

COLEÇÃO ABIA

FUNDAMENTOS DE AVALIAÇÃO, Nº 1

AVALIAÇÃO EM SAÚDE: QUESTÕES PARA OS PROGRAMAS DE DST/AIDS NO BRASIL

MARIA INES BAPTISTELLA NEMES

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

Luis Felipe Rios

PROJETO GRÁFICO

Juan Carlos Raxach

Wilma Ferraz

EDITORACÃO ELETRÔNICA

E PRODUÇÃO GRÁFICA

Wilma Ferraz

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 citadas a fonte e a autora.*

SUMÁRIO

APRESENTAÇÃO	5
POR QUE AVALIAR?	9
PARA QUÊ AVALIAR?	11
O QUÊ AVALIAR?	13
É POSSÍVEL AVALIAR?	16
COMO AVALIAR?	21
DESAFIOS PARA A AVALIAÇÃO DOS PROGRAMAS EM HIV/AIDS:	
ALGUMAS SUGESTÕES À GUIA DE CONCLUSÃO	23
BIBLIOGRAFIA CITADA	26
BIBLIOGRAFIA CONSULTADA	27

APRESENTAÇÃO

A série “Fundamentos de Avaliação” objetiva fomentar o debate acerca da avaliação em HIV/AIDS que se faz urgente no momento em que a epidemia mundial completa seu 20º ano e o Brasil se depara com a fase conclusiva do 2º empréstimo do Banco Mundial para o desenvolvimento do Programa de AIDS governamental. Diante desse quadro, a preocupação atual, tanto para ONGs, sociedade civil e Ministério da Saúde, é a continuidade e aprofundamento das ações iniciadas nos últimos anos, o que requer avaliações sistematizadas dos trabalhos implementados até o momento.

Em 2000, a Associação Brasileira Interdisciplinar de AIDS (ABIA) realizou, com apoio da Fundação Ford e do Ministério da Saúde, o seminário “Avaliação em DST/AIDS”. O encontro cumpriu o objetivo de discutir os processos de avaliação em saúde no âmbito de programas de DST/AIDS no Brasil e de ações implementadas pela sociedade civil. A ocasião suscitou falas oriundas de diferentes atores que vêm realizando avaliações no seu cotidiano de trabalho.

O texto *Avaliação em Saúde: Questões para os Programas de DST/AIDS no Brasil*, que apresentamos a seguir, fornece valiosas pistas para os processos de avaliação em saúde de um modo geral, tendo a função de apresentar aos profissionais e instituições envolvidos no trabalho com HIV/AIDS metodologias disponíveis atualmente e fundamentar os textos e discussões que virão a seguir.

Reconhecemos que tais textos podem trazer informações científicas por vezes inconciliáveis com a prática de nossas organizações. Mas devemos investir num debate multi-setorial, envolvendo ONGs, OGs e universidades a fim de repensar nossas práticas, aprimorando nossas avaliações que já vêm sendo feitas. O grande desafio será reconhecer a pressão cada vez maior (pelo menos em termos históricos) sobre a necessidade de avaliação (vinda dos financiadores, claramente, mas também dentro do nosso próprio trabalho, para garantir que estamos conseguindo alcançar nossas próprias metas), de certa forma concordando, mas insistindo em fazer avaliações nas quais acreditamos em vez de passivamente aceitar o que é imposto de fora.

Richard Parker
Diretor-Presidente da ABIA

Veriano Terto Jr.
Coordenador Geral da ABIA



AVALIAÇÃO EM SAÚDE: QUESTÕES PARA OS PROGRAMAS DE DST/AIDS NO BRASIL¹

MARIA INES BAPTISTELLA NEMES²

Embora questionamentos a respeito dos benefícios das intervenções em saúde sejam bastante antigos, é possível tomar a década de 50 como marco inicial da consolidação da questão da avaliação em saúde enquanto objeto de investigação científica, dentro do movimento que estabeleceu as intervenções em saúde como “campo” científico. Este movimento teve como principais provocadores a grande expansão dos serviços de saúde, o rápido desenvolvimento tecnológico da prática médica, os enormes custos das novas técnicas e a crescente responsabilidade dos governos no financiamento da atenção à saúde. O novo campo científico, de limites pouco nítidos, teve seus objetos criados a partir da necessidade de se obter maior conhecimento acerca das várias formas de organização da

atenção bem como acerca da eficácia, efetividade, eficiência e idoneidade das intervenções em saúde.

Já nos anos 60 e 70, as investigações se ampliaram e sofisticaram e progressivamente foram delimitando “vertentes” e “subcampos”. Hoje, é possível delimitá-las em três grandes grupos mais reconhecíveis na literatura internacional: a avaliação tecnológica (focada sobretudo na segurança, efetividade e custo de tecnologias de produto ou de processo), a avaliação de programas (focada em conjuntos articulados de atividades voltados para uma população alvo), e a avaliação de qualidade (focada em serviços assistenciais de saúde, especialmente médicos).³

Todas essas vertentes tiveram origem e desenvolvimento nos países centrais (principalmente nos EUA). Até hoje, a sua produção, tanto no âmbito da academia quanto no das

¹ A versão preliminar deste texto foi discutida no seminário *Avaliação em DST/AIDS*, realizado em setembro/2000 pela ABIA. A versão atual procurou incorporar as sugestões dos participantes, a quem agradecemos, especialmente Betina Durovni, Otávio Cruz Neto e Ruben Mattos.

² Médica Sanitarista, Professora do Departamento de Medicina Preventiva da Faculdade de Medicina da USP. Tel/fax: (11) 3062 68 22; e-mail: mibnemes@usp.br.

³ Estas delimitações não são tão precisas. Há muitas interfaces entre elas. Há ainda muitos estudos que não caberiam exatamente nestas denominações. O critério de classificação, contudo, dá conta das maiores tendências tanto em termos científicos quanto dos macro-contextos históricos específicos de origem e desenvolvimento de cada uma das vertentes (Novaes, 1999).

instituições públicas ou privadas, realiza-se predominantemente nestes países.

Embora a importância de estudá-las seja evidente para os que desejam trabalhar em avaliação, não me pareceu que uma análise deste tipo fosse aqui tão importante. Em parte, porque seria muito difícil realizá-la nos limites deste texto e, é claro, pelos meus próprios limites. A razão mais importante, entretanto, é o fato destas vertentes, enquanto macro tendências científicas oriundas de específicos contextos históricos, fazerem-se representar muito pouco no Brasil.

No nosso país, muito embora se reconheçam os mesmos determinantes históricos mais gerais (expansão de tecnologias, custos etc.) que determinaram a emergência do campo da avaliação nos países centrais, este reconhecimento é muito mais recente e ocorre em contextos históricos muito diferentes. Entre nós a problemática da avaliação ganhou maior impacto e abrangência apenas nos últimos anos. Há, de fato, uma recente assunção da importância da avaliação no discurso dos formuladores e gestores de práticas de saúde. Temos notado isto na universidade, na demanda crescente por cursos de pós-graduação em avaliação, e nos serviços públicos, que nos chamam com maior frequência para treinamentos ou palestras que visam “ensinar metodologias” para equipes de técnicos interessados em avaliar as atividades dos programas e serviços de saúde.

Apesar disso, é fato a ainda pequena incorporação de sistemáticas de avaliação nos serviços públicos e privados brasileiros (Novaes, 1999).

A insipiência da prática avaliativa entre nós e a importante diferença de contextos históricos que levou, por exemplo, à inexistência no Brasil de um campo profissional específico para “avaliadores”⁴ fazem com que as vertentes de

⁴ Como nos EUA e Canadá, onde há associações profissionais de avaliadores.

avaliação internacionais representem para o Brasil “*idéias fora do lugar*” que, para serem incorporadas, necessitariam de re-interpretação e re-significação (Novaes, 1999:28). Isto tem me levado a fundar minhas palestras menos na discussão das vertentes avaliativas “clássicas” e muito mais no que costumo chamar de “as grandes perguntas da avaliação”. Afirmo sempre que para que um processo avaliativo tenha sucesso, é preciso que tenhamos respostas razoáveis para as seguintes perguntas:

- Por que avaliar?
- Para quem avaliar?
- O que avaliar?
- É possível avaliar?
- Como avaliar?

Estas perguntas podem ser vistas como dizendo respeito a questões mais gerais que se colocam para a prática científica, relacionando-se com as polaridades ciência e tecnologia, produção e utilização de conhecimento e sujeito e objeto da investigação. Vou tomá-las, entretanto, apenas na dimensão que elas suscitam problemáticas práticas para a avaliação em saúde nas específicas condições históricas brasileiras.

É importante ressaltar que trataremos exclusivamente do campo da avaliação em saúde assim como ele se constituiu em outros países e vem se constituindo no Brasil, isto é, focado, sobretudo, nas questões “mais técnicas” dos serviços e programas de saúde. Não serão enfatizadas, portanto, dimensões da saúde “mais políticas” como, por exemplo, as do planejamento ou das políticas institucionais.⁵

⁵ Esta afirmação não deve ser mal entendida. Ao dizer que trataremos das dimensões “mais técnicas”, não estamos negando seu caráter político. Apenas ressaltamos que o campo e os métodos da avaliação em saúde têm historicamente concentrado-se nestas dimensões. Isto não quer dizer que atinjam a “ideal neutralidade” tal como o discurso científico “padrão” apregoa. A avaliação em saúde lida com o *trabalho* em saúde. *E o trabalho em saúde é político ao ser técnico*. A avaliação, ao refletir

POR QUE AVALIAR?

