

COLEÇÃO ABIA

FUNDAMENTOS DE AVALIAÇÃO, Nº 2

**AVALIAÇÃO EM HIV/AIDS:
UMA PERSPECTIVA INTERNACIONAL**

GLENN SESSIONS

ASSOCIAÇÃO BRASILEIRA INTERDISCIPLINAR DE AIDS

RIO DE JANEIRO

2001

Associação Brasileira Interdisciplinar de AIDS - ABIA

Rua da Candelária, 79/10º andar - Centro

Cep: 20091-020 - Rio de Janeiro - RJ

Telefone: (21) 2223-1040

Fax: (21) 2253-8495

E-mail: abia@ax.apc.org

<http://www.alternex.com.br/~abia>

DIRETORIA

Diretor-Presidente: RICHARD PARKER

Diretora Vice-Presidente: SONIA CORRÊA

Secretário Geral: OTÁVIO CRUZ NETO

Tesoureiro: JOSÉ LOUREIRO

Coordenador Geral: VERIANO TERTO JR.

REVISÃO TÉCNICA

Ivia Maksud

REVISÃO DA PUBLICAÇÃO

Marta Torres

**PROJETO GRÁFICO, EDITORAÇÃO ELETRÔNICA
E PRODUÇÃO GRÁFICA**

Wilma Ferraz

Juan Carlos Raxach

APOIO

Fundação John D. e Catherine T. MacArthur

Fundação Ford

IMPRESSÃO

Gráfica Lidador

2ª TIRAGEM

1.500 exemplares

*É permitida a reprodução total ou parcial do artigo desta publicação,
desde que citados a fonte e o autor.*

SUMÁRIO

APRESENTAÇÃO	5
UMA NOTA SOBRE O OBJETIVO DESTE TEXTO	7
O QUE É AVALIAÇÃO?	8
PORQUE A AVALIAÇÃO É IMPORTANTE	9
UMA BREVE HISTÓRIA DA PRÁTICA DE AVALIAÇÃO DE PROGRAMA	10
TIPOS DE AVALIAÇÃO	11
AVALIAÇÃO FORMATIVA	11
AVALIAÇÃO DE PROCESSO	12
AVALIAÇÃO DE RESULTADO	13
AVALIAÇÕES DE EFICIÊNCIA	13
AVALIAÇÃO DE IMPACTO	14
UMA NOTA SOBRE A TERMINOLOGIA DE AVALIAÇÃO	15
TÉCNICAS PARA COLETAR E ANALISAR DADOS PARA AVALIAÇÃO	15
FONTES DE DADOS QUANTITATIVOS	16
ENTREVISTAS ESTRUTURADAS (LEVANTAMENTOS)	16
QUESTIONÁRIOS AUTO-APLICADOS	17
DADOS SECUNDÁRIOS	17
FONTES DE DADOS QUALITATIVOS	18
ENTREVISTAS EM PROFUNDIDADE	18
GRUPOS DE DISCUSSÃO	19
OBSERVAÇÃO DIRETA	19
CLIENTE SIMULADO	20
OBSERVAÇÃO ETNOGRÁFICA	21
TRIANGULAÇÃO	21

TIPOS DE DESENHOS DE PESQUISA	22
DESENHOS DE PESQUISA PARA AVALIAÇÕES DE PROCESSO	22
DESENHOS DE PESQUISA PARA AVALIAÇÕES DE RESULTADO	23
<i>Desenho experimental aleatório</i>	23
<i>Desenho quase-experimental</i>	25
<i>Desenho não-experimental</i>	27
UMA NOTA SOBRE A ESCOLHA DE DESENHOS DE PESQUISA	28
DESENHOS DE PESQUISA PARA AVALIAÇÃO FORMATIVA	29
AVALIAÇÃO PARTICIPATIVA	30
PONTOS NA AVALIAÇÃO DE PROGRAMA	31
TIPOS DE AVALIADOR: INTERNO E EXTERNO	31
FINALIDADE GERAL DA AVALIAÇÃO DE PROGRAMA	32
ADAPTABILIDADE CULTURAL DOS MÉTODOS DE AVALIAÇÃO DE PROGRAMA	32
CONCLUSÃO	33
BIBLIOGRAFIA	33

APRESENTAÇÃO

O texto *Avaliação em HIV/AIDS: uma perspectiva internacional* foi desenvolvido especialmente para ser apresentado no seminário “Avaliação em DST/AIDS”, organizado pela Associação Brasileira Interdisciplinar de AIDS com o apoio da Fundação Ford e do Ministério da Saúde. Dando continuidade ao debate instaurado pelo texto *Avaliação em Saúde: Questões para os programas de DST/AIDS no Brasil*, de Maria Ines Baptistella Nemes (Coleção ABIA, Série Avaliação, nº 1), que buscou contextualizar a prática de avaliação em saúde de um modo mais amplo, no país, o texto de Gleen Sessions se debruça especificamente sobre questões concernentes ao trabalho com AIDS. O autor, fazendo uma revisão da literatura internacional existente sobre avaliação e prevenção, disponibiliza para gerentes de programas, organizações comunitárias e todos aqueles que trabalham com HIV/AIDS técnicas avaliativas utilizadas em programas mundiais de prevenção, ponderando a necessidade de retradução e adaptação das mesmas antes de serem utilizadas na realidade brasileira. Esperamos que esse texto possa contribuir para o desenvolvimento de avaliações mais eficazes em direção às ações relacionadas ao combate da epidemia de HIV/AIDS no contexto brasileiro.

Richard Parker
Diretor-Presidente da ABIA

Veriano Terto Jr.
Coordenador Geral da ABIA



AVALIAÇÃO EM HIV/AIDS: UMA PERSPECTIVA INTERNACIONAL

GLENN SESSIONS

Quando a epidemia de HIV/AIDS foi reconhecida pela primeira vez, há quase vinte anos, a maioria dos programas comunitários de educação e prevenção foram implementados rapidamente para reduzir o comportamento de alto risco e evitar a transmissão do HIV. Devido à premência de se fazer qualquer coisa para reduzir a rapidez de disseminação da epidemia, os recursos disponíveis foram usados primariamente para desenvolver intervenções específicas e material educacional visando as populações de risco. Compreensivelmente, a avaliação dos programas não constituía uma prioridade para a maioria dos esforços de prevenção do HIV durante os primeiros anos da epidemia, quando a necessidade mais imediata era agir rapidamente para salvar vidas. Os programas de prevenção não dispunham de tempo ou recursos para realizar atividades sistemáticas de planejamento e pesquisa que, de maneira ideal, deveriam preceder o desenho e o desenvolvimento de um programa. Hoje em dia, apesar de um aumento nos esforços de avaliação dos programas nos últimos anos, especial-

mente na América do Norte e na Europa, permanece ainda uma necessidade crítica de implementar atividades de avaliação contínuas ao nível da comunidade. Avaliar a efetividade dos programas de prevenção do HIV/AIDS e de intervenções específicas é a única maneira ao alcance dos gerentes de programa e dos responsáveis pela definição de políticas que permite tomar decisões acertadas a respeito de seus programas.

UMA NOTA SOBRE O OBJETIVO DESTE TEXTO

O objetivo principal desta revisão é encorajar os responsáveis por tomar decisões, os gerentes de programa e as equipes de programas comunitários de prevenção e educação em HIV/AIDS a refletir sobre a avaliação e sua relevância para suas organizações. Espera-se que este texto sirva de recurso para todos que trabalham com HIV/AIDS e desejam aprender mais sobre os vários tipos de avaliação que possam ser aplicáveis aos seus programas. Com isso em vista, esta revisão pretende “desmistificar” parte da

terminologia de avaliação e discutir os diferentes tipos de desenho de pesquisa e técnicas comumente usadas da prática da avaliação em todo o mundo. Para elucidar esses conceitos e abordagens foram usados muitos exemplos internacionais de avaliação de programas e atividades de pesquisa. A maioria dessas citações foi extraída da literatura corrente sobre prevenção do HIV e avaliação e refere-se principalmente a atividades de avaliação dispendiosas e tecnicamente exigentes, que em geral estão fora do alcance da maioria das organizações comunitárias. Os exemplos usados foram selecionados com o objetivo de ilustrar as diferentes técnicas e tipos de avaliação que são discutidas no texto desta revisão, e não servem como um modelo real de como uma organização comunitária “deveria” realizar a avaliação de seu trabalho. As atividades de avaliação precisam ser realistas e adaptadas às necessidades e aos recursos da organização, se esperamos que ofereçam informações significativas e úteis que possam no final ajudar na realização de um programa mais efetivo.

O QUE É AVALIAÇÃO?

Não existe uma definição comumente aceita de avaliação; freqüentemente ela é definida de acordo com os diferentes contextos e situações onde é usada. Conseqüentemente, muitas definições foram desenvolvidas por diferentes disciplinas e tipos de programa. Entretanto, para fins desta revisão, *avaliação* será definida simplesmente como um “processo sistemático para determinar até que ponto um programa ou intervenção atingiu os objetivos pretendidos” (1). Aplicável à maioria das atividades organizadas de saúde pública, esta definição sugere vários componentes-chave de uma avaliação de programa bem-sucedida: objetivos e metas cuidadosamente definidos no programa ou intervenção a ser avaliado, um sistema estruturado e consistente para

coletar, analisar e relatar as informações, um método claramente definido para medir o grau de mudança resultante do programa ou intervenção e uma estimativa final para determinar se o objetivo final da avaliação foi alcançado. A avaliação visa em primeiro lugar dirigir a ação, determinar se as atividades do programa atendem aos seus objetivos declarados, se são apropriadas e se estão sendo efetivamente implementadas (2). Partindo desta perspectiva, a avaliação dos programas de prevenção do HIV permite que os responsáveis pelas decisões meçam sistematicamente a efetividade das estratégias e intervenções para reduzir a transmissão do HIV, baseados em descobertas empíricas em lugar de abordagens mais subjetivas.

Como foi rapidamente mencionado, um componente essencial de qualquer atividade de avaliação é identificar as metas e objetivos do programa ou intervenção antes que uma atividade de avaliação possa ser realizada. Uma *meta* pode ser definida como “uma declaração, comumente geral e abstrata, de um estado desejado para o qual um programa é dirigido” (3). A meta de um programa é uma declaração geral sobre o que ele deseja alcançar a longo prazo, em geral sem um prazo ou critérios mensuráveis. A meta geral da maioria dos programas de prevenção é reduzir a taxa de transmissão do HIV em uma determinada comunidade ou alguma população-alvo (p. ex. reduzir a incidência de HIV/AIDS entre profissionais do sexo em Tegucigalpa, Honduras). As metas de um programa também podem ser mais focalizadas, relacionadas a intervenções específicas (p. ex. aumentar o uso do preservativo entre profissionais do sexo em Tegucigalpa, Honduras). Os *objetivos*, ao contrário, são declarações específicas, operacionalizadas, detalhando as realizações desejadas de um programa (3). São mais imediatos que as metas, entretanto contribuem para que a meta a longo prazo seja atingida, e são descritos em termos que podem ser

medidos com um período de tempo específico durante o qual os resultados serão atingidos. Os objetivos dos programas de prevenção do HIV geralmente são apresentados em termos dos resultados esperados das atividades ao nível do programa (p. ex. aumentar 20% em seis meses o número de profissionais do sexo em Tegucigalpa, Honduras, que informam ter usado preservativo em seu último encontro sexual).

Os *indicadores* são variáveis que indicam o progresso que foi feito em relação a uma meta ou objetivo. Muitas das maiores organizações de desenvolvimento internacional classificam os indicadores usados em avaliação de programas em diferentes tipos: *insumo* (recursos do programa), *processo* (atividades do programa), *rendimento* (resultados do programa) e *resultado* (resultados ao nível da população) (4). Muitas vezes denominados indicadores de desempenho, os indicadores fornecem aos gerentes e avaliadores critérios mensuráveis para determinar como seus programas estão em relação à obtenção dos resultados esperados. Os indicadores em geral fornecem dados quantitativos em forma de percentagem, valor médio ou número absoluto (p. ex. a percentagem de profissionais do sexo que informam ter usado preservativo em seu último encontro sexual), mas podem também ser expressos em um formato sim/não (p. ex. um estudo-piloto de profissionais do sexo terminado: sim/não).

Conjuntos de indicadores, que em geral estão diretamente ligados a atividades, objetivos e metas do programa, são uma parte essencial de um sistema bem-sucedido de monitoração e avaliação de programa. Embora não possam ser claramente separadas, a monitoração em geral se preocupa com a implementação contínua de um programa e a avaliação se ocupa da efetividade e do impacto do programa (5). *Monitoração* é definida como o rastreamento rotineiro de informações prioritárias sobre um programa e seus efeitos planejados (6). Um sistema de monitoração

de programa pode incluir vários tipos de fontes de informação, como estatísticas de gerenciamento e serviço, relatórios financeiros, distribuição de material, levantamento de clientes e dados qualitativos. Ao nível do programa, a monitoração efetiva permite que os gerentes estimem o progresso de seu programa e introduzam aperfeiçoamentos para atingir os objetivos e metas do programa (p. ex. após monitorar o rendimento em meados do programa, intervenções objetivando profissionais do sexo e uso de preservativo são intensificadas). A monitoração do programa é frequentemente denominada *avaliação de processo*, e será examinada em detalhe mais adiante.

