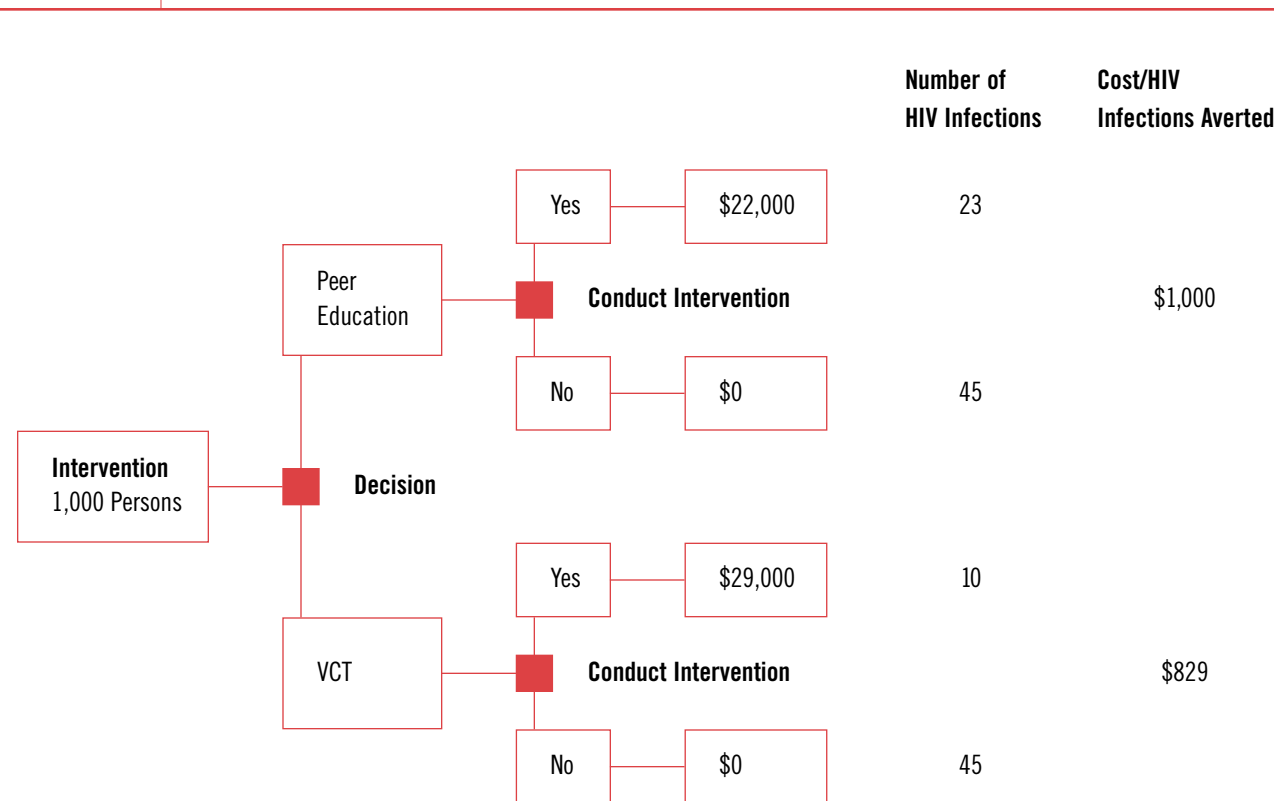
An example of an important sensitivity analysis that needs to be conducted with each intervention is the likely range of effectiveness that will be generated from the intervention. One good possibility is to use the confidence intervals of the outcome of interest, such as HIV incidence, to set the high and low values for the sensitivity analysis. In situations where one models the outcome, confidence intervals can be used for the input parameters, such as with percent condom use with the AVERT model.

Presenting the Results

Once all of the requisite analysis is completed, it is important to develop a policy presentation of the results. In doing this, it is important to consider the audience for the information. Overly technical presentations to persons not familiar with cost analysis can result in a poor response. It is also helpful to give concrete examples and to present the following details, as recommended by Haddix and colleagues¹:

- the study question;
- the study perspective, time frame, and analytic perspective;
- the assumptions used to build the model and estimate outcomes;
- a description of the interventions;
- evidence of the effectiveness of the interventions;
- identification of the relevant costs, including whether productivity costs are included, and the discount rate used;
- results of the analysis showing the comparisons made;
- results of the sensitivity analysis;
- discussion of the results that incorporate the social and policy perspective; and
- recommendations for action.

FIGURE 17-1 Simple decision tree for two interventions



CONCLUSION

Cost-effectiveness analysis can provide important insights into the utility of HIV intervention programs. Care needs to be taken to select an analytic approach that matches the specific research question of interest and available data. When conducting a cost-effectiveness analysis of an HIV intervention, it is important to carefully

develop the analysis plan as has been described in this chapter. The approaches outlined here describe basic methods for conducting a cost-effectiveness analysis. More sophisticated techniques not covered in this chapter are also available. It is recommended that use of advanced techniques be done cautiously and with the technical assistance of persons with experience in this area.

REFERENCES

- 1 Haddix AC. Prevention effectiveness: a guide to decision analysis and economic evaluation. New York: Oxford University Press; 1996.
- 2 Rehle T, Saidel T, Hassig S, et al. AVERT: a user friendly model to estimate the impact of HIV/sexually transmitted disease prevention interventions on HIV transmission. *AIDS* 1998;12(Suppl 2):S27-S35.
- 3 Weinstein MC, Graham JD, Siegel JE, Fineberg HV. Cost-effectiveness analysis of AIDS prevention programs: concepts, complications, and illustrations. In: Turner CF, Miller HG, Moses LE, editors. *Confronting AIDS: sexual behavior and intravenous drug use*. Washington (DC): National Academy Press; 1989. p. 471-499.
- 4 Coates T, Furlonge C, Mwakagile D, et al. Validation of self-reported sexual risk behavior with STD incident rates: results from the voluntary HIV counseling and testing study. From: XIth International Conference on AIDS. Geneva; 1998. Abstract no.: 14107.
- 5 Kauth MR, St. Lawrence JS, Kelly JA. Reliability of retrospective assessments of sexual HIV risk behavior: a comparison of biweekly, three-month, and twelve-month self-reports. *AIDS Educ Prev* 1991;3(3):207-214.
- 6 At-Risk [program]. 3.5.2 version. Newfield (NY); 1997.
- 7 Drummond M, Stoddard G, Torrence G. *Methods for the economic evaluation of health care programs*. New York: Oxford University Press; 1987.
- 8 Marseille E, Kahn JG, Mmiro F, et al. Cost effectiveness of single-dose nevirapine regimen for mothers and babies to decrease vertical HIV-1 transmission in sub-Saharan Africa. *Lancet* 1999;354(9181):803-809.
- 9 Gilson L, Mkanje R, Grosskurth H, et al. Cost-effectiveness of improved treatment services for sexually transmitted diseases in preventing HIV-1 infection in Mwanza Region, Tanzania. *Lancet* 1997;350(9094):1805-1809.
- 10 Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996.
- 11 Murray CLJ, Allan DL, editors. *The global burden of disease*. Cambridge (MA): Harvard University Press; 1996.



FAMILY HEALTH INTERNATIONAL

www.fhi.org

WASHINGTON OFFICE

HIV/AIDS Prevention and Care Department
2101 Wilson Boulevard, Suite 700
Arlington, VA 22201
USA

703 516-9779 Telephone

703 516-9781 Fax

HEADQUARTERS

PO Box 13950
Research Triangle Park, NC 27709
USA

919 544-7040 Telephone

919 544-7261 Fax