Avaliar é julgar, atribuir valor. Por que esta aparente obviedade deve ser ressaltada?

O campo da saúde – especialmente o campo da Saúde Coletiva – é um campo extremamente comprometido com o pragmatismo, com a utilidade. Por isso, dele sempre se diz que é um campo de “conhecimentos e práticas”. Suas investigações são particularmente e muito freqüentemente dedicadas à análise de práticas concretas. Nestas análises, o valor da efetividade (definida, por exemplo, pela alteração da morbimortalidade, da melhoria da qualidade de vida etc.) está sempre presente, ainda que como idéia reguladora. São análises, portanto, muito comprometidas com a valoração. Neste sentido, contém freqüentemente, de modo subjacente ou explícito, algum nível de julgamento, podendo ser tomadas, neste sentido, como “avaliações”.

Entretanto, as investigações que têm delimitado mundialmente o que pode ser chamado de “campo da avaliação em saúde” possuem enfoques específicos, extremamente pragmáticos, desenvolvidos em contextos históricos particulares.

A pequena experiência avaliativa no Brasil, a inexistência de avaliadores profissionais entre nós, e a óbvia dificuldade (e impropriedade) de “importar” contextos históricos são responsáveis, em parte, pela pouca nitidez do campo da avaliação e pelo baixo reconhecimento da especificidade da pesquisa avaliativa. Assim, apesar das propostas e discursos, a pequena experiência avaliativa e a baixa sistematização permanecem.

É extremamente freqüente entre nós tomar algumas análises ou mesmo estudos descritivos enquanto “avaliações”. Refiro-me a relatórios de atividades ou de alcance de metas, descrições de programas, relatórios “de qualidade” etc. Tal

sobre este trabalho, também é assim. Para enfatizar este ponto, anotamos os termos “mais técnico” ou “mais político” entre aspas.

como afirmei, para as investigações do nosso campo em geral, estes estudos estão baseados em idéias reguladoras de efetividade, eficiência ou otimização. Não alcançam, entretanto, *o julgamento* que é o conceito operativo central da avaliação.

Usando livremente um conceito de Aguillar & Ander-Egg (1994), vou chamar estes estudos de “Quase-Avaliações”. “Quase”, porque apesar de conterem idéias de valoração, não chegam a emitir *juízos de valor*. Suas evidências – mesmo quando colhidas com objetividade e confiabilidade – não foram tomadas como construtoras de parâmetros ou indicadores avaliativos. Esta questão é semelhante ao que ocorre com a pesquisa científica em geral. Vamos “ao campo” para afirmar, infirmar e/ou aprimorar hipóteses derivadas de teorias: o real só nos “diz” aquilo que perguntamos.

Em avaliação é o mesmo. O “ciclo avaliativo” necessita da elaboração de parâmetros que, fundados em teorias acerca da intervenção, possam produzir *indicadores* válidos para o julgamento. Por isso, é possível dizer que se o passo mais difícil – e verdadeiramente criativo – da pesquisa científica está na formulação de hipóteses, o passo mais difícil da avaliação é o da formulação de indicadores.

Decorre imediatamente disto a afirmação – nunca por demais repetida – de que *o método* só pode ser *o mais adequado ao objeto construído* pela investigação. Se a “apologia do método” ainda é um pecado mortal na pesquisa científica, mas muito freqüente para nosso campo em geral, imagine-se na avaliação, cujo compromisso pragmático é muito mais imediato!

Assim, por exemplo, a aproximação avaliativa com o cuidado médico individual, sistematizada por Donabedian nos anos 60, que recomenda a elaboração de indicadores de estrutura, processo e resultado, tem sido freqüentemente tomada como uma “fórmula

metodológica” quase mágica e que sempre pode ser utilizada para avaliar “tudo”.

Se a abordagem avaliativa for “Donabediana”, então, é preciso ressaltar que uma “quase-avaliação” só começa a “virar” uma “avaliação mesmo”, quando conseguimos estabelecer articulações de caráter causal aceitável entre indicadores de estrutura, processo e resultado. Tenho sido tão enfática neste ponto que sempre respondo à freqüente pergunta “isto é estrutura ou processo?” com a afirmação “não é preciso enquadrar na fórmula, o mais importante é estabelecer relações e associações pertinentes”.

Preocupo-me, entretanto, com um possível exagero deste raciocínio. A assunção da importância da avaliação entre nós tem inegáveis potenciais éticos e não se deve solapar o ânimo dos que estão realmente interessados em avaliar. Uma quase-avaliação que apenas sistematize situações, por exemplo, pode contribuir muito para a transparência e inteligibilidade de processos, condição necessária para o início de qualquer processo avaliativo.

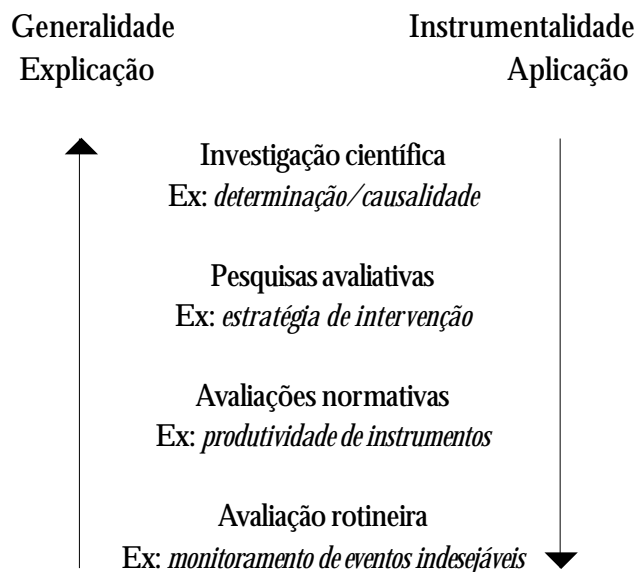
Além disso, desqualificar as “quase-avaliações” é impertinente do ponto de vista científico. Uma “quase-avaliação” bem conduzida pode produzir suficientes evidências práticas para a localização de eventos sentinela (por exemplo, como de um paciente sem os exames rotineiros prescritos pelo programas) ou ainda para identificar processos imediatamente monitoráveis (como a prescrição de alguma medicação incorreta) e mesmo, em alguns casos, evidências que auxiliem a montar futuras avaliações (como a localização de diferenças entre processos de trabalho entre agências de um mesmo programa).

Uma avaliação – uma “boa” avaliação – deve julgar. Uma “boa” e “completa” avaliação julga e explica. Mas julgar e, sobretudo, explicar, é atributo da investigação científica.

Alguns autores procuram dar conta desta questão dividindo a atividade de avaliação em dois campos: a avaliação normativa, destinada a julgar o funcionamento de uma dada intervenção de acordo com normas pré-estabelecidas, e a pesquisa avaliativa, destinada a julgar e explicar todas as relações entre os componentes de uma intervenção de acordo com teorias sobre o programa (Conrandiopoulos, 1997). Uma outra denominação chama a primeira de avaliação normativa e a segunda de avaliação causativa (Chen, 1990).

Não me parece muito útil este nominalismo, exceto quando puder auxiliar o raciocínio que, de resto, diz respeito a todas as investigações científicas e não apenas às avaliações. Um maior grau de abstração, atributo das investigações de “grandes” aportes teórico-metodológicos, gera maior explicação e menor aplicação. No pólo oposto, pesquisas mais aplicadas têm menor poder explicativo sendo, porém, mais instrumentais. O esquema a seguir mostra o raciocínio para um dado problema de saúde exemplificando os objetos possíveis para cada um dos níveis.

Para problema de saúde X



Quando defrontada com esta questão com alunos, uso este esquema muito mais para ressaltar:

- qualquer um destes momentos exige um certo nível de formulação de hipóteses que sustentará a formulação e medição de indicadores,
- as avaliações “rotineiras” ou mais normativas podem fornecer importantes elementos para a pesquisa e, finalmente,
- que precisamos estar sempre atentos para o grau de explicação que conseguimos, sobretudo para não nos “conformamos” com ele.

PARA QUÊ AVALIAR?

A avaliação tem um enorme compromisso pragmático. Avalia-se para mudar, para aprimorar. Entretanto, é justamente este o ponto que recebe mais críticas. O problema do baixo impacto das avaliações aparece freqüentemente na literatura e no discurso de agentes do trabalho e, às vezes, de associações de usuários.