PORQUE A AVALIAÇÃO É IMPORTANTE

Os responsáveis pelas políticas, os gerentes de programa, os profissionais da área de saúde e os ativistas comunitários vêm reconhecendo cada vez mais que a monitoração e a avaliação dos programas de prevenção do HIV beneficiam tanto a comunidade como a organização. Há vários motivos pelos quais a monitoração e a avaliação são importantes. Primeiro, proporcionam uma abordagem sistemática para estimar e melhorar a efetividade de um programa de prevenção. Segundo, a monitoração e a avaliação podem identificar os processos pelos quais um determinado resultado é obtido, de modo que estratégias e intervenções bem-sucedidas possam ser reproduzidas em outras situações. Terceiro, ajudam os responsáveis pelas decisões a determinar se os recursos de um programa estão sendo alocados judiciosamente ou não. E quarto, atividades de avaliação cuidadosa podem demonstrar o sucesso organizacional à comunidade e aos responsáveis pelas políticas, além de órgãos financiadores atuais e potenciais, que frequentemente exigem provas dos resultados dos esforços de um programa como condição para continuar dando apoio.

Embora a utilidade de uma avaliação dos componentes dos programas de prevenção do HIV seja geralmente aceita, os gerentes de programa das organizações comunitárias muitas vezes relutam em implementar atividades de avaliação. Frequentemente isso se deve a percepções errôneas das práticas e propósito da monitoração e avaliação de um programa. Uma preocupação comum é a de que a avaliação possa desviar recursos financeiros e humanos valiosos dos serviços de prevenção direta. Entretanto, os tipos mais básicos de avaliação (p. ex. de processo) podem apresentar resultados úteis a partir de fontes de dados já existentes, com pouco esforço ou despesa adicionais. Para os tipos mais exigentes de avaliação (p. ex. formativa ou de resultado), outras estratégias podem ser usadas para reduzir custos, como colaborar com outras organizações para realizar atividades de avaliação cooperativa ou identificar outras fontes de financiamento (7). Além disso, ao determinar o que funciona e o que não funciona em um programa, os esforços de avaliação beneficiam tanto a população-alvo do programa como sua equipe, ao permitir que seja feita uma alocação mais eficiente de recursos limitados.

Outra preocupação da maioria das organizações comunitárias é a falta de experiência da equipe para projetar e realizar uma avaliação de programa. Embora muitos tipos de avaliação exijam pesquisadores treinados e experientes, outros tipos de avaliação são baseados em métodos práticos e contínuos de avaliação que absorvem a equipe do programa, membros da comunidade e outras pessoas envolvidas. Quando as exigências técnicas de uma avaliação proposta demandam assistência externa, entretanto, as organizações podem buscar outras fontes de conhecimento sobre avaliação, que podem ser outras organizações envolvidas com HIV/AIDS, secretarias municipais de saúde, fundações da área e universidades locais (7, 8).

Como muitas pessoas não compreendem bem o propósito e os benefícios da monitoração e da avaliação, a resistência apresentada tanto pela equipe do programa como pelos membros da comunidade constitui outra barreira bastante comum para a realização de um programa de avaliação bem-sucedido. Essas pessoas podem estar preocupadas com o tempo e os custos da avaliação, e se ela pretende beneficiar o órgão financiador mais do que o programa ou a comunidade. Os membros da equipe podem também estar receosos de que estejam sendo julgados e seus empregos estejam em risco. Para atender a tais preocupações e criar um ambiente de apoio à avaliação, os gerentes de programa e avaliadores precisam pensar em abordagens que envolvam e dêem poder à comunidade e à equipe do programa no processo de avaliação. Tais estratégias incluem: estabelecer grupos consultivos na comunidade para contribuir para o planejamento e a implementação da avaliação, realizar debates educativos sobre assuntos relacionados, selecionar uma metodologia que estimule a participação da equipe no processo de avaliação e divulgar os resultados preliminares e finais da avaliação entre a equipe e os membros da comunidade (9). Este comportamento de colaboração aumenta a probabilidade de que um trabalho de avaliação seja eficiente, além de intensificar sua credibilidade.

UMA BREVE HISTÓRIA DA PRÁTICA DE AVALIAÇÃO DE PROGRAMA

Embora a avaliação sistemática dos programas de educação e saúde pública já exista no mundo industrializado desde o começo do século XX, a moderna avaliação de programa teve início primariamente nos Estados Unidos durante a Grande Depressão dos anos 30 e no período seguinte à Segunda Guerra Mundial. Com o crescimento dramático do dispêndio governamental em programas sociais de grande

escala durante esse período, houve uma pressão cada vez maior por parte de políticos e autoridades do governo para a aplicação das técnicas de “gerenciamento científico”, que estavam sendo usadas pela indústria americana e os militares, às instituições de serviço humanitário para fins de planejamento e tomada de decisões. Conceitos e métodos extraídos da economia, ciência política, sociologia e outras ciências foram usados para desenvolver e refinar métodos de pesquisa social, como pesquisas por amostragem e procedimentos estatísticos, e os departamentos de ciência social das universidades tornaram-se os centros do trabalho inicial em avaliação de programa (3).

Ao final dos anos 50, programas de avaliação em grande escala estavam sendo realizados não só nos Estados Unidos, na Europa e em outros países industrializados, mas também em nações menos desenvolvidas. Em todo o mundo, programas de planejamento familiar, nutrição, assistência médica, desenvolvimento agrícola e comunitário incluíam componentes de avaliação. Além disso, quantidades crescentes de literatura sobre a prática de pesquisa de avaliação foram produzidas em todo o mundo e os métodos de pesquisa com coleta e análise de dados tornaram-se cada vez mais sofisticados, especialmente com a introdução da tecnologia informática (3).

Nos Estados Unidos, atividades de avaliação de programa tiveram seu crescimento mais rápido durante as eras Kennedy e Johnson, de meados ao fim dos anos 60, com os programas sociais de “guerra à pobreza” e as iniciativas da “Grande Sociedade” financiados pelo governo federal. Amplos recursos foram destinados para combater o desemprego, o crime, a deterioração urbana, proporcionar acesso à assistência médica e ao tratamento da saúde mental. A avaliação de programa era usada principalmente para estimar os ganhos desses programas. Entretanto, os problemas complexos que cons-

tituíam o alvo dessas iniciativas ambiciosas não desapareceram e muitos dos programas foram considerados mal planejados e administrados, limitados em sua efetividade final. Como resultado, durante os anos 70 e 80, o público americano, incentivado pelo discurso de políticos conservadores, tornou-se cada vez mais desconfiado do apoio governamental em grande escala dado a iniciativas sociais. A pressão política tornou-se cada vez maior para que os programas sociais apresentassem provas de sua efetividade para receber recursos. Esses fatos tiveram um efeito significativo na ênfase dada à prática atual de avaliação nos Estados Unidos e, provavelmente, no resto do mundo. Desde os anos 70, os esforços de avaliação voltaram-se primeiramente para a demonstração da efetividade dos programas sociais para garantir a financiadores, autoridades do governo e ao público em geral que seus investimentos valiam a pena (3, 10). Assim, embora a avaliação tenha se tornado uma parte amplamente aceita do planejamento e gerenciamento de programas sociais no mundo industrializado durante aquele período, a ênfase principal estava na responsabilidade fiscal e na comprovação da efetividade. Esta tendência na avaliação de programa continua até os dias atuais. Entretanto, houve um aumento marcante do reconhecimento dos benefícios da boa prática de avaliação entre gerentes de programa e responsáveis por políticas, não só para demonstrar a efetividade e os resultados, mas também para proporcionar outros tipos de informação que possam ajudar os programas sociais a melhorar a qualidade de seus serviços.

TIPOS DE AVALIAÇÃO

AVALIAÇÃO FORMATIVA

A maioria dos especialistas em avaliação faz distinção entre cinco tipos de avaliação:

formativa, de processo, de resultado, de eficiência e de impacto. Devido aos desafios metodológicos e às necessidades de recursos, os dois últimos tipos de avaliação em geral não são realizados pelas organizações comunitárias. Esta categorização é baseada em quando a avaliação está sendo realizada, que tipo de informação está sendo coletada e a finalidade geral da avaliação. A *avaliação formativa* é um tipo especial de avaliação antecipada que ocorre durante e após um programa ter sido desenvolvido, mas *antes* que seja amplamente implementado. É usada para entender a necessidade real de um programa ou de uma determinada intervenção, para tomar decisões experimentais sobre como implementar a intervenção ou melhorá-la (11). A avaliação formativa pode ajudar na seleção e desenho de atividades e serviços apropriados para o programa, antes da implementação, para que consigam atingir melhor suas metas e objetivos.

A informação necessária para uma avaliação formativa inclui tanto dados quantitativos como qualitativos que podem ser extraídos de uma variedade de fontes inclusive estatísticas de saúde pública ou dados censitários, questionários de pesquisa e grupos de discussão com membros da população-alvo, entrevistas com pessoas-chave da comunidade e revisão da literatura científica (11). Em um programa de prevenção de HIV/AIDS, por exemplo, a avaliação formativa é usada frequentemente para estimar a provável efetividade das campanhas de comunicação de massa antes de seu lançamento para o grande público. Outros exemplos: medições básicas de conhecimentos e atitudes de uma população-alvo em relação a **comportamentos de prevenção** do HIV/AIDS, determinação das necessidades de um sistema de distribuição de preservativos em uma determinada comunidade, teste prévio do desenho de questionários, entrevistas em profundidade com a equipe e os participantes de um programa para identificar antecipada-

mente os problemas de implementação e teste-piloto de um programa de intervenção, no todo ou só de algumas partes, antes de sua implementação rotineira. Deve-se observar que as avaliações formativas podem ser relativamente simples, ou complexas e dispendiosas, dependendo do desenho de pesquisa e dos métodos usados para determinar a intervenção ou programa a ser implementado.

AVALIAÇÃO DE PROCESSO

A *avaliação de processo* é o tipo mais básico de avaliação. As avaliações de processo estudam as maneiras pelas quais os serviços ou intervenções são realizados; são planejadas para descrever o que está acontecendo e não para estabelecer se funciona ou não (12). Envolve três perguntas básicas: O que está sendo feito? A quem está sendo feito? Como está sendo feito? O foco está na monitoração das atividades do programa e se estão sendo dirigidas às pessoas certas. Este tipo de avaliação geralmente começa com a implementação da intervenção e continua durante sua execução para determinar se a intervenção está funcionando como foi planejado e, se não, para direcionar as ações necessárias para melhorar o trabalho (13).

Na maior parte dos casos, a avaliação de processo requer poucos recursos adicionais do programa e pode utilizar registros administrativos existentes e outros sistemas de informação sobre as atividades em curso no programa, sendo parte do desenho de pesquisa. Por exemplo, os dados do programa usados em uma avaliação de processo de um programa de prevenção de HIV/AIDS podem incluir: o número de clientes que fizeram o teste de HIV e receberam aconselhamento, o número de pessoas treinadas para diagnosticar e tratar DSTs, o número de preservativos vendidos ou distribuídos, ou as características demográficas dos clientes do programa.

Além de estatísticas rotineiras de serviço e dados de indicadores, este tipo de avaliação também pode utilizar vários outros métodos para coletar dados, como discussões temáticas de grupo, entrevistas e levantamentos entre a equipe e os participantes do programa, observação direta de atividades do programa, relatórios feitos pelo pessoal do programa, e estudos narrativos de caso. Os resultados das avaliações de processo são em geral descritivos e incluem detalhes do que acontece de fato quando a intervenção ou serviços estão sendo realizados. Usando uma mistura dessas várias técnicas, uma avaliação de processo é extremamente útil para explicar como e porque um programa “funciona” ou não funciona; entretanto, não pode demonstrar se um programa está sendo realmente efetivo.

AValiação DE RESULTADO

O tipo mais comum de avaliação é a *avaliação de resultado*. A avaliação de resultado determina a efetividade de um programa, demonstrando até que ponto conseguiu produzir os resultados desejados e atingir os objetivos declarados (14). Responde às perguntas: “O programa ou intervenção funciona?” e “É capaz de fazer diferença?”. Em relação aos programas comunitários de prevenção do HIV, a avaliação de resultados geralmente focaliza os efeitos de um programa relativamente a curto prazo e se ele modificou o conhecimento, as crenças, intenções ou comportamentos de uma população-alvo específica. Muitos tipos diferentes de resultados podem ser estimados, inclusive cognitivos (p. ex. conhecimento sobre o HIV e seus modos de transmissão), de atitude (p. ex. percepções sobre pessoas com HIV/AIDS) e de comportamento (p. ex. mudanças no comportamento pessoal ou de grupo) (5).