É possível delimitar alguns dos problemas que freqüentemente determinam este pequeno impacto:

- pequena margem e/ou baixa relevância do foco de avaliação: ocorre comumente quando os “avaliadores” escolhem indicadores mais fáceis de obter mas que dizem pouco acerca do valor da intervenção para os trabalhadores e a população alvo (como informações advindas de sistemas rotineiros que geram dados de rendimento e produtividade, por exemplo);
- excessiva margem/abrangência do foco de avaliação: pólo oposto ao anterior, este se caracteriza pelo “querer avaliar tudo”. Ocorre quando se somam grandes quantidades de dados sem que haja uma relação consistente entre eles (como, por exemplo, informações de “estrutura” somadas as “de processo”: número de recursos humanos e materiais, número de atendimentos etc.);

- centramento em resultados: ocorre quando o foco da avaliação diz respeito unicamente aos resultados esperados da intervenção (como, por exemplo, taxas de mortalidade). Nestes casos, o processo fica indefinido. Esta é a típica “black box evaluation”, somente baseada em “inputs” e “outputs” (Chen, 1990). Falaremos muito dela adiante, mas, por enquanto vale ressaltar que o centramento em resultados pode levar a vieses importantes como a omissão de efeitos não esperados (e por vezes, perversos) do programa, da tecnologia ou serviço de saúde (como seria o caso de um programa de aconselhamento em HIV/AIDS aumentar certos comportamentos de risco ou a desejável descentralização de um programa provocar queda na qualidade técnica do cuidado médico);

- centramento em processos específicos: pólo oposto ao anterior, este se caracteriza por uma busca detalhada de relações processuais que acabam por ser somente aplicáveis a situações institucionais e tecnológicas pouco generalizáveis (como por exemplo, o centramento do foco nos sistemas de referência e contra-referência de serviços de saúde).

Os “teóricos” da avaliação costumam recomendar uma série de requisitos metodológicos para o enfrentamento desses problemas. Para eles, uma avaliação deve ter validade interna, objetividade, confiabilidade e validade externa ou capacidade de generalização.⁶ Estes são, entretanto, requisitos metodológicos de qualquer investigação científica.

Há outro requisito, mais específico, e muito valorizado por algumas das correntes de avaliação de programas, que é a “responsividade”

⁶ Estes requisitos, embora diferentemente valorizados, são defendidos nas avaliações de programa predominantemente quantitativas (*internal validity, external validity, reliability, objectivity*), nas predominantemente qualitativas (*true value, applicability, consistency, neutrality*) e nas avaliações da qualidade (*causal validity, sensitivity, specificity, relevancy, timeliness, inclusivity*) (Chen, 1990).

(*responsiveness*).⁷ Este me parece o mais interessante porque suscita uma questão que vai diretamente ao ponto central da relação entre produção e utilização das avaliações.

Uma avaliação “responsiva” é aquela que interessa, que é importante, para todos os grupos envolvidos, para os que eu costumo chamar de “sócios” da avaliação.

Se todos os requisitos metodológicos acima apontados implicam alguns conflitos, a questão da “responsividade” *versus* a capacidade de generalização é para a avaliação e sua utilização a polaridade mais importante.

Parece evidente que a maior especificidade do foco de avaliação favoreça a responsividade e desfavoreça a capacidade de generalização. Tenho visto muitas vezes este conflito expressado entre gestores/formuladores e equipes locais. Os primeiros tendem a estabelecer “grandes” indicadores generalizáveis para muitos sítios.⁸

O staff local acusa estes indicadores de pouca sensibilidade para as questões locais “realmente relevantes”. Este dilema é também responsável por assunções que colocam em pólos opostos a avaliação interna e a avaliação externa (polêmica muitas vezes traduzida, nos nossos serviços de saúde, pelo “desprezo” das equipes locais aos avaliadores “externos” representados pelas informações e boletins que, preenchidas

localmente, alimentarão os “pacotes” de avaliação dos níveis gerenciais e centrais).

Evidentemente, não haverá solução propriamente metodológica para esta questão. Teoricamente, as avaliações podem incidir em qualquer um dos pontos da relação problemática → objeto de trabalho → agentes e instrumentos → finalidades de uma intervenção em saúde. Será o objeto tomado para a avaliação, o objetivo e foco da avaliação e sua finalidade (para quê e para quem?) que determinará sua maior responsividade ou maior capacidade de generalização. Não é o caso de se aderir a um ou outro pólo (como já ocorreu muito na área de avaliação de programas sociais no EUA). Somente parece possível dizer que idealmente dever-se-ia equilibrar avaliações mais generalizáveis (tendencialmente mais “externas”) com avaliações mais responsivas (tendencialmente mais “internas”).

Finalmente, é importante ressaltar que a “baixa utilização” dos produtos da avaliação nunca será resolvida apenas no âmbito metodológico. Ocorre aqui uma outra ordem de determinação.

A avaliação em saúde tem sido predominantemente um processo que almeja julgar e explicar práticas sociais sob uma só racionalidade: a científica. Seu poder será sempre limitado por isto. Ademais, as avaliações em saúde estão fundadas principalmente na epidemiologia e secundariamente em disciplinas administrativas. Não é possível esperar, portanto, que possam dar conta do terreno de múltiplas determinações e de conflitos éticos-políticos no qual atuam os

⁷ Uma discussão muito interessante sobre os valores da avaliação pode ser encontrada em Ayres, 1996. Além da profundidade e concisão, este artigo tem a vantagem de utilizar ações preventivas em AIDS para desenvolver o raciocínio. Neste texto, elegemos para comentar apenas o valor da responsividade, que nos parece o que mais tem tensionado as práticas avaliativas entre nós.

⁸ Estes tipos de indicadores muito genéricos aparecem frequentemente em modelos para avaliação construídos por agências internacionais para aplicação em países pobres. “(...) *diversas organizações internacionais (...) estimulam a utilização em projetos com propostas metodológicas estruturadas em ‘pacotes’ (...), sendo sua presença dominante nos países com pouca densidade científica local. A justificativa apresentada para essas propostas fechadas e geralmente simplificadas é a grande utilidade de processos*

avaliativos para a adequada identificação de prioridades em saúde, a necessidade de um aprendizado rápido e de adaptação a contextos de recursos (menos qualificados e/ou mais escassos)” (Novaes, 1999:28-9).

Apesar da simplificação e da inespecificidade estes “pacotes” podem ser úteis em determinados contextos, como por exemplo, em inícios de processos avaliativos que envolvam grande extensão de sítios pesquisados.

formuladores e gestores das políticas de saúde, por exemplo. De outro lado, na “outra ponta” das intervenções em saúde, está o trabalhador direto. Também para este a avaliação somente é capaz de apreender algumas dimensões do trabalho. Não se pode esperar dos quadros normativos que sustentam a avaliação em saúde respostas unívocas para dimensões tão importantes e presentes no trabalho como as interações humanas.

Mas se a avaliação não “pode tudo” ela certamente “pode muito”. Considero que a recente assunção da importância da avaliação em saúde entre nós deve ser saudada pela sua inegável potencialidade ética. Isto é tão mais importante quando lembramos que as práticas de saúde em geral – e a instituição médica em particular – têm historicamente se mantido alheias a formas mais efetivas de controle social.

Parece-me ainda que esta potencialidade será tão mais desenvolvida quanto mais pensarmos a avaliação como decodificadora da realidade e, neste sentido, como produtora de insumos para o debate político. Neste caso, sua maior densidade científica interessa menos como produtora imediata de verdades e mais como modo de maximizar a objetividade do debate.

O QUÊ AVALIAR?

A resposta mais natural a essa pergunta é: queremos avaliar se a intervenção funciona ou não funciona, ou seja, se produz os resultados e os impactos esperados.⁹ Esse pode ser conside-

⁹ Na literatura em saúde o termo “resultado” refere-se a efeitos obtidos em indivíduos (como a melhoria da qualidade de vida em uma avaliação do cuidado médico) e o termo “impacto” diz respeito a modificações em indicadores de base populacional atribuíveis a uma determinada intervenção (como a diminuição da taxa de infecção em uma avaliação de um programa de controle de tuberculose). Em avaliação, habitualmente, o “impacto” refere-se à diferença observada nos efeitos (nos indivíduos ou na população) obtidos por um programa *em relação* a outro programa (ou não programa). Já os resultados (“outcome”) referem-se aos efeitos esperados de um programa e medem sua

rado como o objetivo mais “nobre” e mais ambicioso de uma avaliação – julgar que efeitos foram produzidos e em que medida eles foram produzidos. Entretanto, na área social em geral, mas especialmente na área da saúde, nem sempre é possível fazer esse tipo de avaliação.

Um primeiro problema é *definir o que são “os efeitos”*, ou seja, ter clareza *do que se espera* como consequência da intervenção tomada sob avaliação. Seria esperado que programas ou intervenções bem planejadas tivessem isso *a priori* muito bem definido – a partir de um determinado conjunto de ações se espera atingir tais ou quais resultados. No “melhor dos mundos” um programa deveria já ter definido, ainda no seu planejamento, mecanismos para medir resultados de uma forma válida e confiável. Entretanto, não é bem isso que ocorre na maior parte das intervenções em saúde. Somente quando se focaliza a avaliação em ações muito específicas e muito bem recortadas a delimitação é mais fácil. Tal seria o caso, por exemplo, da avaliação de um tipo de procedimento cirúrgico. Nessas condições, tanto o tipo de resultado esperado como a forma de medi-lo, assim como a relação de determinação entre a intervenção e as mudanças encontradas, podem estar previamente bem definidos, dada a possibilidade de controle, ainda que sempre relativa, do conjunto das condições intervenientes.

A definição da situação também é mais fácil para programas que têm finalidades estratégicas muito bem definidas. Por exemplo, se avaliarmos o impacto na mortalidade da clientela de pacientes com AIDS acompanhados em dois serviços diferentes, utilizando a mortalidade como medida do resultado da assistência recebida. Nesse caso, a diminuição da letalidade da doença é um objetivo estratégico bem definido e muito claro para um serviço assistencial.

efetividade. Nesta parte do texto usarei o termo “efeito” para designar genericamente o alcance das finalidades propostas para uma determinada intervenção.

Entretanto, mesmo quando temos resultados bem definidos, há a questão da *imputabilidade*, ou seja, como atribuir esses resultados àqueles cuidados? O que se quer avaliar, afinal, é o cuidado e, no entanto, o resultado depende de uma série de outras coisas que não são atribuíveis ao cuidado. Quais seriam, por exemplo, as condições clínicas de admissão desses pacientes nos serviços estudados? Em primeiro lugar, para qualquer intervenção sobre uma doença, as condições de “entrada” do doente são um dos principais determinantes para a capacidade de resposta às intervenções. É evidente que este efeito deve ser controlado para que se possa avaliar o papel causal do cuidado no resultado. Mas além desta condição, os resultados estão relacionados a um conjunto de outros determinantes, como a adesão ao tratamento, as condições sociais, o acesso a cuidados continuados e complementares e muitos outros.

Ademais, é preciso lembrar que nesse exemplo o resultado “mortalidade” tem uma relação lógica bastante plausível com o tratamento. Esta é uma importante condição para a definição de resultados esperados: a associação plausível e confiável com a intervenção. Muitas vezes esta associação não está bem determinada, como acontece em programas com múltiplos e/ou inespecíficos objetivos, como ocorre justamente com o Programa de Controle da AIDS cujos objetivos passam também pela qualidade de vida, autonomia, *empoderamento*, e isto só para a área de assistência aos infectados (na “área da prevenção” as coisas se complicam bem mais...).

Este é um grande problema para as avaliações em saúde: mesmo quando temos resultados bem definidos e passíveis de aferição confiável, como atribuí-los à intervenção?¹⁰

¹⁰ Os vários métodos estatísticos de análise multivariada (controle de variáveis de confusão, análise de componentes etc.) podem auxiliar muito o esclarecimento da “imputabilidade”. Evidentemente, porém, tal como nos métodos qualitativos, as

Embora tenha sido desenvolvida para o cuidado individual em geral, médico ou realizado por um outro profissional de saúde, a famosa tríade de Donabedian – estrutura, processo e resultado – procura dar conta justamente de responder a esse tipo de questão. Essa é a grande potencialidade explicativa dessa abordagem – procurar identificar quais as variáveis de estrutura que *provavelmente* permitem um bom processo e quais as variáveis de processo que *provavelmente* levam a um bom resultado. Identificados os indicadores de estrutura e processo, atribuir o resultado à intervenção torna-se muito mais plausível.

Até aqui falamos da dificuldade na definição e imputabilidade de resultados de programas *implantados*. Muitas vezes, porém, temos dúvidas acerca da própria implantação. Algumas vezes a alocação de recursos para um programa é tão pequena que inviabiliza qualquer estrutura mínima para o funcionamento. Nestes casos extremos, não é possível avaliar nada. Todavia, a situação mais freqüente é de “programas” com nível tão baixo de definição tecnológica que não se sabe o quê definiria uma adequada implantação. Há, entre nós, muitos exemplos extremos desta indefinição: “programas” cuja normatização limita-se a recomendações para tratamentos de doenças, por exemplo, ou ainda, à simples declarações de intencionalidades.¹¹

explicações dependem de uma “boa” teoria, a mesma que também guia a coleta de dados.

¹¹ Na definição de Champagne et al. (1986:1), “um programa de saúde pode ser definido como um conjunto de recursos que proporcionam serviços para uma população-alvo de um modo organizado em relação ao tempo e ao espaço, com a finalidade de conseguir objetivos estabelecidos relacionados com problemas de saúde específicos”. É importante precisar mais: “Um programa de saúde pode incluir, por exemplo, saneamento e controle da poluição, vacinação, “check-ups” periódicos, assistência médica à demanda, etc, etc. Do ponto de vista das atividades, o programa adquire variadas táticas, conservando, porém, seu núcleo de finalidade original que se dá sob e sobre coletivos, através de mediações menores ou maiores com o coletivo” (Nemes, 1995:6).

Nem sempre, felizmente, a situação é tão extrema. É possível afirmar, contudo, que grande parte dos nossos programas tem baixo grau de definição tecnológica acabando por assumir na prática formas muito diferentes de implantação.

Este também é o caso quando a pergunta da avaliação é mais complexa e diz respeito a níveis mais “macro” de intervenção, como por exemplo seria o caso ao avaliarmos a recente diferença observada entre a queda da mortalidade, após a introdução dos antiretrovirais, entre a capital e o interior de São Paulo. Este é tipicamente o caso em que se deve realizar uma avaliação da implantação. No caso das diferenças observadas na variação das taxas de mortalidade, sempre caberá perguntar o quanto o programa de assistência especializada a pacientes com HIV/AIDS (e o tratamento com antiretrovirais) de fato se implantou no interior, ou, o quanto isso se deu com extensão suficiente.

Avaliar a implantação tenta evitar um importante erro em avaliação: medir os efeitos de uma intervenção cujo grau de operacionalização não é suficiente ou não é minimamente satisfatório. Quando é que se deve realizar uma análise de implantação? Em primeiro lugar, para saber se a implantação de fato se efetivou dentro dos parâmetros previstos pelo programa, para que, a seguir, se possa julgar seu funcionamento. Por exemplo, para se avaliar as diferentes taxas de mortalidade entre capital e interior tem que se saber antes se a assistência está se dando em ambos os lugares dentro de padrões mínimos de aceitação, ou seja, se podemos considerar que o grau de implantação alcançado é satisfatório, tanto na capital como no interior, para que seus resultados possam ser comparados. Esse seria um tipo de análise de implantação prévio à avaliação dos resultados e que estaria centrado na verificação do grau em que a intervenção proposta de fato vem ocorrendo.

Um outro tipo de análise de implantação,

mais centrado no processo através do qual as ações se desenvolvem e com maior poder explicativo, é o que realiza simultaneamente uma avaliação de resultados individuais e/ou de impactos e a análise da influência da variação da implantação nos efeitos observados e a análise do contexto. Por exemplo, em programas pouco específicos, em que não há uma delimitação muito clara de como deve ser a implantação, como ocorre com a área de prevenção do Programa de AIDS. Esses programas possuem uma grande multiplicidade de estratégias e atividades disponíveis, fazendo com que os tipos de ações que são implantadas e o modo de operá-las sejam muito variáveis.

Por outro lado, há outras situações em que mesmo que se estabeleça um padrão tecnológico bem definido para as intervenções, as variáveis de contexto acabam por alterar os resultados alcançados, como seria o caso ao compararmos programas em que as ações se dão de modo centralizado com outros em que se dão de forma descentralizada. Ou ainda, a situação que ocorre, por exemplo, com a distribuição de preservativos no Estado de São Paulo – onde os processos de implantação em cada serviço são muito variados, o que se diversifica ainda mais se considerarmos as ações das ONGs nessa distribuição.

Quando não seria necessário realizar uma análise de implantação? Primeiro, quando se tem um grau aceitável de conhecimento do que se quer implantar, o Programa possui definições tecnológicas claras e específicas, existe uma experiência prévia em sua utilização, se conhece os efeitos esperados e estes variam muito pouco em relação a variáveis contextuais. Este é tipicamente o caso de programas de vacinação. Mesmo programas mais complexos, como o Programa de Controle da Hanseníase no Brasil, por exemplo, podem atingir estágios mais “maduros” permitindo maior centramento da avaliação em indicadores de impacto propriamente ditos.

Até aqui nós falamos de avaliações de programas cujos objetivos estão bem definidos, e cujo cumprimento de finalidades pode ser tomado por efeitos bastante específicos. Não é o que acontece com a maioria dos programas – eles são extremamente complexos, envolvem serviços diferentes, esperam resultados múltiplos e, muitas vezes, resultados pouco específicos. Isso ocorre sobretudo nos programas mais novos, dedicados a questões emergentes. Como é caso dos Programas de HIV/AIDS, onde os objetivos tendem a ser ambiciosos e genéricos, e, ainda que algumas de suas tecnologias sejam bem definidas, sua implantação e seu poder de intervenção e impacto estão extremamente vinculados a variáveis de processo e de contexto.

A avaliação aqui é definida como parte de um projeto que procura auxiliar os formuladores de políticas, gerentes e agentes, a desenvolver objetivos e resultados esperados para um programa de forma mais clara. Nesse sentido, a avaliação pode contribuir para identificar objetivos que estejam “encobertos”, que não estejam claramente enunciados, assim como para desenvolver consensos sobre quais são os objetivos mais operacionalizáveis.