Em comparação com as avaliações de processo, uma avaliação de resultado é mais difícil

de realizar porque mede os resultados de um programa de intervenção. Pode também exigir muito trabalho e ser cara, dependendo da complexidade do desenho de pesquisa e da intervenção a ser avaliada. Por exemplo, em sua forma mais simples, uma organização comunitária poderia fazer uma avaliação de resultado de uma atividade simples, como um seminário de prevenção do HIV. Poderia tentar medir quanto o grupo-alvo em questão aprendeu durante o seminário usando técnicas como testes prévios e posteriores e entrevistas com os participantes. Entretanto, outras avaliações de resultado são muito mais complexas do que isso, especialmente se a intervenção envolve um certo número de medições dos resultados do programa e uma população-alvo maior, ou tentativas de medir a mudança de comportamento. Ademais, além de mostrar até onde um programa está atingindo seus objetivos, as avaliações de resultado deveriam ser capazes de demonstrar quais as atividades do programa foram diretamente responsáveis pelos efeitos observados e descartar quaisquer outras explicações plausíveis para os resultados. Para poder atribuir as mudanças nos níveis de conhecimento e, particularmente, nos comportamentos de uma população-alvo somente a um programa ou intervenção específica é preciso um desenho experimental extremamente complexo e, sem assistência externa, este tipo de pesquisa para avaliação está além das capacidades da maioria das organizações comunitárias. Outros tipos menos rigorosos de avaliação de resultado, entretanto, podem ser realizados por organizações comunitárias a partir de uma grande quantidade de variáveis, proporcionando informações valiosas sobre a efetividade do programa.

AValiações DE EFICIÊNCIA

As *avaliações de eficiência* examinam a relação entre os resultados e os custos de um pro-

grama a fim de determinar se os recursos financeiros estão sendo gastos de forma eficiente. Proporcionam aos gerentes de programa e responsáveis pelas políticas informações para a tomada de decisões sobre alocação de recursos. Os dois tipos principais de avaliação de eficiência são a análise de custo-benefício e a análise da eficiência em termos de custo. Ambas oferecem uma estrutura de referência para relacionar os custos aos resultados do programa. A *análise de custo-benefício* exige que os resultados do programa, além dos custos, sejam expressos em termos monetários. Assim, os custos monetários são comparados com os resultados monetários. É usada para determinar se os benefícios econômicos de um programa social têm mais valor para a sociedade do que seus custos financeiros e permite que diferentes tipos de programa (p. ex. saúde, educação, agricultura) com resultados diferentes sejam comparados porque todos os resultados são medidos usando o mesmo padrão monetário. Em um programa de prevenção do HIV, a análise de custo-benefício pode focalizar a diferença entre os custos do programa e o dinheiro poupado com despesas de tratamento para as pessoas que de outra forma teriam se tornado infectadas sem a intervenção do programa. Em muitos casos, entretanto, a aplicação desta abordagem aos programas de prevenção do HIV é tecnicamente desafiadora e também controversa, devido à dificuldade de atribuir valores monetários a certos resultados de saúde e em particular à própria vida humana.

A *análise da eficiência em termos de custos* é usada com mais frequência na avaliação dos programas comunitários de prevenção de doenças do que a análise de custo-benefício. Neste tipo de avaliação apenas os custos do programa são calculados em termos monetários; os resultados são expressos em termos substantivos ou físicos, tais como o custo necessário para prevenir uma infecção pelo HIV ou treinar um agente de saúde.

Permite que os responsáveis pelas decisões determinem se o benefício (ou resultado) de um programa de prevenção ou intervenção específica de prevenção do HIV vale o que custa. Por exemplo, uma avaliação da eficiência em termos de custo de um programa de prevenção de DST/HIV entre mulheres profissionais do sexo em Nairóbi, Quênia, estimou que o programa evitou de 6 mil a 10 mil novos casos entre clientes e parceiros sexuais de clientes a um custo de US\$8 a US\$12 por infecção de HIV evitada (15). Pode também ser usada para comparar a eficiência de diferentes estratégias de prevenção para atingir o mesmo resultado. Nos Estados Unidos, um estudo de eficiência em termos de custo, realizado entre usuários de drogas injetáveis, determinou que o custo de evitar uma infecção pelo HIV era menor para uma troca de seringas e um programa de aconselhamento/educação prolongado (cerca de US\$4 mil) do que um programa de aconselhamento e testes de HIV (cerca de US\$13 mil) (16). Estes exemplos demonstram a utilidade da análise de eficiência em termos de custo nos programas de prevenção do HIV. Entretanto, este tipo de avaliação é complexo e exige pessoal com um alto nível de experiência técnica. Semelhante à avaliação de resultado, a análise de eficiência em termos de custo também mede a efetividade de um programa, mas estende a análise acrescentando a medição dos custos do programa. Devido às complexidades metodológicas, a maioria dos especialistas em avaliação acha que as organizações comunitárias devem adiar as avaliações de eficiência até que tenham adquirido mais experiência com avaliações formativas, de processo e de resultado (7, 10).

AVALIAÇÃO DE IMPACTO

A *avaliação de impacto* é uma avaliação que focaliza os efeitos finais generalizados de um programa sobre uma população-alvo mais ampla,

além dos participantes diretos imediatos do programa (9). Neste tipo de avaliação, um programa de prevenção poderia tentar determinar se suas atividades, ou uma intervenção específica, foram realmente capazes de reduzir a incidência do HIV ou aumentar o uso de preservativos em uma comunidade. Para serem conduzidos eficazmente, estudos de impacto requerem grandes quantidades de recursos financeiros e profissionais. Além disso, são enormes os desafios metodológicos, especialmente os relacionados à atribuição programática. É praticamente impossível atribuir qualquer mudança de comportamento a esforços de um único programa, já que as populações ou comunidades em questão também foram expostas à mídia e a outras fontes de informação sobre HIV/AIDS, amigos e parentes com HIV e AIDS, e outras atividades nacionais e locais de prevenção, além do programa que está sendo avaliado. Consequentemente, estudos de impacto de programas de prevenção de HIV a nível comunitário são raros, e, à exceção de umas poucas organizações, jamais se esperaria que organizações trabalhando a nível da comunidade conduzissem esse tipo de avaliação.

UMA NOTA SOBRE A TERMINOLOGIA DE AVALIAÇÃO

É importante mencionar que a terminologia na literatura sobre avaliação não é usada consistentemente e às vezes pode ser muito confusa. Alguns autores, por exemplo, podem empregar “resultado” e “impacto” indiferentemente, enquanto outros podem de fato usar estes termos de forma oposta àquela em foram apresentados nesta revisão (11). A expressão “avaliação somativa” pode significar coisas diferentes para diferentes pessoas e às vezes é usada no mesmo contexto que avaliação de “resultado” ou de “impacto”, ou ambas. E finalmente, a

expressão “eficiência em termos de custo” é freqüentemente usada na literatura para referir-se tanto à análise de “eficiência em termos de custo” como à análise de “custo-benefício”. Recentemente, entretanto, tem havido uma tentativa informal dos especialistas em avaliação para padronizar a terminologia. Esta revisão emprega os termos mais comumente usados encontrados na literatura sobre avaliação.

TÉCNICAS PARA COLETAR E ANALISAR DADOS PARA AVALIAÇÃO

A atividade de avaliação depende da disponibilidade de dados extraídos de uma variedade de fontes. De uma forma ideal, a avaliação de um programa deveria incluir metodologias tanto quantitativas quanto qualitativas para a coleta e análise dos dados. A escolha do tipo de abordagem depende da atividade do programa a ser avaliada. A maioria dos especialistas em avaliação concordaria que os diferentes métodos quantitativos e qualitativos deveriam ser considerados complementares, e não mutuamente exclusivos. Entretanto, tem-se discutido qual a melhor abordagem. Grande parte desta controvérsia teve origem em concepções errôneas sobre a finalidade de cada abordagem. Os métodos quantitativos, que dependem de dados que podem ser contados (isto é, números), pretendem produzir dados factuais confiáveis sobre efeitos e resultados generalizáveis a populações maiores; enquanto os métodos qualitativos, que coletam dados sob forma narrativa, pretendem fornecer dados altamente detalhados e válidos ao nível do programa, que possam ser usados para entender, em vez de quantificar, fatores como a forma pela qual o programa é percebido e como ou porque ele consegue seus efeitos (19). As principais diferenças entre os métodos quantitativos e qualitativos são resumidos no quadro a seguir (20, 21):

Diferenças entre os Métodos Quantitativo e Qualitativo	
Quantitativo	Qualitativo
Descreve “Quantos?” e/ou “Com que frequência?”	Descreve “Como e Por quê?”
Fornecer dados numéricos e estatísticos que permitem interpretações similares feitas por diferentes avaliadores	Fornecer dados sobre percepções, crenças, valores etc., que podem ser interpretados de várias maneiras por diferentes avaliadores
Exige amostras grandes, de preferência com seleção aleatória	Permite amostras mais limitadas, geralmente sem seleção aleatória
Geralmente requer pessoal com experiência em métodos estatísticos	Exige a participação de pessoal com experiência na análise de dados qualitativos, e não estatísticos
Os resultados podem ser generalizados para o total da população-alvo	Os resultados não podem ser generalizados com confiabilidade estatística
Usa principalmente perguntas fechadas com respostas limitadas	Usa mais perguntas abertas
Fornecer respostas menos “profundas” sobre assuntos delicados (p. ex. sexuais)	Permite respostas mais “profundas” sobre assuntos delicados (p. ex. sexuais)

Para obter um resultado da melhor qualidade, devem ser usadas metodologias tanto quantitativas quanto qualitativas na prática de avaliação abrangente de programas. Em última análise, contudo, a seleção das estratégias de avaliação depende dos objetivos, plano e recursos disponíveis.

FONTES DE DADOS QUANTITATIVOS

ENTREVISTAS ESTRUTURADAS (LEVANTAMENTOS)

As fontes mais comuns de dados quantitativos na avaliação de um programa são as entrevistas estruturadas (ou levantamentos), os questionários auto-aplicados, as estatísticas sobre os serviços do programa e outras fontes secundárias de dados. As *entrevistas estruturadas* são realizadas por um entrevistador treinado, que usa um

questionário padronizado para aplicar um conjunto de perguntas predeterminadas exatamente da mesma maneira e na mesma seqüência a uma amostra de pessoas, em geral selecionada aleatoriamente a partir de uma população-alvo específica (22). As respostas às perguntas são fechadas, quer dizer, cada resposta é antecipada e pré-codificada, o que permite que os dados coletados possam ser analisados por computador. Se a amostra é grande o bastante para permitir análise estatística, as informações podem ser expandidas para uma população maior. Esta técnica, também conhecida como estudos de CAP (Conhecimento, Atitudes e Prática), é em geral aplicada a um número maior de informantes, a fim de traçar o perfil de uma população-alvo em termos de conhecimentos, atitudes, crenças e comportamentos. Contudo, os questionários estruturados dos levantamentos podem ser usa-

dos com grupos-alvo menores, como os participantes de um programa de intervenção, para obter informações valiosas sobre um número limitado de questões, embora os resultados não possam ser generalizados.

QUESTIONÁRIOS AUTO-APLICADOS

Os *questionários auto-aplicados* são semelhantes às entrevistas estruturadas porque usam um questionário altamente estruturado com respostas fechadas e previamente codificadas. A vantagem é que esta técnica é mais barata e exige menos tempo que as entrevistas porque não demanda pessoal treinado. Entretanto, os questionários auto-administrados têm muitas limitações. Primeiro, exigem que os informantes leiam e escrevam bem o bastante para preencher eles próprios o questionário. Além disto, mesmo entre informantes alfabetizados, existem outros problemas com os questionários auto-aplicados: as instruções ou perguntas podem ser mais facilmente mal-interpretadas sem a presença de um entrevistador para explicá-las; nem todas as perguntas podem ter sido respondidas; além das dificuldades de incorporar várias seqüências de perguntas (p. ex. "Se a resposta à pergunta 12 é 'sim' passe para a pergunta 13; se não, pule para a pergunta 18"). Apesar de suas limitações, os questionários auto-aplicados podem ser utilizados em certas situações onde o nível de instrução dos participantes, como nos programas de treinamento para prestadores de serviço, não represente um problema (22).

DADOS SECUNDÁRIOS

A informação coletada de acordo com as duas técnicas mencionadas anteriormente é referida como *dados primários*, porque esses dados foram coletados recentemente durante uma atividade de avaliação. Entretanto, pode haver o

que se chama de *dados secundários*, ou seja, dados já existentes que podem ser bastante úteis quando se faz uma avaliação. Uma fonte importante de dados secundários são as *estatísticas dos serviços do programa*. Tratam-se de informações coletadas rotineiramente por muitos programas para documentar os serviços e atividades que oferecem. Os registros do programa podem fornecer informações essenciais para qualquer esforço de avaliação, particularmente para a avaliação de processos. Podem incluir informações sobre o número de participantes em várias atividades do programa, o tipo e número de atividades realizadas, materiais produzidos e distribuídos pelo programa e dados demográficos. Nos programas de prevenção de HIV/AIDS, as estatísticas de serviço ficam registradas de várias maneiras, dependendo do tipo de atividade de intervenção, por exemplo: registros administrativos de clínicas de DSTs, relatórios de atividades de prevenção, registros de treinamento de agentes de saúde e voluntários, planilhas de intervenções, materiais educativos produzidos e divulgados, preservativos distribuídos ou vendidos etc. A maioria dos programas usa algum tipo de sistema de registros para monitorar suas atividades e informar sobre resultados aos órgãos que financiam essas atividades. Contudo, sem uma monitoração cuidadosa da qualidade, os dados sobre serviços do programa freqüentemente podem ser registrados de forma incompleta ou imprecisa e por isto podem não ser confiáveis para fins de avaliação. Apesar disso, as estatísticas de serviços que foram padronizadas e estão sendo coletadas e agregadas de maneira contínua podem representar uma fonte importante de informações na avaliação de um programa.

Outras fontes de dados quantitativos secundários são os censos e as estatísticas de saúde pública, além dos estudos epidemiológicos, de vigilância sanitária e de CAP realizados por ór-

gãos governamentais estaduais ou municipais ou por organizações não-governamentais regionais (ONGs). Esses dados podem oferecer informações valiosas sobre aspectos demográficos da população em geral, além de informações relativas à AIDS/HIV/DSTs em uma área-alvo específica, tais como o número de casos de AIDS notificados, taxas de prevalência de DSTs, incidência estimada do HIV e medições básicas sobre o nível dos conhecimentos e comportamentos de risco em determinadas populações. Para fins de avaliação, estes tipos de dados podem representar uma forma barata de obter informações básicas sobre uma população-alvo e promover o estabelecimento de contatos e o intercâmbio de informação entre organizações. Conjuntos de dados de pesquisa e de população já existentes, entretanto, são mais úteis para fins de planejamento do programa, e não para atividades práticas de avaliação, que exigem fontes primárias de dados (7).

FONTES DE DADOS QUALITATIVOS

No início da epidemia de HIV/AIDS, os programas de prevenção com atividades de monitoração e avaliação em geral usavam métodos quantitativos para avaliar a eficiência das intervenções sobre o comportamento. Foram coletados dados sobre comportamentos e atitudes sexuais entre populações-alvo para quantificar resultados específicos dos programas, como o uso de preservativo, número de parceiros sexuais e níveis de conhecimento sobre HIV/AIDS. Entretanto, devido a questões complexas relacionadas a mudanças de comportamento e à natureza delicada da sexualidade individual, as informações quantitativas não conseguiam explicar inteiramente porque algumas intervenções eram eficazes e outras não. Consequentemente, estratégias de avaliação de programas começaram cada vez mais a utilizar abordagens qualitativas

para tentar compreender o comportamento individual e os fatores sociais, econômicos, culturais e políticos que o afetam.

ENTREVISTAS EM PROFUNDIDADE

As técnicas qualitativas mais amplamente usadas na avaliação de um programa são: entrevistas em profundidade, entrevistas com grupos de discussão e observação. As *entrevistas em profundidade* são conversas individuais entre o entrevistador e o entrevistado, este último selecionado por atributos determinados (p. ex. cliente, prestador de serviço, líder comunitário, informante-chave etc.). O entrevistador tem um conjunto de perguntas que servem de orientação para a discussão, mas possui flexibilidade para formular as perguntas com suas próprias palavras e na seqüência que parecer mais adequada ao curso da entrevista. O entrevistador também faz perguntas relativas aos comentários do entrevistado e dá seguimento a quaisquer tópicos que possam ter interesse. Ao contrário do questionário quantitativo, cujas respostas são simplesmente sim ou não, as perguntas neste tipo de entrevista são abertas, o que permite todo tipo de resposta possível. Na avaliação dos programas de prevenção, as entrevistas individuais representam um método de coletar informações mais aprofundadas sobre atitudes, motivações e sentimentos dos participantes do programa sobre questões delicadas relacionadas ao HIV/AIDS, além de permitir a obtenção, junto aos clientes, membros da equipe e informantes-chave, de um entendimento valioso para desenvolver e melhorar intervenções específicas do programa. Ademais, este método pode ser usado para coletar dados qualitativos para avaliação de processos ou avaliação formativa de mensagens nos meios de comunicação de massa, material educativo ou idéias preliminares para intervenções entre populações ou pequenos grupos (7). A desvantagem

principal da entrevista em profundidade é que ela exige um entrevistador treinado e experiente, com conhecimentos sobre o programa e os objetivos da avaliação, que seja capaz de dar seguimento às respostas com perguntas-chave detalhadas e, ao mesmo tempo, seja capaz de evitar impor suas opiniões ao entrevistado, de forma a não influenciar as respostas. Em segundo lugar, as entrevistas em profundidade exigem um investimento substancial de tempo e dinheiro para obter um número relativamente reduzido de entrevistas, que precisam ser gravadas e transcritas posteriormente para análise (23).

GRUPOS DE DISCUSSÃO

Os *grupos de discussão* são uma maneira mais eficiente de coletar dados qualitativos do que as entrevistas em profundidade com indivíduos, porque este método exige menos tempo e pessoal para obter informações entre um grupo de pessoas, em vez de uma só pessoa de cada vez, aumentando assim o tamanho da amostra (24). Os grupos de discussão são entrevistas cuidadosamente planejadas com seis a 12 pessoas de uma população-alvo, para discutir um tópico específico com um moderador treinado. De forma semelhante à das entrevistas em profundidade, o moderador usa um conjunto de perguntas preestabelecidas para dirigir a discussão, mas não precisa seguir uma determinada ordem ou uma forma de expressão nas perguntas. Os participantes do grupo são selecionados para compor um grupo o mais homogêneo possível, em relação a características individuais (isto é, sexo, idade, educação etc.) e experiência quanto ao tópico determinado, para proporcionar uma atmosfera confortável para uma discussão aberta. O moderador tenta minimizar a influência dos membros mais dominantes e falantes, encorajando todos a participarem da discussão e expressarem seus próprios pontos de vista no contexto

dos pontos de vista dos demais. Através da interação dos participantes, o objetivo dos grupos de discussão é gerar informações detalhadas e significativas sobre o tópico selecionado. Na avaliação dos programas de prevenção do HIV, os grupos de discussão são usados para várias finalidades: obter informações sobre o conteúdo e linguagem a serem usados nos questionários de levantamentos de CAP; testar materiais educativos ou mensagens para os meios de comunicação de massa antes de sua produção ou disseminação em grande escala; determinar as percepções de uma população-alvo sobre as operações de um programa de prevenção; entender melhor os achados de uma avaliação quantitativa; ou determinar as mudanças que ocorreram em uma população-alvo como resultado de um determinado programa ou intervenção (20).

OBSERVAÇÃO DIRETA

Enquanto a base das entrevistas em profundidade e dos grupos de discussão consiste em fazer perguntas e analisar respostas, as técnicas de observação envolvem a coleta de dados através da observação de uma atividade ou comportamento individual no ambiente social em que está ocorrendo. Os dados de observação permitem que o avaliador consiga entender um programa ou intervenção de uma forma que não é inteiramente possível quando se usam apenas as percepções de outras pessoas, obtidas em entrevistas (24). Na avaliação de um programa, as técnicas mais frequentemente usadas para coleta deste tipo de dados quantitativos são a observação direta, a abordagem simulando ser um cliente e a observação etnográfica. A *observação direta* é a observação feita no local de uma atividade do programa, conduzida por um observador experiente, para monitorar se a atividade está sendo bem realizada. As informações coletadas

dependem da atividade que está sendo observada, mas podem incluir a ordem em que os eventos acontecem e o tempo destinado para cada um, as reações e comportamentos dos participantes ou as ações do pessoal do programa. A coleta de dados deve ser registrada sistematicamente e de maneira uniforme, embora vários métodos possam ser utilizados. Por exemplo, o observador pode usar uma lista de controle preestabelecida para estimar o desempenho de um prestador de serviço, o que simplifica a análise e aumenta a probabilidade de diferentes observadores apresentarem informações consistentes (3). Em outros casos, o observador toma notas detalhadas e depois prepara uma descrição narrativa do que foi observado. Algumas informações podem ser quantificadas (p. ex. listas de controle de desempenho); entretanto, como em todos os métodos qualitativos, a coleta e análise de dados de observação podem ser muito subjetivas e os resultados dependem bastante da habilidade e experiência do observador. Outra limitação do método de observação direta é que a presença de um observador é geralmente intrusiva e pode alterar os comportamentos da equipe do programa ou dos participantes em observação. No entanto, a maioria dos avaliadores acham que este tipo de observação pode ser empregado sem afetar a validade da avaliação, se houver sensibilidade em relação ao problema do efeito causado pelo observador (24). Apesar das limitações, técnicas de observação direta podem ser uma fonte valiosa de dados na avaliação de um programa de HIV/AIDS. Por exemplo, esta abordagem poderia ser utilizada para avaliar a qualidade de uma apresentação educativa para a comunidade, um seminário de treinamento para membros da equipe ou voluntários de um programa, respostas a ligações feitas para uma “linha direta” de informação sobre HIV/AIDS, ou uma sessão de aconselhamento sobre sexo seguro ou práticas de injeção de drogas.

CLIENTE SIMULADO

Em resposta à crítica de que a equipe ou os participantes de um programa podem agir de forma diferente se estiverem conscientes de estarem sendo observados, os pesquisadores desenvolveram outra técnica de observação para avaliar programas sociais. A abordagem do *cliente simulado* ou “*misterioso*” é uma técnica em que uma pessoa se apresenta como cliente para determinar a qualidade dos serviços que estão sendo prestados pela equipe do programa, que não sabe que está sendo observada. A vantagem desta técnica é que um programa ou atividade pode ser avaliado em “condições normais” pelo observador desconhecido, a fim de obter informações a partir da mesma perspectiva dos clientes. Neste método, os dados de observação são anotados ao deixar o local da atividade, seja pelo observador a quem o serviço foi prestado, ou por um entrevistador que interroga o cliente simulado. Assim como na observação indireta, uma lista de controle é usada frequentemente para permitir que certos dados sejam quantificados. Este método, entretanto, é usado com menos frequência do que a observação direta devido a várias limitações. Primeiro, existe uma dúvida sobre se os clientes simulados podem registrar com precisão todas as suas observações após deixar o local da atividade. Segundo, se o cliente simulado for detectado pela equipe do programa, a observação é interrompida e o futuro uso desta técnica na atividade de avaliação pode ficar ameaçado. E finalmente, a equipe de gerenciamento do programa pode sentir-se pouco à vontade com esta metodologia, porque se parece com “espionagem”. Esta atitude, contudo, pode ser minimizada, informando-se antes à equipe que esta técnica será utilizada, sem especificar a data exata (20). Na avaliação dos programas de prevenção do HIV/AIDS, a abordagem do cliente simulado pode ser útil para ava-

liar os serviços de teste de HIV, aconselhamento e os prestadores de serviço em clínicas de DSTs.

OBSERVAÇÃO ETNOGRÁFICA

A *observação etnográfica* desenvolveu-se como um método de coleta de dados na avaliação de programas a partir das tradições da antropologia. Como na pesquisa antropológica, a coleta de dados etnográficos envolve trabalho de campo intensivo, no qual o pesquisador participa e observa os eventos de todo dia na comunidade ou grupo sendo estudado durante um período de tempo relativamente longo. O propósito é o pesquisador estabelecer uma relação de harmonia e confiança com os objetos do estudo e observar comportamentos, redes sociais, e eventos em um ambiente natural. Este método, entretanto, geralmente exige um longo período de tempo para ser realizado porque depende do desenvolvimento de relações de confiança entre o pesquisador e a população-alvo. No entanto, esta interação dá ao pesquisador acesso a informações detalhadas sobre comportamentos, o contexto em que ocorrem e o vocabulário usado para descrevê-los (25). Na prevenção do HIV, a observação etnográfica é uma ferramenta particularmente útil para reunir informações sobre populações de difícil acesso, como usuários de drogas injetáveis, profissionais do sexo e homens que fazem sexo com homens (HSH). Ela tem sido usada de forma eficiente em pesquisa formativa para desenvolver melhores questionários quantitativos, material educativo e estratégias de intervenção para essas populações de alto risco. Por exemplo, um estudo antropológico de homens mexicanos confirmou que muitos homens que praticam sexo com homens podem não se considerar *gays* ou bissexuais, porque na relação anal desempenham o papel ativo, e não o passivo, que pelos padrões sociais mexicanos é considerado homossexual. Consequentemente,

concluiu-se que as medidas de prevenção do HIV/AIDS deveriam ser desenvolvidas para atingir esta população de risco que não se identifica como sendo *gay* (25, 26). Em outro exemplo, observação etnológica e outras técnicas qualitativas foram utilizadas durante um período de seis meses entre profissionais do sexo em Bombay, Índia, para coletar informações sobre comportamentos de risco, a fim de desenvolver um plano estratégico para atividades de prevenção de HIV/DSTs (27). A coleta de dados etnográficos é diferente na pesquisa de HIV/AIDS porque oferece uma estrutura básica para entender melhor os comportamentos sexuais e de uso de drogas em um contexto cultural, que inclui aspectos sociais, econômicos e políticos.