Mais importante ainda, a avaliação pode ser um auxiliar para desenvolver teorias que articulem coerentemente as finalidades com as ações propostas. Isto é especialmente importante em programas que lidam com indicadores de efetividade muito “longínquos” como “a diminuição da infecção pelo HIV na população”. Muito embora este seja um objetivo consensual óbvio, é menos um indicador e muito mais um horizonte. Como bem coloca Ayres (1996:38): *“como chegar até este resultado? O que é e o que não é negociável na ‘economia vital’ dos sujeitos para alcançar este fim? (...) Quais as formas mais corretas, verdadeiras e autênticas de atingi-lo?”*.

Há algumas metodologias como o método Delphi ou outros métodos de discussão e de

estabelecimento de consensos que podem contribuir para estas reflexões e ampliar o acesso a processos e resultados objetiváveis. Ressalte-se que programas que lidam imediatamente com dimensões tão humanas – como a “economia vital” dos sujeitos na prevenção da infecção pelo HIV – devem dedicar muito mais atenção aos processos comunicativos.¹² A experiência das avaliações na área educacional, que lamentavelmente tem “conversado” muito pouco com a área da saúde, pode contribuir muito nesta discussão.

É preciso lembrar finalmente que a definição de “resultados esperados” sempre será um consenso provisório. Isto vale para qualquer programa. Programas de saúde lidam imediatamente com sujeitos. Não se pode, ou melhor, não se deve, esperar respostas unívocas dos vários sujeitos envolvidos. As várias respostas encontradas devem ser revistas também como argumento do diálogo entre o sujeito-serviço ou programa de saúde e os demais. Se isto é aplicável para qualquer programa de saúde, mesmo para aqueles de objetivos muito específicos, claros e consensualmente óbvios, é, em muito maior grau, para os que lidam com questões mais complexas e mais humanas.

É POSSÍVEL AVALIAR?

Os livros de avaliação colocam que o modo ideal para avaliar o impacto de uma intervenção é medir seu efeito em um grupo, quando comparado a um outro grupo que não foi alvo da intervenção. Em ambos os grupos as pessoas-alvo do programa são distribuídas aleatoriamente. Mas nem é preciso explicar a enorme dificuldade de conseguir isto na vida

¹² Um exemplo de proposição de indicadores para objetivos comunicacionais em atenção primária à saúde pode ser encontrada em Sala, Nemes e Cohen, 1998.

real.¹³ Basta lembrar que seria moralmente impossível submeter grupos de pessoas a, por exemplo, permanecerem sem orientação acerca do uso de preservativos, por pertencerem a um “grupo-controle”.

Na vida real, no cotidiano das ações de saúde, a situação é ainda mais complexa: temos intervenções “maiores” e/ou “melhores”, cuja desigualdade distributiva permite um acesso diferenciado para alguns grupos e não para outros, sobretudo em um país com diferenças sociais e regionais tão acentuadas quanto o nosso.

Ambas as assunções acima – as questões éticas envolvidas e as diferenças na quantidade e qualidade das intervenções entre distintos grupos populacionais – não impossibilitam avaliações válidas, só as tornam mais complicadas metodologicamente.

Para nossa realidade, creio que avaliações “combinadas” de processo e impacto podem ser uma importante alternativa metodológica para contornar estas dificuldades. Já apontamos a necessidade de conhecermos o processo para a as questões de imputabilidade. Aqui reiteramos que se avaliar o impacto das intervenções é um objetivo muito importante, tão importante quanto é entender o porquê de impactos diferentes. Mesmo nos programas mais maduros, as

¹³ O desenho de avaliações de impacto descrito há mais tempo, nos anos 20, e até hoje considerado o “mais poderoso” nas avaliações de impacto, é o chamado desenho experimental: compara-se os efeitos em um grupo-alvo obtidos por um programa com os efeitos semelhantes obtidos em outro grupo comparável, mas que não foi alvo do programa, um grupo controle. Para ser válido, este desenho exige várias condições metodológicas, sendo uma das mais importantes a aleatorização dos alvos, ou seja, a distribuição aleatória das pessoas alvo do programa entre o grupo com programa e o grupo-controle. A dificuldade em realizar este desenho na vida real levou os avaliadores, mais tarde, nos anos 60, a adotar maciçamente os chamados desenhos “quase-experimentais” nos quais a comparação grupo-alvo/grupo-controle permanece, mas, entre outras mudanças, a aleatorização prévia é “substituída” por técnicas estatísticas de regressão aplicadas posteriormente. Ou seja, técnicas que permitem controlar variáveis (como idade, sexo etc.) pós

avaliações de impacto, quando conjugadas com as avaliações causativas de processo, podem esclarecer muito mais. Analisar o processo, “abrir a caixa preta”, é uma necessidade, particularmente quando o impacto obtido é menor do que o esperado.

Quando se tem uma descrição adequada dos programas e se pode identificar seus componentes é possível comparar impactos e, o mais importante, associá-los a componentes do programa. O desenho combinado da avaliação exige que se conheça muito bem o programa, tanto suas finalidades, que permitem a avaliação de impacto, quanto seus componentes, o conjunto de ações que o compõe, conhecimento básico para a avaliação de processo.

Conforme vimos assinalando, essa clareza de objetivos e de atividades nem sempre é possível, sendo mais facilmente encontrada em programas mais “maduros” já submetidos anteriormente a avaliações normativas e a avaliações tanto de implantação como de processo de forma mais rotineira.

Vencidas as dificuldades apontadas, esse desenho se mostra muito poderoso em seu poder explicativo. Coerente com a ambição de simultaneamente avaliar processos e resultados, enfrenta o desafio de “abrir a caixa preta”, não se limitando a identificar apenas a entrada e a saída, ou o “input” e o “output”, e sim procurando associar os resultados (output) com os componentes do programa.

estudo. Esta foi também a evolução dos estudos epidemiológicos que muito se beneficiaram das técnicas de regressão. Tanto no campo da avaliação, quanto no campo da epidemiologia, os alcances e limites destes estudos controlados *a posteriori* têm sido muito discutidos. Vale assinalar, no entanto, que os problemas relativos ao uso de técnicas de regressão pós estudo possuem características e limites diferentes quando são utilizados em estudos epidemiológicos ou em estudos avaliativos, especialmente em função da distinta natureza de seus objetos, o que já foi muito bem apontado por Campbell em 1979.

A técnica metodológica mais apropriada para esse tipo de desenho é a da adição ou combinação de componentes – a articulação dos componentes de um programa comparada com a articulação de outro programa, ou com a adição de componentes. Por exemplo, uma avaliação que compare um programa que utiliza em suas intervenções duas técnicas educativas e um outro que faz uso dessas duas técnicas e de mais uma terceira. Existem técnicas estatísticas, como, por exemplo, a análise multivariada de componentes principais que podem ajudar no estabelecimento desses desenhos.

Mas é preciso definir muito bem quais são os componentes, ou qual é a combinação, a adição, entre eles, e ter critérios de corte, sobretudo quando se tem um número grande de variáveis. Por exemplo: comparar o uso consistente de preservativos (resultado) entre um grupo de jovens submetido a várias intervenções (marketing social, trabalhos nas escolas, formação de multiplicadores por ONGs), com um outro grupo de jovens submetido apenas a um tipo de intervenção (marketing social).

Outro exemplo poderia ser o da avaliação combinada da adesão do usuário ao tratamento com antiretrovirais (resultado) com componentes da qualidade da atenção (processo), comparando-se grupos submetidos a serviços com qualidade diferente, considerando-se componentes combinados como boa estrutura física e boa qualidade técnica do cuidado médico em um serviço com outro, com boa estrutura física e boa qualidade do cuidado médico, ao que se soma uma assistência integrada prestada por equipe multiprofissional.

Apesar de toda sua adequação para situações reais de programas, como os que nós vivemos, as análises de componentes são apenas metodologias que serão mais ou menos explicativas na dependência da definição e plausibilidade da associação entre o componente

e o efeito, ou, em outros termos, na dependência da adequação da teoria sobre o programa. Não haverá explicação válida com componentes mal definidos e/ou muito “agregados”.

Por exemplo: a qualidade do trabalho médico em um serviço de assistência, ou a qualidade do trabalho de aconselhamento em um serviço de testagem sorológica. Nesses casos, o componente “qualidade” precisa ser muito bem definido mediante indicadores com poder discriminatório, ou seja, que possuam sensibilidade e especificidade aceitáveis para apreender diferentes graus de qualidade.

Falar de qualidade suscita uma outra questão para a possibilidade da avaliação: o trabalho dos profissionais. Nossos programas dependem muito de trabalhadores especializados, com alto grau de autonomia técnica. Essa autonomia, característica substantiva do trabalho médico, é extensível, ainda que em graus variados, aos outros profissionais da área de saúde.

Avaliar implica julgamentos de valor com base em normas, parâmetros e padrões que descreverão a “boa prática” dos profissionais. Como descrevê-la objetivamente se ela depende tanto de decisões pessoais?