TRIANGULAÇÃO

Como foi possível perceber pelas descrições acima, cada tipo de método quantitativo e qualitativo possui seus pontos fortes e fracos. Portanto, para compensar os pontos fracos de um único tipo de coleta de dados, seria ideal utilizar sempre que possível muitas fontes de informação, tanto quantitativa como qualitativa, na avaliação e pesquisa de programas. O uso de vários métodos e fontes de dados para avaliar uma intervenção ou um programa específico é chamado de *triangulação*. Além de compensar as fraquezas inerentes a uma metodologia única, a triangulação pode aumentar a validade e a credibilidade dos achados durante uma avaliação, se os resultados produzidos pelos diferentes métodos forem os mesmos (3, 24). Além disso, a combinação de vários tipos de coleta de dados pode proporcionar um retrato mais completo da intervenção ou programa que está sendo avaliado.

Por exemplo, a triangulação de dados foi usada para fazer uma avaliação de processo e de resultado de uma intervenção entre pares para prevenção de DST/HIV realizada por uma orga-

nização comunitária, que escolheu como alvo uma área de um parque público em Londres, Reino Unido, freqüentada por homens que vão lá para encontrar parceiros sexuais masculinos. A finalidade principal da intervenção foi promover práticas sexuais mais seguras entre os homens que se encontram no parque, através da distribuição de preservativos e material educativo. Foram empregados vários métodos de pesquisa na avaliação, inclusive qualitativos (isto é, observação direta, entrevistas em profundidade, grupos de discussão) e quantitativos (isto é, levantamentos, dados de distribuição de preservativos, registros administrativos) (28). Este tipo de abordagem de triangulação, utilizando vários métodos, permitiu uma perspectiva abrangente dessa intervenção, o que de outra forma poderia não ter sido possível devido às dificuldades de realizar pesquisas entre essa população difícil de atingir. Em Lima, Peru, foi feita uma triangulação para avaliar a percepção de risco de HIV/AIDS em uma amostra de mais de 500 adolescentes. A pesquisa quantitativa de CAP mostrou que o HIV/AIDS não era percebido como um risco pela amostra e que o uso do preservativo era limitado. Contudo, foi através do uso de dramatizações, entrevistas em profundidade e grupos de discussão que os pesquisadores conseguiram entender melhor “por que” esses resultados ocorreram (29). Combinar os vários tipos de técnicas de coleta de dados representa uma estratégia importante na pesquisa e avaliação de programas e pode oferecer um maior entendimento dos esforços de prevenção.

TIPOS DE DESENHOS DE PESQUISA

DESENHOS DE PESQUISA PARA AVALIAÇÕES DE PROCESSO

Além da seleção das técnicas para coleta e análise de dados, outro aspecto crucial de qual-

quer atividade de avaliação é determinar o desenho geral da pesquisa. Ambas as decisões dependem das necessidades e recursos do programa e do tipo de avaliação a ser feita. Avaliações de processo e de resultado exigem diferentes tipos de desenho de pesquisa, enquanto as avaliações formativas usam uma mistura desses desenhos. Como já foi mencionado, a avaliação de processo consiste de descrição e documentação de que tipo de serviços foram prestados e a quem. O desenho de pesquisa para este tipo de avaliação pode incluir tanto dados quantitativos (p. ex. levantamentos da equipe e participantes do programa, estatísticas de serviços prestados) e dados qualitativos (p. ex. grupos de discussão, entrevistas, observação), que de forma ideal seriam medidos a intervalos periódicos durante toda a duração do programa, para garantir que seus objetivos estejam sendo realizados. Por exemplo, uma avaliação de processo em seis clínicas de saúde reprodutiva em Nairóbi, Quênia, foi realizada para determinar a prestação de serviços e a qualidade da assistência, através da avaliação de prestação de informações sobre uso de preservativo, DSTs e HIV/AIDS e distribuição de preservativos. Usando observações de cliente-prestador e entrevistas de liberação de clientes, foi identificado que as informações esperadas estavam sendo dadas de forma inconsistente e, por esta razão, deu-se início ao treinamento dos prestadores de serviço e a um sistema para monitoração da distribuição de preservativos. Os resultados da avaliação de acompanhamento revelaram que a prestação dos serviços nas clínicas havia melhorado significativamente após as ações corretivas (30).

Além de seu papel no aperfeiçoamento das intervenções de um programa, um processo bem planejado de avaliação pode oferecer percepções valiosas da relação entre a intervenção e o resultado. Por exemplo, uma avaliação de processo e de resultado foi realizada para medir os efeitos

de um programa de prevenção de HIV/AIDS entre trabalhadores agrícolas na área rural do Zimbábue. O objetivo geral da intervenção na comunidade era diminuir o risco de infecção por DST/HIV através do sexo sem proteção e aumentar a participação dos trabalhadores agrícolas em atividades de prevenção da AIDS. Dois dos indicadores selecionados para medir o resultado do programa foram a distribuição de preservativos e a incidência de DSTs. No fim do período de estudo, dados relativos a ambos os indicadores de resultado apresentaram resultados positivos (a saber, aumento da distribuição de preservativos e queda da incidência de DSTs). Entretanto, informações da avaliação de processo, que incluíam relatórios de atividades entre os trabalhadores, resumos de grupos de discussão, registros de distribuição de preservativos e registros de clínicas, demonstraram que outros fatores afetaram os resultados. A análise dos dados de processo revelaram que os preservativos grátis do programa atraíram trabalhadores de lugares fora da área de intervenção e que outros preservativos foram levados para serem revendidos. Da mesma forma, o declínio da incidência de DSTs refletiu a falta de tempo durante os meses de pico da lavoura para freqüentar clínicas, e não uma queda real de infecção por DSTs (31). Este exemplo ilustra como o processo de avaliação pode ser usado para complementar a avaliação de resultado e oferecer um quadro mais claro de como um programa de intervenção funciona ou não funciona na realidade.

DESENHOS DE PESQUISA PARA AVALIAÇÕES DE RESULTADO

Desenho experimental aleatório

A avaliação de resultado é desenhada para estimar a efetividade de intervenções específicas de um programa e tenta responder às perguntas: “O programa ou intervenção funciona?” e “É

capaz de fazer diferença?”. Existem três tipos usuais de desenhos de pesquisa usados em avaliações de resultado: desenhos experimentais aleatórios ou “verdadeiros”, desenhos quase-experimentais e desenhos não-experimentais. É um fato amplamente aceito entre os especialistas em avaliação tradicionais que o desenho experimental aleatório de pesquisa é o padrão-ouro para avaliar os efeitos de um programa ou intervenção. No *desenho experimental aleatório*, os indivíduos de uma população-alvo são designados aleatoriamente para fazer parte do grupo experimental que sofre a intervenção ou “tratamento” ou para o grupo de controle que não sofre a intervenção. O grupo de controle permite uma estimativa sobre o que teria acontecido ao grupo experimental se a intervenção que está sendo avaliada não tivesse sido implementada. A aleatorização é obtida jogando-se uma moeda ou usando-se qualquer outro método probabilístico para determinar que todas as pessoas terão uma chance igual de serem designadas para cada grupo. Se utilizado corretamente, com uma amostra suficientemente grande, pode-se supor com alguma certeza que o grupo experimental e o grupo de controle serão comparáveis distribuindo-se fatores extrínsecos igualmente em cada grupo. Em outras palavras, a aleatorização garante que o grupo experimental e o grupo de controle serão equivalentes em relação a todos os fatores que não sejam a exposição à intervenção ou ao programa que está sendo avaliado. Portanto, quaisquer diferenças observadas entre os dois grupos após a intervenção ser introduzida teoricamente são devidas com maior probabilidade aos efeitos da intervenção, e não a uma designação tendenciosa de pessoas para cada um desses grupos (3).

Dois dos desenhos experimentais aleatórios usados mais comumente são o *desenho dos grupos de controle com teste prévio e posterior* e o *desenho do grupo de controle só com teste posterior*. No primeiro

desenho, dados-base (isto é, teste prévio) são coletados para o grupo experimental e para o grupo de controle antes que a intervenção do programa seja implementada. O grupo experimental sofre então a intervenção, enquanto o grupo de controle não sofre e os dados (isto é, o teste posterior) são coletados novamente após um período de tempo considerado suficiente para que a intervenção apresente seu resultado pretendido. Estes dois conjuntos diferentes de dados originados em cada grupo são então comparados e analisados a fim de avaliar a efetividade da intervenção. Por exemplo, este desenho foi usado em um estudo para determinar os efeitos de uma intervenção em um pequeno grupo, visando reduzir os comportamentos de risco em relação ao HIV e aumentar o uso do preservativo entre um grupo de 157 adolescentes afro-americanos do sexo masculino. Após o preenchimento de um questionário prévio auto-administrado, o grupo foi distribuído aleatoriamente para sofrer uma intervenção sobre redução de risco em relação a AIDS ou participar de um grupo de controle sobre orientação profissional. Questionários de acompanhamento foram distribuídos entre os dois grupos imediatamente após a intervenção e novamente três meses mais tarde para medir quaisquer mudanças em comportamentos de risco ou uso de preservativo (32).

O desenho do grupo de controle só com teste posterior é semelhante ao que foi discutido acima, com exceção de não coletar dados-base prévios porque se supõe que a aleatorização produziu grupos experimental e de controle que são equivalentes antes da intervenção do programa. Por exemplo, 426 pacientes inscritos em uma clínica de DSTs nos Estados Unidos foram designados aleatoriamente para um grupo experimental ou para um grupo de controle por um período de cinco semanas, dependendo do seu prazo de inscrição. O grupo experimental sofreu uma intervenção de aconselhamento em grupo

para promover o uso de preservativo e o grupo de controle recebeu os serviços de rotina disponíveis na clínica. Nove meses depois, os registros médicos de ambos os grupos foram revistos e ficou determinado que os homens que participaram da intervenção apresentaram uma taxa significativamente mais baixa de reinfecção por DSTs do que os que não participaram (33). Desta forma, o desenho de grupo de controle só com teste posterior apenas mede o efeito de uma intervenção no grupo experimental comparando aquele grupo com o de controle. Uma desvantagem deste tipo de desenho, entretanto, é que, ao contrário do desenho de grupo de controle com testes prévio e posterior, ele não permite a determinação da magnitude da mudança dentro do grupo experimental em resultado da intervenção, porque não são coletados dados-base (22).

Como já mencionamos, o ponto de vista tradicional entre a maioria dos especialistas em avaliação é que os experimentos aleatórios constituem o melhor desenho para estimar os efeitos de uma intervenção específica. No entanto, suas limitações são também amplamente reconhecidas mesmo pelos defensores mais ardorosos deste tipo de desenho. Primeiro, os desenhos experimentais aleatórios podem ser difíceis de implementar por razões políticas e práticas, na medida que exigem muito tempo, experiência e recursos financeiros. Se realizada corretamente, a aleatorização envolve números relativamente grandes de pessoas para garantir que os grupos de intervenção e de controle sejam comparáveis e para permitir resultados estatísticos significativos. Além disso, em lugar de realizar apenas uma medição dos efeitos do programa imediatamente após a intervenção, muitos desenhos experimentais incluem múltiplos estudos de acompanhamento dos participantes na avaliação, o que pode proporcionar informações valiosas sobre os efeitos a longo prazo da intervenção.

Entretanto, isso também acarreta maiores custos e complexidade na avaliação. Se os objetos do estudo pertencem a uma população difícil de atingir, os estudos de acompanhamento provavelmente serão impossíveis. Outra consideração é que, para obter medidas válidas e precisas do resultado, um desenho experimental aleatório freqüentemente exige um alto grau de controle das atividades do programa durante o tempo em que a avaliação está sendo realizada. Um controle rigoroso de muitas das variáveis envolvidas em um experimento aleatório (p. ex., seleção dos objetos do estudo, participação da equipe e da gerência do programa, alocação dos recursos etc.) pode não ser possível em muitas situações.

Em segundo lugar, há numerosos problemas metodológicos que podem surgir no uso de um desenho experimental aleatório. Participantes que desistem tanto do grupo experimental como do grupo de controle podem ter um efeito de tendenciosidade sobre os resultados do estudo de avaliação. Por exemplo, pode ocorrer um atrito diferenciado no grupo experimental se a intervenção que está sendo avaliada exigir muito esforço por parte dos participantes. Em conseqüência, os efeitos do programa poderiam ser superestimados porque só as pessoas altamente motivadas continuariam no grupo. Outro problema metodológico potencial do desenho experimental aleatório é o efeito de Hawthorne, no qual os objetos da experiência mudam seu comportamento de maneiras que acreditam ser esperadas deles. Isto pode afetar a validade e os resultados gerais de uma avaliação. Outra ameaça para validade de uma avaliação é a contaminação. Isto ocorre quando o grupo de controle é exposto a algum aspecto da intervenção pela difusão de informações oriundas do grupo experimental e pode resultar na subestimação dos efeitos reais do programa. O atrito diferenciado, o efeito de Hawthorne e a contaminação são apenas três dos vários problemas metodológicos

que podem ocorrer no uso de experimentos aleatórios. Em conseqüência, este tipo de desenho de pesquisa exige o envolvimento de um avaliador profissional e geralmente está além da capacidade da maioria das organizações comunitárias sem assistência externa.

E terceiro, o uso de desenhos experimentais com freqüência dá origem a questões éticas controversas. Nos programas relacionados ao HIV/AIDS o argumento ético mais comum contra o uso de aleatorização é o de que, ao designar alguns objetos do estudo para o grupo de controle, a intervenção ou serviço do programa estará sendo negada a pessoas que poderiam ser beneficiadas. Tem havido casos, especialmente no mundo em desenvolvimento, nos quais pessoas que não receberam benefícios em virtude da avaliação de um projeto de AIDS, e dependendo da intervenção sendo avaliada, isto pode ser um ponto complicado. Contudo, em qualquer atividade de pesquisa ou avaliação de um programa, outras estratégias, como o tratamento posterior do grupo de controle, podem ser usadas para evitar que um grupo de controle não receba qualquer tipo de tratamento, o que pode ser particularmente prejudicial para os participantes. Serviços eficazes de prevenção, que de outra forma estariam disponíveis, nunca devem ser negados a qualquer pessoa para fins de avaliação. Se isto não for possível, devem ser usados outros desenhos para a pesquisa.

Desenho quase-experimental

Pelas razões discutidas acima, um desenho experimental aleatório freqüentemente não é apropriado ou factível, e portanto outros tipos de desenhos de pesquisa precisam ser considerados. Os *desenhos quase-experimentais* são desenhos de pesquisa em que indivíduos ou outras unidades do estudo não são designados aleatoriamente para o grupo de intervenção ou para o grupo de controle. Existem diversas variações do desenho

quase-experimental; contudo, o tipo mais comumente usado (a saber, o desenho do grupo de controle não-equivalente, com testes prévio e posterior) é idêntico ao desenho experimental aleatório com testes prévio e posterior, com exceção do fato de não se utilizar aleatorização para designar as pessoas para o grupo experimental ou para o grupo de controle. Em vez disso, após um grupo experimental ser identificado, ele é “pareado” com um grupo de controle, de modo que as pessoas em cada grupo sejam tão semelhantes quanto possível em termos de características e experiências que possam estar relacionadas ao resultado do programa ou intervenção sendo avaliado. Essas características podem incluir sexo, idade, situação socio-econômica, etnia, localização geográfica, comportamentos de risco e outras que possam ser relevantes para o estudo. Como no desenho experimental aleatório, os efeitos do programa ou intervenção são então medidos comparando-se as diferenças entre o grupo de intervenção e o de controle (ou comparação).

Um exemplo de desenho não-equivalente com teste prévio e posterior pode ser visto na avaliação de uma intervenção visando mulheres profissionais do sexo na Índia. Foram recrutadas de forma não-aleatória, no local de uma intervenção, 334 profissionais do sexo e mais 207 outras profissionais de um local de controle similar na área de bordéis de Bombay. Depois que as mulheres de cada grupo foram examinadas para detectar HIV, sífilis e hepatite B (ou seja, o teste prévio), o grupo experimental sofreu uma intervenção durante seis meses, incluindo vídeos educativos, discussões com pequenos grupos e material educativo ilustrado para promover o uso de preservativos. Foram feitos novos exames de sangue (ou seja, o teste posterior) em todas as participantes um ano após as atividades de intervenção serem encerradas e as taxas de soroc conversão foram significativamente diferentes

nos dois grupos (34). Usando o desenho quase-experimental, os resultados dessa avaliação demonstraram claramente a eficácia da intervenção. Além disso, o exemplo mostra a versatilidade do desenho quase-experimental em avaliações de programas de prevenção que visam populações onde não é possível a aleatorização devido a características especiais de grupos difíceis de atingir.

Como os experimentos aleatórios, o desenho quase-experimental pode usar comunidades, em lugar de indivíduos, como a unidade do estudo em uma avaliação de resultado. Por exemplo, este desenho foi usado para avaliar os efeitos de uma intervenção para prevenção do HIV ao nível da comunidade sobre comportamentos de uso de preservativo entre mulheres de baixa renda, primariamente afro-americanas, em quatro comunidades nos Estados Unidos. A seleção das comunidades para intervenção foi baseada em critérios predeterminados que incluíam a presença de mulheres em alto risco, indicado pelas altas taxas de uso de drogas e DSTs, tamanho da população da comunidade e a presença de poucas atividades de prevenção do HIV. As comunidades pareadas para comparação foram escolhidas para serem tão semelhantes quanto possível às comunidades da intervenção, atendendo aos mesmos critérios. Além disso, foram usados dados censitários para garantir que as comunidades de intervenção e de comparação eram comparáveis na distribuição por idade, composição racial e situação econômica. Usando um plano de amostragem, as mulheres foram então entrevistadas individualmente em cada comunidade pareada de intervenção e de comparação, antes e depois da intervenção, que consistia de múltiplas atividades para promover o uso de preservativo durante um período de dois anos. Foram feitas análises estatísticas dos dados de entrevista para medir os efeitos da intervenção, comparando-se a

mudança ao longo do tempo no uso de preservativos pelas mulheres da comunidade de intervenção e comparação (35).

Como foi visto no exemplo acima, os quase-experimentos freqüentemente precisam compensar as diferenças entre os grupos experimental e de controle, pareando características relevantes e/ou usando um método estatístico apropriado. De outra forma, poder-se-ia argumentar que os efeitos observados em uma avaliação de resultado estão associados com as diferenças nas características dos dois grupos sendo comparados, em vez de serem um resultado direto da intervenção. Este fator, conhecido como tendência (ou vício) de seleção, constitui um problema potencial maior nos desenhos quase-experimentais, quando comparados com o desenho aleatório em que se pressupõe que variáveis que não são as da intervenção sendo avaliadas estão distribuídas igualmente no grupo experimental e no de comparação por atribuição aleatória. Assim, a validade de um desenho quase experimental depende de uma seleção cuidadosa dos grupos experimental e de controle, para assegurar que ambos os grupos são tão similares quanto possível em relação a características relevantes que podem afetar o resultado da intervenção. Além do mais, como muitos desenhos quase-experimentais usam um grupo de controle (ou de comparação), eles podem estar sujeitos aos mesmos problemas metodológicos (a saber, atrito diferenciado, efeito de Hawthorne, contaminação) que o desenho experimental aleatório. No entanto, quando corretamente desenhados e analisados, os quase-experimentos podem oferecer provas poderosas dos resultados de um programa ou intervenção. Contudo, a experiência técnica e os custos, relativos ao pareamento e à análise estatística, que geralmente são necessários em um desenho quase-experimental bem-executado, provavelmente serão proibitivos para a maioria das organizações comunitárias.

Desenho não-experimental

O terceiro tipo de desenho de pesquisa usado em avaliações de resultado é o *desenho não-experimental* (às vezes chamado de estudo de observação). Para as organizações comunitárias, este desenho é mais prático do que os desenhos aleatórios e os quase-experimentais porque em geral é menos dispendioso e tecnicamente exigente. Um desenho não-experimental não envolve comparação entre um grupo experimental e um grupo de controle para medir os efeitos de uma intervenção. Ao contrário, no tipo mais comum de desenho não-experimental, conhecido como *estudo de antes e depois*, medições feitas antes da intervenção são comparadas com medições similares feitas após a intervenção na mesma população-alvo, para determinar se ocorreu alguma mudança. Por exemplo, este desenho foi usado para realizar uma avaliação de resultado de uma intervenção de educação sanitária/redução de risco entre usuários de drogas injetáveis. O comportamento de risco no uso de seringas foi medido em entrevistas estruturadas imediatamente antes da intervenção, que consistiu de aconselhamento individual com orientadores que não faziam parte do grupo, e três a cinco meses após a intervenção. Os resultados indicaram estatisticamente quedas significativas nos comportamentos de risco; entretanto, esses achados só podem sugerir que a intervenção foi efetiva. Como na maioria dos desenhos não-experimentais, não é possível atribuir os resultados deste estudo somente à intervenção. Inúmeros outros fatores poderiam ter influenciado as diferenças entre as medições feitas antes e depois da intervenção, inclusive os efeitos de outras atividades de intervenção contra a AIDS, a exposição aos meios de comunicação de massa, a interação entre pessoas no grupo de pares, amadurecimento dos participantes e os efeitos educativos e sensibilizadores do próprio trabalho de medição.

O exemplo acima ilustra uma limitação do

desenho não-experimental, na medida em que sem grupos de controle é muito mais difícil – a maioria dos técnicos em avaliação diria que é impossível – mostrar uma relação causal entre a intervenção de um programa e uma mudança observada em uma população-alvo, devido a múltiplos fatores que poderiam afetar as variáveis de resultado ao longo do tempo. Contudo, nas avaliações de resultado em que a causalidade não é um objetivo, o uso de um desenho não-experimental é apropriado. Por exemplo, este desenho foi usado para avaliar a eficácia da intervenção de um grupo de aconselhamento em sete sessões, visando reduzir comportamentos sexuais de risco entre homens *gays* nos Estados Unidos. Foram coletados dados relevantes antes, imediatamente após a intervenção e novamente oito meses mais tarde, entre os 15 participantes do estudo. A avaliação foi realizada para determinar se os mesmos resultados positivos poderiam ter sido obtidos em sete sessões, em comparação com um estudo de intervenção anterior com objetivos semelhantes, mas que foi realizado em 12 sessões. Usando um desenho experimental aleatório e um tamanho maior de amostra, a avaliação anterior demonstrou uma relação causal entre a intervenção e o resultado bem-sucedido no grupo-alvo (36, 37). Desta forma, para fins do segundo estudo, no qual a causalidade não era um objetivo, um desenho não-experimental era mais que suficiente para fornecer informações valiosas sobre como melhorar um programa de intervenção.

Outro exemplo mostra a força de um estudo não-experimental na avaliação de resultado. Um programa de prevenção do HIV entre pares para aumentar o uso do preservativo entre profissionais do sexo foi desenvolvido e avaliado usando um desenho não-experimental, em uma área urbana no norte da Tailândia. A intervenção com duração de um ano, que incluía sessões de treinamento repetidas com grupos pequenos,

conduzidas por educadores do grupo, visava fazer profissionais do sexo e donos de bordel insistirem no uso obrigatório de preservativos pelos clientes. Medições prévias e posteriores foram baseadas em observação direta por um cliente simulado treinado que solicitou e verificou se uma subamostra de profissionais do sexo iria insistir no uso do preservativo. A proporção de mulheres que recusaram fazer sexo sem preservativo aumentou significativamente em seguida à intervenção e, de acordo com os pesquisadores, este projeto demonstrativo influenciou o governo tailandês a adotar o programa de bordéis com 100% de uso do preservativo (38).