Em relação ao trabalho dos médicos, é possível dizer que embora este trabalho permaneça necessariamente dependente de julgamentos pessoais e de decisões arriscadas, a transformação tecnológica da medicina (Schraiber, 1998) ampliou significativamente as possibilidades de controle “externo” da prática. A polaridade entre a autonomia técnica e a maciça utilização de técnicas materiais de diagnóstico e terapêutica abrem um grande espaço de “negociação” possível entre a singularidade do ato médico individual e seu controle técnico externo. Grande parte deste controle é exercido ou no mínimo regulado pela própria corporação profissional mediante vários mecanismos como, por exemplo, o consenso para terapias antiretrovirais

do programa de AIDS. Esses mecanismos normatizam fortemente o trabalho médico, ao menos nas dimensões mais restritivamente técnicas como o uso de tecnologias materiais.

Entretanto, o encontro com o médico não é, evidentemente, apenas a “receita certa ou o exame certo”. É por demais conhecida a importância dos “outros conteúdos” do encontro médico-paciente. As chamadas tecnologias “não materiais”, especialmente as práticas discursivas, são aqui extremamente importantes.

Se elas são importantes para os médicos, são ainda mais importantes para outros profissionais, especialmente os psicólogos e assistentes sociais, cujo conteúdo técnico do trabalho depende muito pouco de tecnologias materiais.¹⁴ Isto também é evidente para as práticas mais propriamente educativas, como as de prevenção, nas quais atuam muitos tipos de “professores” além dos profissionais já citados como auxiliares de enfermagem, educadores de rua, redutores de danos etc. Vale lembrar que os médicos e professores são reconhecidos como as duas categorias profissionais de maior autonomia técnica.¹⁵

Apesar da institucionalização profissional, profissionais como psicólogos ou assistentes sociais contam com mecanismos de controle técnico externos (normas emanadas dos conselhos e das corporações) menos definidos e específicos do que os médicos. Isto implica que o jul-

gamento de seu trabalho é mais difícil uma vez que a própria abordagem terapêutica ou educativa é muito mais dependente do critério pessoal do profissional. Por outro lado, as normas da corporação são insuficientes para estabelecer julgamentos para além de padrões éticos genéricos. Se para profissões constituídas existe esta insuficiência, para os demais trabalhadores ainda não exatamente profissionalizados e muito especialmente nas práticas mais propriamente educativas, estas normas são praticamente inexistentes.

O trabalho em saúde e o trabalho educativo, sempre dependerão de autonomia técnica. Isto não quer dizer, contudo, que não possam ser submetidos – ao menos até certo ponto – a avaliações. O positivo exercício da criatividade e a responsabilidade do julgamento pessoal não podem alienar o sujeito-técnico dos outros sujeitos do trabalho,¹⁶ neste caso, dos sujeitos “representados” no processo avaliativo: os clientes ou a população-alvo do programa, os formuladores de políticas, os gerentes.

Os tensionamentos entre a necessidade de controle técnico e social e a necessária autonomia técnica do trabalho em saúde representam, por outro lado, desafios positivos para a avalia-

¹⁴ “... (os agentes do trabalho em saúde detém) possibilidades concretas para exercerem-se como sujeitos, executando escolhas e realizando vontades que não são essencialmente técnicas, mas expressões de valores éticos e políticos através da ação técnica (...). Isto significa uma expressiva presença do agente de trabalho na condição de sujeito social, já que esta forma de inserir-se no ato de trabalho demanda intenso e complexo exercício de subjetividade, denotando a forte base intelectual e subjetiva (da) técnica (Schraiber e Nemes, 1996:113).

¹⁵ Não se deve confundir a autonomia profissional com a capacidade criativa e intelectual de todo trabalhador enquanto ser humano. A autonomia significa o exercício destas capacidades conjugado ao poder de formular o projeto de trabalho, detendo, assim, o domínio sobre este (Schraiber e Nemes, 1996:112).

¹⁶ Assim, se da primeira perspectiva acima enunciada extraímos a característica, positiva a nosso ver, de possibilitar ao agente do trabalho que se coloque como sujeito em seu trabalho, da segunda, porém, decorre a característica de colocar-se como o **exclusivo sujeito no trabalho e único sujeito do trabalho**. Instaura-se, na relação intersubjetiva médico-paciente, uma desigualdade, alienando o médico da compreensão do outro como igualmente sujeito da sociedade. Com graus variáveis o mesmo ocorre quando se trata de relações entre médicos e outros profissionais de saúde e, até, entre médicos de diferentes especialidades, em que uns são mais “pares” ou iguais entre si que outros. As diferentes autoridades técnicas, portanto, se transmutam em desiguais autonomias técnicas, promovendo no trabalhador coletivo hierarquias de autoridade profissional. Por isso dizemos que a autonomia técnica promove **autonomias hierárquicas** no plano institucional, autonomias agora quanto ao poder que as dimensões administrativo-gerenciais do trabalho coletivo conseguem ter, para submeter à sua supervisão e controle cada trabalhador individual (Schraiber e Nemes, 1996:116).

ção. O próprio processo avaliativo pode ser um momento de reflexão sobre a prática¹⁷ e de estabelecimento negociado de indicadores de “boa prática”.

A autonomia é, sem dúvida, um desafio para a avaliação. É preciso lembrar, contudo, de outra dimensão do trabalho dos profissionais de saúde, que também têm implicações sérias para a avaliação, e que representa o “outro lado” da autonomia. Se a autonomia evidencia a relatividade das estruturas também é preciso salientar o percurso oposto, isto é, lembrar os fatores que limitam os exercícios de subjetividade: as restrições que a organização do trabalho impõe aos sujeitos. Estas podem ser tantas que a situação de trabalho parece o oposto da autonomia: os agentes do trabalho reclamam dos “padrões mínimos” impostos pela instituição que os impediria de usar todos os instrumentos “cientificamente adequados” a serviço do cliente. A prática, assim, parece estar sempre aquém do “padrão aceitável”, o que impediria, de antemão, qualquer processo avaliativo.¹⁸ Claro que, lamentavelmente existem condições de trabalho tão inadequadas nas quais isto é real. Abaixo de um certo nível de estrutura não dá para julgar nenhum processo, aliás, nestes casos será impossível avaliar qualquer programa.

¹⁷ “Considerando-se, ademais, que os profissionais precisam reconhecer a complementaridade e proceder ativamente à articulação das ações, o que se têm como questão é a forma de realizar as necessárias autonomias técnicas através dessa complementaridade e interdependência. Espera-se, aqui, um duplo movimento por parte da equipe de saúde, pois tanto a autonomia técnica dos profissionais quanto a articulação das ações, são necessárias para a eficiência e eficácia dos serviços. Movimento que requer o exercício da autonomia técnica de forma interdependente, autonomias reconhecidas e negociadas pelos distintos agentes, por meio de sua interação” (Schraiber, Peduzzi, Sala, Nemes, Castanheira, Kon, 1999:234).

¹⁸ Agregue-se a isto o fato de que raramente este padrão “aceitável” encontra-se estabelecido. Ao contrário, o discurso do técnico, neste casos, legitimado pela ciência, sempre busca padrões “máximos”.

Felizmente, isto não acontece sempre e muitos programas contam com um “mínimo aceitável” de estrutura. Mesmo nestes casos, entretanto, enfrentam muitas vezes a cruel oposição entre princípios éticos-normativos como o princípio da extensão/universalização “versus” o princípio da “melhor” qualidade possível. Estas oposições de fato tenciona o processo avaliativo. Traduzidas no discurso dos trabalhadores pela triste expressão “aqui fazemos o mínimo”, esta oposição resulta na cisão entre os planos ético, político e técnico do trabalho, reduzindo o plano ético a características pessoais do trabalhador e podendo paralisar qualquer tentativa de negociação entre a administração ou a gerência para o estabelecimento de parâmetros avaliativos (nestes casos, a gerência é empurrada à condição de responsável exclusiva pelo plano político, enquanto o trabalhador direto resta a responsabilidade exclusiva pelo plano técnico). Somente com base no enfrentamento claro e negociado deste desafio será possível o diálogo acerca das definições de parâmetros de avaliação aceitáveis. Na ausência deste diálogo, os processos avaliativos podem ser tomados pelos profissionais como, no mínimo, constrangedores, quando não como persecutórios e punitivos.¹⁹

¹⁹ Isto é especialmente importante nas vertentes de avaliação e garantia de qualidade: “Muito embora os teóricos da qualidade afirmem que mais importante que achar o culpado é achar e analisar o erro, até hoje os controles de qualidade mais tradicionais tendem, no mínimo, a ser restritivos, estanques e não vinculados a outros níveis da organização. Tanto esta noção punitiva tem base real, que há uma crença disseminada de que é mais fácil trabalhar com qualidade no setor privado do que no setor público, justamente pela relativa impunidade dos trabalhadores no setor público (...) (cabe assinalar que o abandono do caráter punitivo destes mecanismos, a assunção de que falhas são inevitáveis e de que ninguém erra por prazer, são a base da lealdade da organização para com seus agentes. E, evidentemente, são o ponto de partida para uma forma de avaliação que se pretenda dialógica entre a gerência e o trabalhador direto (...)) (só possível) de exercer se a gerência estiver, de fato, inteirada e ‘aliada’ ao trabalho direto” (Schraiber, Peduzzi, Sala, Nemes, Castanheira, Kon, 1999: 237).