UMA NOTA SOBRE A ESCOLHA DE DESENHOS DE PESQUISA

Durante vários anos, tem havido um debate entre os técnicos em avaliação de formação acadêmica sobre a maneira mais apropriada de avaliar a eficácia de intervenções contra o HIV/AIDS. De um lado, argumenta-se que o desenho experimental aleatório é o “padrão-ouro” na prática de avaliação e que todas as outras abordagens em avaliação de resultado não são confiáveis devido à sua falta de rigor científico. Outros afirmam que o “modelo científico” (isto é, aleatorização) não é apropriado para avaliação de programa e que “vários métodos de pesquisa (incluindo estudos observacionais, desenhos quase-experimentais, projetos de demonstração e avaliações de intervenções e mudanças sociais) são necessários para esclarecer estratégias efetivas de prevenção do HIV” (39). Mesmo os proponentes mais ansiosos, que defendem uma gama mais ampla de metodologias de avaliação, concordariam que um estudo experimental prospectivo bem-desenhado pode proporcionar provas poderosas de eficácia na avaliação de intervenções sobre o comportamento. No entanto, eles argumentariam que muitos dos critérios necessários

para realizar tais estudos limitam sua utilidade primariamente a programas de uma única intervenção em ambientes razoavelmente controlados (10). Ademais, ao não usar métodos qualitativos, que oferecem uma perspectiva mais ampla da intervenção que está sendo avaliada, os desenhos experimentais limitam o escopo da avaliação; e portanto estão geralmente mal equipados para compreender os múltiplos fatores individuais e ao nível da comunidade relacionados à mudança no comportamento sexual. Esses críticos sustentariam que outros tipos de estudo são mais factíveis por razões práticas e metodológicas, especialmente considerando os recursos necessários para realizar um experimento aleatório. Esta controvérsia sobre os méritos relativos de um desenho experimental aleatório, contudo, é relativamente supérflua para as organizações comunitárias. Desenhos de pesquisa rigorosos, inclusive os quase-experimentais, são dispendiosos e exigem muito tempo, além de experiência técnica que a maioria dessas organizações não possui e só em raras ocasiões, com significativo apoio externo, se poderia esperar de um programa de prevenção ao nível da comunidade a realização de desenhos de pesquisa tão elaborados.

DESENHOS DE PESQUISA PARA AVALIAÇÃO FORMATIVA

As avaliações formativas são realizadas antes da implementação em grande escala de uma intervenção ou programa. Durante este tipo de avaliação, as informações são coletadas e apresentadas aos responsáveis pelas decisões para que possam entender melhor a necessidade da intervenção e ajudar no desenho e desenvolvimento do programa. A seleção do desenho de pesquisa mais apropriado nas avaliações formativas depende dos objetivos propostos pela intervenção que vai ser avaliada e, é claro, dos recursos disponíveis para o programa. As avaliações formativas

usam no desenho da pesquisa abordagens semelhantes às utilizadas tanto pela avaliação de processo como pela de resultado, além de técnicas de coleta de dados e fontes similares. Como nas avaliações de processo, este desenho de pesquisa usa uma variedade de técnicas quantitativas e qualitativas para reunir informações relevantes sobre o desenvolvimento do programa. Por exemplo, usando múltiplos métodos de coleta de dados e informações, uma avaliação formativa pode ser realizada para identificar necessidades culturalmente específicas de prevenção do HIV em um grupo minoritário ou uma população difícil de atingir (este tipo de estudo para determinar as necessidades não atendidas de uma determinada população, comunidade ou área geográfica é freqüentemente denominado de estimacão de necessidades, na literatura sobre avaliação). Nos estágios iniciais do desenho de um programa, a avaliação formativa pode também usar um desenho de pesquisa que lembra uma avaliação de resultado para estimar se uma determinada estratégia de intervenção funciona melhor que outra. Por exemplo, antes de publicar e divulgar um folheto de prevenção de HIV/AIDS, uma organização comunitária pode designar aleatoriamente algumas pessoas para examinar diferentes versões ou projetos do folheto e comparar suas preferências. Ou poderia ser desenvolvido um questionário prévio e posterior sobre conhecimentos, atitudes e comportamentos pretendidos para testar a efetividade de cada folheto. Esses desenhos experimentais e não-experimentais em pequena escala podem ser úteis também na previsão dos efeitos de uma intervenção, assim como na seleção das estratégias de prevenção mais efetivas antes de sua implementação geral.

Um exemplo de estudo de pesquisa formativa bem-executado são os Projetos de Demonstração Comunitária em AIDS, que visaram grupos difíceis de atingir (p. ex., usuários de

drogas injetáveis (UDIs), parceiras sexuais de UDIs, mulheres profissionais do sexo, homens não identificados como *gays* que fazem sexo com homens e jovens em situações de alto risco) em cinco cidades dos Estados Unidos. As metas dos projetos para mudança de comportamento eram o uso consistente de preservativo e, quando apropriado, o uso consistente de água sanitária para limpar equipamento de injeção, entre a população-alvo. Além da fase de pesquisa formativa, as atividades de pesquisa no projeto de cinco anos incluíram avaliações de processo e de resultado para avaliar a efetividade do uso de voluntários da comunidade para realizar a intervenção teórica que envolvia a disseminação de mensagens e materiais de prevenção. Em relação à pesquisa formativa, ela focalizou a identificação, o acesso e o entendimento da população-alvo através do uso de fontes de dados secundários relevantes, observação etnográfica, grupos de discussão e entrevistas semi-estruturadas com pessoas que tinham contato ou acesso aos grupos-alvo (isto é, “*gatekeepers*”) e com membros dos diferentes grupos-alvo. A pesquisa forneceu informações valiosas sobre as populações-alvo em cada um dos locais, o que foi usado para desenhar e desenvolver o programa de intervenção (40, 41). Deve-se mencionar, entretanto, que este esforço em grande escala foi realizado em condições ideais, com considerável apoio técnico e financeiro, e que seria extremamente difícil para qualquer organização comunitária adaptar este modelo em sua totalidade. Entretanto, o método sistemático apresentado neste estudo poderia oferecer uma estrutura útil para as organizações que estivessem pensando em realizar uma atividade de avaliação formativa.

AVALIAÇÃO PARTICIPATIVA

Em reação contra as metodologias correntes de avaliação que dominam a literatura, várias

abordagens alternativas foram desenvolvidas. Um método de pesquisa que tem sido usado cada vez mais por organizações comunitárias, especialmente nos países em desenvolvimento, é a avaliação participativa. Esta abordagem consiste em engajar ativamente as pessoas objeto do programa, a equipe e os voluntários e outros membros da comunidade na atividade de avaliação. A premissa básica é de que ninguém conhece melhor um programa do que aqueles que o utilizam ou o estão implementando. Pretende-se que seja um esforço igualitário, no qual a perspectiva do avaliador recebe a mesma prioridade que a de outras pessoas envolvidas na atividade. O objetivo geral de uma avaliação participativa é oferecer informações relevantes e uma compreensão profunda de um programa ou intervenção para melhorar sua implementação e seus resultados, ao mesmo tempo em que é prático, útil e traz poder para aqueles que realizam a avaliação, no sentido de promover o desenvolvimento pessoal e organizacional. Existem outros benefícios dessa abordagem colaborativa, tais como reduzir a desconfiança e o medo de atividades de avaliação, aumentar a conscientização e o compromisso com o programa, ampliar o conhecimento, fortalecer as parcerias, permitir diferentes perspectivas e aumentar a possibilidade de que as descobertas serão utilizadas (8, 10).

Um exemplo de avaliação participativa foi aquela realizada por uma organização não-governamental (ONG) em Uganda, utilizando curandeiros tradicionais na atividade de prevenção de HIV/AIDS e prestação de assistência. Facilitada por consultores técnicos, a avaliação foi um esforço de colaboração envolvendo a equipe do projeto, os curandeiros, membros das comunidades onde os curandeiros trabalhavam, responsáveis pelas políticas e outros interessados. Várias técnicas quantitativas e qualitativas para coleta de dados foram utilizadas, tais como um levantamento dos curandeiros tradicionais,

grupos de discussão envolvendo membros da comunidade, entrevistas com informantes-chave e observação direta de sessões de aconselhamento e eventos educativos na comunidade orientados pelos curandeiros. A avaliação apresentou vários achados sobre a efetividade do programa de intervenção e as atitudes dos participantes na avaliação, inclusive os próprios curandeiros, sobre o papel de curandeiros tradicionais como orientadores treinados e educadores comunitários. A conclusão foi que o processo participativo da avaliação permitiu que os membros da equipe, que desenharam e implementaram a atividade, aceitassem e “admitissem” os resultados, assim como estimassem sua própria contribuição na conquista das metas organizacionais. Ademais, “o processo de coletar e analisar nossos próprios dados, e discutir os achados positivos e negativos, permitiu que a organização reavaliasse sua estratégia de intervenção e focalizasse mais efetivamente as necessidades da comunidade e dos curandeiros tradicionais” (42).

PONTOS NA AVALIAÇÃO DE PROGRAMA

TIPOS DE AVALIADOR: INTERNO E EXTERNO

A maioria das organizações comunitárias reconhece a necessidade de monitoração contínua e avaliação das atividades de prevenção do HIV para saber se as intervenções do programa estão atendendo efetivamente às necessidades de seus clientes, ou para identificar onde podem ser introduzidos aperfeiçoamentos. No entanto, não existe consenso sobre a melhor maneira de se fazer uma avaliação de programa, além de várias questões relacionadas a isso. Uma questão é quem estaria mais preparado para realizar uma avaliação. Alguns sustentam que só aqueles que já trabalham para a organização podem entender integralmente suas necessidades e objetivos, bem como as operações diárias do programa. Além

disso, um avaliador interno estaria mais familiarizado com os demais membros da equipe e da comunidade, teria acesso a recursos organizacionais e mais oportunidade de um retorno informal por parte dos clientes e outras pessoas associadas com o programa. Por outro lado, argumenta-se que os avaliadores externos, que são contratados fora da organização, representam uma escolha melhor porque em geral possuem conhecimentos mais amplos sobre avaliação do que os avaliadores internos, podendo assim trazer uma perspectiva diferente para o processo porque não estão diretamente comprometidos com a organização. Proponentes desta abordagem sustentam às vezes que os avaliadores externos são mais “objetivos” que aqueles que pertencem à organização. Deve-se observar, no entanto, que “...a avaliação nunca é uma atividade neutra e objetiva. Decisões precisam ser tomadas sobre os elementos do programa a serem focalizados, os indicadores intermediários e de rendimento a serem examinados e também os aspectos do trabalho que não serão avaliados devido à falta de tempo e de dinheiro” (5). Ademais, os avaliadores podem ter opiniões ou preconceitos a respeito de determinadas populações afetadas pelo HIV (p. ex. homens que fazem sexo com homens, usuários de drogas injetáveis, profissionais do sexo), além de suas próprias agendas pessoais, tais como a necessidade de gerar renda e publicar artigos, o que também pode afetar os resultados de uma avaliação.

Uma terceira opção na seleção de um avaliador combina as qualidades dos tipos de avaliador interno e externo e parece uma avaliação participativa. Um membro (ou membros) da equipe faz a avaliação e um consultor externo dá assistência em alguns dos aspectos técnicos (p. ex. desenho da avaliação, análise estatística) e ajuda a reunir informações especializadas. Isto permite que a avaliação tenha uma perspectiva externa, mantendo ainda o benefício de um avaliador

interno com seu conhecimento de primeira mão sobre o programa (10). A decisão final na seleção de um avaliador depende de vários fatores, inclusive a disponibilidade de recursos, o conhecimento técnico da equipe e o objetivo da avaliação. Entretanto, quer a escolha recaia em um avaliador interno, quer em um externo, ou ainda em uma combinação de ambos, a seleção deveria ser determinada pela consideração sobre que tipo de avaliador seria mais capaz de proporcionar informações válidas e significativas sobre o programa que está sendo avaliado, e ao mesmo tempo desenvolver a capacidade dentro da organização para realizar no dia-a-dia e de forma apropriada a monitoração e a avaliação de atividades do programa.

FINALIDADE GERAL DA AVALIAÇÃO DE PROGRAMA

Outro ponto importante é a finalidade geral da avaliação de programa no trabalho com HIV/AIDS. Durante as últimas décadas, doadores e responsáveis por políticas exigiram cada vez mais que os programas sociais, incluindo mais recentemente os programas de prevenção do HIV, realizassem atividades de avaliação para produzir provas da efetividade ou dos “resultados” de suas iniciativas. Existem várias razões para esta ênfase recente dada à “responsabilidade” dos programas sociais, inclusive recursos limitados, problemas sociais cada vez mais complexos e interrelacionados, mudanças nas prioridades e políticas institucionais e governamentais, e variação da opinião pública sobre até onde um programa social deve apoiar populações desamparadas e vulneráveis. Consequentemente, muitos avaliadores, gerentes de projeto e outros responsáveis pelas decisões começaram a encarar a avaliação de programa somente como uma forma de demonstrar a efetividade ou de “provar” o mérito de um programa, e a avaliação é geralmente dirigida para mostrar que um programa

ou intervenção funciona, em vez de servir também para melhorar a maneira como funciona (10). O resultado desta perspectiva histórica limitada é que os esforços de avaliação feitos pela maioria dos programas de prevenção do HIV, como outros programas sociais, focalizam em geral a apresentação de dados quantitativos para medir resultados selecionados, através do uso de estatísticas de serviço ou indicadores de desempenho, a fim de atender às exigências de relatório das instituições financiadoras. Embora seja obviamente necessário demonstrar a efetividade e estimar os resultados de um programa, além de atender às demandas dos doadores, focalizar exclusivamente resultados quantificáveis é uma visão limitada e enganosa. Avaliações convencionais de resultado, sozinhas, não podem levar a compreender como e por que os programas funcionam, para quem e em que circunstâncias, nem ajudar a perceber como se pode aperfeiçoar um programa. A prática correta da monitoração e da avaliação envolve múltiplas abordagens através de uma combinação dos diferentes tipos de avaliação (p. ex. de resultado e de processo) e os vários métodos de coleta de dados (p. ex. quantitativos e qualitativos) para melhorar a efetividade do programa e seu futuro crescimento.

ADAPTABILIDADE CULTURAL DOS MÉTODOS DE AVALIAÇÃO DE PROGRAMA

Apesar do fato de mais de 90% das infecções globais pelo HIV ocorrerem no mundo em desenvolvimento (43), a grande maioria dos estudos de pesquisa e avaliação de programa relatados na literatura corrente vem da América do Norte e da Europa. Isso não deve causar surpresa se considerarmos a quantidade de recursos alocados a todos os aspectos do trabalho em AIDS, inclusive tratamento, pesquisa científica e prevenção, nos países industrializados em com-

paração com o mundo em desenvolvimento. Entretanto, os países com a mais alta prevalência de HIV são aqueles com menos recursos para realizar atividades de avaliação, embora pudessem conseguir mais benefícios se pudessem orientar mais efetiva e eficientemente seus esforços de prevenção. Assim, uma pergunta importante ainda permanece: as abordagens de avaliação desenvolvidas no mundo industrializado seriam adaptáveis a diferentes situações culturais nos países em desenvolvimento? Obviamente, as normas e valores sociais, tanto entre grupos de risco como na população em geral, variam de lugar para lugar. Essas diferenças nas condições individuais, sociais, culturais e econômicas afetam não só o desenvolvimento de estratégias específicas de intervenção, mas também o desenho e implementação de atividades de avaliação de programa. Por exemplo, a avaliação de um programa de intervenção visando rapazes *gays* em São Francisco não usaria as mesmas abordagens que seriam usadas em Bangkok. Da mesma forma, a avaliação formativa de uma campanha de educação em AIDS entre adolescentes em Paris exigiria técnicas diferentes de coleta de dados de outra em Karachi. Os métodos de avaliação precisam ser adaptados ao determinado grupo-alvo, na situação sociocultural na qual o programa de prevenção está sendo realizado. Certas abordagens de avaliação, entretanto, são mais adaptáveis que outras. Por exemplo, a pesquisa etnográfica, as avaliações participativas e outras abordagens conduzidas por pares provavelmente seriam aplicáveis em qualquer situação cultural, embora algumas das abordagens convencionais possam ser muito mais restritivas. É essencial, portanto, examinar criticamente as abordagens de avaliação que foram desenvolvidas para o mundo industrializado e modificá-las para adequá-las às diferentes populações e aos diferentes contextos do mundo em desenvolvimento (44).

CONCLUSÃO

Nos primeiros anos da epidemia de AIDS, as agências de financiamento apoiaram programas comunitários de prevenção no mundo inteiro sem exigir provas de sua efetividade. Entretanto, agora que a epidemia entra em sua terceira década, isto mudou. Devido à crescente competição entre vários programas sociais, inclusive programas de tratamento da AIDS, pelos recursos disponíveis e a ênfase maior na responsabilidade fiscal, os financiadores estão exigindo cada vez mais que os programas demonstrem os resultados de suas atividades de prevenção para continuar o financiamento. A boa prática de avaliação, contudo, não deveria ser encarada apenas como uma exigência de agências de financiamento em certas épocas do ano. As atividades contínuas de monitoração e avaliação oferecem uma ferramenta para melhorar os programas, garantindo que as intervenções estão sendo feitas com eficiência e por isto beneficiam tanto a organização como a comunidade. Por estes motivos, é da maior importância que as organizações comunitárias aumentem sua capacidade de avaliar suas atividades de prevenção. Uma chave para o esforço de avaliação bem-sucedido é o envolvimento de pessoas que se importam com o programa de prevenção, inclusive equipe, voluntários e indivíduos da população-alvo. Proporcionando perspectivas diferentes, cada grupo pode trazer uma contribuição valiosa para melhorar o programa e, em última análise, limitar a disseminação da epidemia.

BIBLIOGRAFIA

1. Center for Aids Prevention Studies. *Glossary of Evaluation Terms*. Center for Aids Prevention Studies. San Francisco, CA; 1997.
2. Mertens TE, Crael, M. Evaluation of HIV/STD Prevention, Care and Support: An Update on

- WHO's Approaches. *AIDS Education and Prevention*; 1997; 9 (2): 133-145.
3. Rossi PH, Freeman HE. *Evaluation: A Systematic Approach* (5th ed.). Newbury Park, CA: Sage Publications; 1993.
 4. Veney JE, Gorbach P. *Definitions for Program Evaluation Terms*. Chapel Hill, NC: *Evaluation Project* (working paper series n° wp-tr-01); 1993.
 5. Aggleton P, Young A, Moody D, Kapila M, Pye M. *Does It Work? Perspectives on the Evaluation of HIV/AIDS Health Promotion*. London: Health Education Authority; 1992.
 6. Boerma T. *National AIDS Programs: A Guide to Monitoring and Evaluation*. UNAIDS. Geneva; 2000 (in press).
 7. National AIDS Fund. *Evaluating HIV/AIDS Prevention Programs in Community-Based Organizations*. National AIDS Fund. Washington DC; 1993.
 8. Framework for Program Evaluation. *Morbidity and Mortality Weekly Reports: Recommendations and Reports*; 1999, (48 RR-11): 1-35.
 9. Academy for Educational Development. *Critical Issues in HIV Prevention Evaluation*. Academy for Educational Development. Washington DC; 1998.
 10. Kellogg Foundation. *Evaluation Handbook*. W.K. Kellogg Foundation; 1997.
 11. Coyle SL, Boruch RF, Turner CF. *Evaluating AIDS Prevention Programs*. Washington DC: National Academy Press; 1991.
 12. Rhodes F, Malotte CK. The Role of Program Evaluation in Community-Based HIV Prevention Programs. In: Corby NH, Wolinski RJ, eds. *Community HIV Prevention: The Long Beach AIDS Community Demonstration Project*. Long Beach CA: The University Press; 1997.
 13. National Council of La Raza. *Understanding Evaluation Techniques: The Building Blocks of Evaluation*. National Council of La Raza. Washington DC; 1993.
 14. Mohr LB. *Impact Analysis for Program Evaluation* (2nded.). Newbury Park CA: Sage Publications; 1995.
 15. Moses S, Plummer FA, Ngugi EN, et al. Controlling HIV in Africa: Effectiveness and Cost of an Intervention in a High-frequency STD Transmitter Core Group. *AIDS*; 1991; 5: 407-411.
 16. Kahn JG, Washington AE, Showstack JA, et al. *Updated Estimates of the Impact and Cost of HIV Prevention in Injection Drug Users*. Report prepared for the Centers for Disease Control. Institute for Health Policy Studies, University of California, San Francisco, CA; 1992.
 17. Celantano DD, Nelson KE, Lyles CM, et al. Decreasing Incidence of HIV and Sexually Transmitted Diseases in Young Thai Men: Evidence for Success of the HIV/AIDS Control and Prevention Program. *AIDS*; 1998; 12: F29-36.
 18. Hearst N, Lacerda R, Gravato N, et al. Reducing AIDS Risk Among Port Workers in Santos, Brazil. *American Journal of Public Health*; 1999; 89(1): 76-78.
 19. Buckner BC, Tsui AO, Hermalin AI, McKaig C. *A Guide to Methods of Family Planning Program Evaluation, 1965-1990*. Chapel Hill, NC: Evaluation Project; 1995.
 20. Bertrand JT, Solis M. *Evaluando Proyectos de Prevencion de VIH/SIDA: Un manual con Enfoque en las Organizaciones No Gubernamentales*. Chapel Hill NC: Measure Evaluation Project; 1999 (in press).
 21. Fuente. *Manual para la Excelencia en la Investigación Mediante Grupos Focales*. Academy for Educational Development. Washington DC.
 22. Fisher AA, Laing JE, Stoeckel JE, Townsend JW. *Handbook for Family Planning Operations Research Design* (2nd ed.). New York, NY: Population Council; 1991.
 23. Bertrand JT, Lakssir A. *Qualitative Methods for Evaluating Family Planning Programs*. Chapel Hill NC: Evaluation Project; 1997 (draft).
 24. Patton MQ. *Qualitative Evaluation and Research Methods* (2nd ed.). Newbury, CA: Sage Publications; 1990.
 25. Miller HG, Turner CF, Moses LE. *AIDS, The Second Decade*. Washington DC: National Academy Press; 1990.

26. Carrier JM. Sexual Behavior and the Spread of AIDS in Mexico. *Medical Anthropology*; 1989; 10: 129-142.
27. Oostvogels R, Karamade A. Ethnography of Prostitution in Kamathipura and Khetwadi Sections of Bombay, India. *International Conference on AIDS* (abstract n° WS-D10-3); 1993.
28. French R, Power R, Mitchell S. An Evaluation of Peer-led STD/HIV Prevention Work in a Public Sex Environment. *AIDS Care*; 2000; 12(2): 225-34.
29. Quintana A, Vasquez E. HIV/AIDS Risk Perception and Adolescents in Lima, Peru. *International Conference on AIDS* (abstract n° 152/14210); 1998.
30. Shumba PS, Teri E, Kamunya R, Duncombe C. *The Urban Reproductive Health Initiative in Kenya: A Process Evaluation*. Watertown, MA: Pathfinder International; 1999.
31. Laver SM, Van Den Borne B, Kok G, Woelk G. Was the Intervention Implemented as Intended?: A Process Evaluation of an AIDS Prevention Intervention in Rural Zimbabwe. *International Quarterly of Community Health Education*; 1997; 16(1): 25-46.
32. Jemmott JB, Jemmott LS, Fong GT. Reductions in HIV Risk-Associated Sexual Behaviors among Black Male Adolescents: Effects of an AIDS Prevention Intervention. *American Journal of Public Health*; 1992; 82(3): 372-377.
33. Cohen DA, MacKinnon DP, Dent C, et al. Group Counseling at STD Clinics to Promote Use of Condoms. *Public Health Reports*; 1992; 107(6): 727-731.
34. Bhave G, Lindan CP, Hudes ES, et al. Impact of an Intervention on HIV, Sexually Transmitted Diseases, and Condom Use among Sex Workers in Bombay, India. *AIDS*; 1995; 9(suppl 1): S21-30.
35. Lauby JL, Smith PJ, Stark M, et al. A Community-Level HIV Prevention Intervention for Inner-City Women: Results of the Women and Infants Demonstration Projects. *American Journal of Public Health*; 2000; 90(2): 216-222.
36. Kelly JA, St. Lawrence JS, Hood HV, Brasfield TL. Behavioral Intervention to Reduce AIDS Risk Activities. *Journal of Consulting and Clinical Psychology*; 1989; 57(1): 60-67.
37. Kelly JA, St. Lawrence JS, Betts R, et al. A Skills-Training Group Intervention Model to Assist Persons in Reducing Risk Behaviors for HIV Infections. *AIDS Education and Prevention*; 1990; 2(1): 24-35.
38. Visrutaratna S, Lindan CP, Sirhorachai A, Mandel JS. 'Superstar' and 'Model Brothel' : Developing and Evaluating a Condom Promotion Program for Sex Establishments in Chiang Mai, Thailand. *AIDS*; 1995; 9(suppl 1): S69-75.
39. Coates TJ, Chesney M, Folkman S, et al. Designing Behavioral and Social Science to Impact Practice and Policy in HIV Prevention and Care. *International Journal of STD and AIDS*; 1995; 7(suppl 2): 2-12.
40. Community Level Prevention of HIV Infection among High-risk Populations: The AIDS Community Demonstration Projects. *Morbidity and Mortality Weekly Reports: Recommendations and Reports*; 1996; 10(RR-6), 1-24.
41. Higgins DL, O'Reilly K, Tashima N. Using Formative Research to Lay the Foundation for Community Level HIV Prevention Efforts: An Example from the AIDS Demonstration Projects. *Public Health Reports*; 1996; 111(suppl 1): 28-35.
42. King R, Kabatesi D, Homsy J. Using Participatory Evaluation to Focus AIDS Interventions with Traditional Healers in Uganda. *International Conference on AIDS* (abstract n° 44309); 1998.
43. UNAIDS. *Report on the Global HIV/AIDS Epidemic*. UNAIDS. Geneva; June 2000.
44. UNAIDS. *Sexual Behavioral Change for HIV: Where Have Theories Taken Us?* UNAIDS. Geneva; 1999.