A implementação de mecanismos de avaliação a partir de definições de fato negociadas entre os profissionais pode ser um mecanismo de resgate da profissionalização, do orgulho e da valorização do trabalho, lamentavelmente pouco frequentes em muitos programas de saúde.

COMO AVALIAR?

As questões suscitadas nos itens anteriores já fazem antever que esta pergunta também não tem respostas “tranqüilas”. Não há, rigorosamente, nenhuma metodologia “pronta” para um ou outro foco de avaliação.

É preciso ressaltar de novo aqui que o método depende do objetivo e do objeto da avaliação. O objetivo depende muito do nível da intervenção que se quer avaliar.

O nível – um projeto, um programa formado por um conjunto de projetos, uma política nacional de saúde etc. – definirá o foco de avaliação e determinará, em grande parte, o método. Assim todos os níveis de um programa de AIDS podem ter em comum uma só finalidade. Por exemplo, reduzir a taxa de transmissão da doença e melhorar a qualidade de vida dos já infectados. Cada um dos níveis organizacionais do programa, entretanto, poderá ter diferentes perguntas para a avaliação, o que implicará diferentes unidades de análise e métodos de avaliação.

Os níveis não são diferentes apenas pelo número de indivíduos-alvo envolvidos. Por exemplo, um foco de avaliação típico de nível organizacional superior nesse caso poderia ser a comparação entre médias de jovens que vêm recebendo atividades educativas em relação a sexo seguro e médias dos que vêm adotando este comportamento nos vários locais de execução do programa. Um desenho adequado para responder a esta questão é um desenho de avaliação quase-experimental e a metodologia de eleição é quantitativa.

Em contraste, a pergunta poderia ser como os jovens-alvo do programa vêm se relacionando com as equipes educativas e quais características deste relacionamento levam a maior autonomia nas escolhas relativas ao sexo. Esta é uma pergunta muito mais típica dos níveis executores do programa para a qual o desenho avaliativo mais adequado é o estudo de caso e a metodologia, a qualitativa.²⁰

Claro que os objetivos das avaliações podem ser, e são em geral, mais complexos, requerendo, muitas vezes, a adequada integração entre metodologias qualitativas e quantitativas. A adequação e maior poder explicativo dos estudos multi-métodos para muitas questões de pesquisa já estão suficientemente aceitos no campo. É interessante apontar, porém, que a integração quanti-quali não deve ser vista exclusivamente pelo poder do qualitativo para levantar questões *para* o quantitativo (como seria, por exemplo, a utilização de grupos focais para levantamento de questões que resultariam em variáveis dependentes em uma avaliação da satisfação dos usuários), ou ainda pela explicação dos achados qualitativos *acerca* das evidências quantitativas obtidas (como a realização de entrevistas em profundidade para esclarecer melhor porque uma variável como idade está associada com um resultado como adesão ao tratamento com antiretrovirais). Embora isto seja correto e muitas vezes suficientemente útil para uma dada questão de avaliação, nem sempre a utilização do qualitativo deve ser vista de modo tão subordinado ao quantitativo.

É preciso assinalar que os métodos qualitativos produzem evidências de qualidade diversa e, às vezes, aparentemente contraditórias aos

²⁰ Os estudos de caso têm sido considerados adequada estratégia de pesquisa quando a investigação reúne três condições: – a principal questão de pesquisa é do tipo “como?” e/ou “por quê?”; – o foco da pesquisa está em eventos contemporâneos; – não há controle sobre eventos relevantes (Yin, 1984).

métodos quantitativos. Assim, por exemplo, na pesquisa avaliativa sobre adesão ao uso de antiretrovirais realizada em São Paulo, a variável sexo não se mostrou associada com a não aderência, ou seja, o sexo feminino não mostrou nenhuma diferença na prevalência de não aderência em relação ao sexo masculino (Nemes, Souza, Kalichman e cols., 1999). Isto quer dizer que se pode aceitar, nos limites do grau de confiança estabelecidos pelo estudo, que *o grupo* de mulheres não difere, em termos de adesão ao tratamento, *do grupo* de homens. Por outro lado, o estudo qualitativo mostrou que as dificuldades ligadas ao modo de andar a vida são extremamente importantes na adesão ao tratamento, mais do que as questões diretamente ligadas ao tratamento. Para as mulheres as dificuldades relacionadas ao modo de andar a vida podem estar em parte determinadas pela condição de gênero, tais como as tarefas domésticas, o cuidar de filhos etc.

Esta produção de verdades diferentes tem a ver, neste caso, com a questão da adequação do método *ao objeto, ou foco da avaliação*, o que suscita imediatamente a questão dos objetos mais “obscuros” para a avaliação. Um dos mais importantes destes objetos é justamente o “*interior técnico*” do trabalho dos profissionais, como apontado acima. Para este tipo de objeto, a metodologia qualitativa se impõe. O uso integrado de entrevistas com os profissionais e observações diretas do trabalho, por exemplo, pode revelar esquemas operativos dos vários espaços de intervenção onde atuam os profissionais com grande autonomia técnica; pode revelar o modelo tecnológico deste trabalho.

Na mesma avaliação acima mencionada foi o uso integrado dessas metodologias que pôde produzir evidências que levaram à conclusão da baixíssima especificidade do trabalho dos serviços de saúde com a questão da adesão ao tratamento, ou em outros termos, da ausência de

instrumentos tecnológicos direcionados para a adesão no trabalho dos médicos e da equipe de enfermagem (Castanheira et al., 1999).

Apesar da grande discussão nos últimos anos sobre os métodos qualitativos na pesquisa em saúde, certas questões metodológicas ainda despertam dúvidas. Se a objetividade na metodologia qualitativa parece mais esclarecida,²¹ a questão da generalização dos dados qualitativos permanece ainda como uma dúvida que aparece de modo muito freqüente entre as pessoas que querem começar algum tipo de avaliação. Embora esta questão esteja muito bem tratada na literatura parece importante enfatizar aqui dois pontos. Primeiro: enquanto as pesquisas amostrais lidam com generalizações *estatísticas*, os estudos qualitativos, tal como os estudos de caso, lidam com generalizações *analíticas*, exatamente como acontece nos estudos experimentais.²² Em segundo lugar: se a hipótese do trabalho foi construída de um modo historicamente consistente e se as evidências empíricas mostrarem a produtividade desta hipótese, as evidências se aplicarão ao conjunto (e não a cada um) dos objetos (por exemplo, um conjunto de serviços de assistência à AIDS onde se está avaliando a prática gerencial). Esta aplicação que não é para cada um, mas para o conjunto, todavia *só funcionará para o nível de abstração em que se está trabalhando* (no exemplo, o resultado poderia ser um *modelo gerencial* capaz de identificar a racionalidade principal da prática gerencial do conjunto dos serviços).

²¹ “o real que (o pesquisador) conhece é aquele que ele realiza, e sua objetividade é uma construção que se dialetiza no processo de objetivização e subjetivização (...) a polaridade complementar entre sujeito e objeto (leva) a necessidade do esforço de objetivização, ou seja, a busca de minimizar as incursões do subjetivismo e do espontaneísmo pelo compromisso interno de discussão teórico metodológica e pelo compromisso social de submeter a abordagem aos critérios da prática (...) A relação dialética entre o esforço lógico e os compromissos sociais (...) é o parâmetro principal da objetivização (...) (que não exclui o sujeito e sua historicidade (...))” (Minayo, 1992:250-1).

Existem muitas metodologias qualitativas e quantitativas que podem auxiliar processos avaliativos. Em avaliações promovidas por programas na vida real muitas vezes será necessário recorrer a especialistas que possam assessorar o processo e auxiliar na escolha das metodologias mais adequadas a cada caso. Nenhuma metodologia, entretanto, é suficientemente esclarecedora se não estiver adequada a uma teoria consistente. Isto vale para os métodos quantitativos e qualitativos.

Ainda em relação ao pólo metodológico é importante lembrar a necessidade – em qualquer abordagem – de garantir a máxima confiabilidade e validade das medidas que compõem os indicadores de avaliação, especialmente quando se trata de avaliações internas nas quais, muitas vezes, o pesquisador é o próprio agente do trabalho que se está avaliando. Neste ponto também o recurso a especialistas pode ajudar muito.

De novo aqui é preciso enfatizar que o necessário cuidado com a metodologia não deve paralisar as propostas de avaliação. Nenhum “pacote metodológico” poderá substituir a fase inicial e fundamental de qualquer avaliação, resumidas nas primeiras perguntas: por que, para quê e o quê avaliar.

Costumo dizer que há um lado muito bom de não termos no Brasil avaliadores profissionais. Isto tem nos obrigado a um esforço de reflexão talvez maior. Alguém com experiência em avaliação pode ajudar, é claro, mas nunca substi-

Isto não nos exime da necessidade de clarificar os passos metodológicos do processo de pesquisa, ao contrário, torna-a mais exigente. Situar-se como sujeito que investiga exige aqui uma “*reeducação moral do pesquisado*” pelo menos até o ponto de permitir uma objetivação do sentido das ações e das relações observadas (Fernandes, 1972:25).

²² Os estudos qualitativos rigorosos não se limitam a uma descrição. Devem ser guiados por um esquema teórico que serve de princípio diretor inclusive para a coleta de dados; só assim podem evitar o privilegiamento exagerado do empírico ou o “erro do concreto mal colocado” (Bruyne et al, s/d).

tuirá o conhecimento do objeto de trabalho, que, afinal, é o fundador de qualquer teoria acerca de uma intervenção. Mas também nunca substituirá o sincero desejo de correr todos os riscos de uma avaliação. Para isto, além de conhecer, é necessário um compromisso moral de base.

Aqui me parece estar a maior fecundidade dos processos de avaliação: a revelação, publicação e renovação deste compromisso.²³ Por isto também a compreensão, aceitação e adesão das equipes de trabalho aos processos de avaliação são tão importantes, mesmo quando se trata de uma avaliação tipicamente “externa”. Por outro lado, os processos avaliativos que envolvem maior participação dos agentes e clientes são potencialmente mais produtivos neste sentido.

DESAFIOS PARA A AVALIAÇÃO DOS PROGRAMAS EM HIV/AIDS: ALGUMAS SUGESTÕES À GUIA DE CONCLUSÃO

Os programas de HIV/AIDS oferecem enormes desafios para a avaliação. Muitos destes desafios são extensivos a outros objetos de intervenção em saúde. Creio mesmo que as especificidades dos programas podem auxiliar a elaboração de novas e melhores estratégias para muitos outros programas, isto porque a intervenção em AIDS torna questões, há muito presentes na saúde pública, agudas e públicas, tais como a dificuldade de integração de práticas coletivas e individuais, a insuficiência dos instrumentos tecnológicos habituais para lidar com várias dimensões do sofrimento humano ou ainda a emergência de dilemas éticos e morais na assistência

²³ Referindo-se à alegoria da criação do *homo* pelo Cuidado Ayres (1996) coloca: “a avaliação é como o Cuidado da alegoria, é a curadora do trabalho sem ser o trabalho em si mesmo. O sopro vital que anima nossa criação não nos pertence, são vidas, valores, experiências, vontades do Outro (...) Enquanto perdura no tempo, enquanto permanece, o trabalho une e dá sentido a esse misto de contingências e liberdades” (p. 40).

aos doentes articulada a uma grande velocidade na incorporação de tecnologias materiais.

As práticas de avaliação podem auxiliar a identificar superações para algumas destas questões. Nesta direção, oferecemos ao debate algumas sugestões gerais de aproximações em avaliação, potencialmente responsivas, relacionadas aos dois grandes conjuntos tecnológicos dos programas, a prevenção e assistência.

A área da prevenção reúne uma série de atributos que dificultam enormemente a avaliação. Na base de todos eles está a conhecida dificuldade em avaliar com as práticas educativas. Aqui, os requisitos metodológicos das “boas avaliações” dos quais tratamos no início (objetividade, validade, responsividade etc.) são tão mais difíceis de alcançar quanto mais “desarmônicos” entre si. Articula-se ainda a isto dificuldades de muitas ordens na definição tanto da “boa prática” quanto de resultados ou impactos. Para dificultar ainda mais, as práticas têm um baixo nível de normatização, além de operarem em múltiplos sítios com instrumentos e agentes muito diversos.

Um processo de avaliação das atividades de prevenção poderia, por exemplo, basear-se em ciclos sucessivos de discussão e de negociação acerca do modelo de avaliação. Estes ciclos se iniciam pela formação de consensos (e “dissensos”) sobre o programa (ou conjunto de atividades) baseados nas representações empíricas sobre experiências e significados do programa para todos os informantes que vivenciam a experiência do programa e seus efeitos: usuários diretos e indiretos; agentes de trabalho, gerentes e administradores, representantes de outras instituições e projetos articulados ao programa, não usuários interessados etc. (por exemplo: jovens escolares; pais e amigos, professores, diretores de escola; representantes de ONGs e de serviços de saúde; educadores não diretamente envolvidos etc). Estas discussões

iriam montando os vários passos da avaliação, estabelecendo respostas provisórias a perguntas como “o que é o programa?”, “quais são suas finalidades?”, “como ele está funcionando?”, “quais são seus principais componentes?”, “estes componentes estão relacionados com as finalidades? como?”. A progressão destas discussões seria alimentada e progressivamente “sofisticada” por informações e análises obtidas por outros métodos: análise de documentos e relatórios, revisão da literatura; entrevistas etc. O objetivo destes ciclos será estabelecer um consenso possível acerca das “grandes perguntas” da avaliação: “por que?”, “para quê?”, “o quê?”, “é possível?”, “como?”, de modo a estabelecer um plano *operativo* para a execução da avaliação. Os ciclos de negociação recomeçam a partir dos resultados obtidos na “primeira” avaliação, questionando o produto da avaliação, sua utilização e a necessidade de novos ciclos e novas avaliações.

Neste tipo de metodologia o “avaliador” é menos um especialista e mais um facilitador do processo, o que não dispensa o concurso de especialistas para questões tais como a factibilidade de alternativas metodológicas ou análise da literatura e de documentos e até mesmo para o concurso de “teorias rivais” acerca dos objetos da avaliação. Estes aportes objetivam aumentar a objetividade e a confiabilidade do modelo de avaliação. É preciso ressaltar, porém, que este tipo de avaliação não pode garantir que todos os principais “requisitos metodológicos” aconteçam simultaneamente. Muito baseada em vivências e enfoques naturalísticos, sempre tenderá a perder em validade externa, ganhando, porém, em responsividade.

Esta sugestão está genericamente inspirada em alguns conceitos centrais das linhas de avaliação que mais valorizam a participação e o “empoderamento” em processos avaliativos, desenvolvidos especialmente para programas educativos nos EUA (Guba & Lincoln, 1989).

Como já afirmado muitas vezes, tais processos estão intimamente relacionados com contextos históricos próprios, podendo, entretanto, servir – com as necessárias adaptações – como inspiradores de projetos de avaliação entre nós.

Em relação à assistência, creio que um desafio muito grande diz respeito à qualidade do cuidado. A qualidade do cuidado está fortemente associada ao impacto esperado dos programas de saúde, especialmente para condições que exigem acompanhamento contínuo da equipe de saúde e que envolvem grande aplicação de tecnologias complexas (materiais e não-materiais), como é o caso da AIDS.

A Coordenação Nacional de DST e Aids tem procurado estabelecer um padrão de qualidade mínima para a assistência mediante definição de normas para o credenciamento de Serviços de Assistência Especializada (SAE). Mesmo considerando que estas normas tratam apenas de padrões mínimos relacionados à estrutura do cuidado, configurando um tipo de acreditação de serviços, ainda assim, a cobertura de sua aplicação tem sido apenas parcial.

De fato, os serviços do sistema público de saúde brasileiro que prestam assistência aos portadores de HIV/AIDS possuem configurações administrativas e institucionais bastante diversas que provavelmente implicam níveis diferentes na qualidade do cuidado prestado aos pacientes e, por conseguinte, diferentes resultados no impacto sobre a população-alvo. Esta hipótese foi fortemente corroborada na avaliação da adesão dos pacientes ao tratamento antiretroviral realizada em 27 serviços de São Paulo: quando estratificados em níveis de qualidade segundo características “clássicas” de estrutura e processo do cuidado,²⁴ mostraram riscos de não-adesão

²⁴ As variáveis utilizadas foram: quantidade e capacitação técnica da equipe; suprimento de ARV e outras medicações importantes; acesso a exames específicos; intervalo de tempo para dispensação, orientações; controle de abandono/faltosos; se-

significativamente diferentes, *independentemente* das características dos usuários ou do tratamento.²⁵

As características básicas de estrutura e processo do cuidado poderiam ser tomadas como ponto de partida para avaliações locais que envolvessem a equipe e os usuários em “ciclos de qualidade”. Os ciclos objetivam estabelecer parâmetros de qualidade que serão adotados e monitorados durante um determinado período após o qual um novo ciclo identificará os resultados favoráveis e desfavoráveis que aconteceram e esta será a base da criação de um novo parâmetro de qualidade, mais ambicioso e sofisticado do que o anterior. Embora genericamente inspirado nos “quality circles” utilizados no Japão no setor industrial nos anos 80, que resultaram na tão criticada “gerência de qualidade total”, tem sido usado com sucesso em alguns programas de saúde (Schillemans & cols., 1992). Novamente aqui é importante assinalar que o aproveitamento desta metodologia ou de outra qualquer deve ser cuidadosamente pensado para cada situação avaliativa. No caso desta metodologia, por exemplo, o que me parece mais produtivo é sua potencialidade para aumentar a responsividade da avaliação e de introduzir novas normas – não restritivamente técnicas – na definição de qualidade.

Agradecimentos

À H Maria Novaes, Elen R. Lodeiro Castanheira e colegas do Grupo QualiAids, do Departamento de Medicina Preventiva da FMUSP, com quem tenho compartilhado reflexões presentes neste texto. A todos os companheiros das avaliações do Programa DST/AIDS de São Paulo.

guimento do consenso na prescrição de ARV; desempenho e acolhimento da equipe; satisfação com o serviço (Nemes e cols, 1999; Castanheira e Nemes, 1999).

²⁵ Os resultados apresentados em São Paulo foram a base de um protocolo de pesquisa nacional que objetiva estabelecer um

BIBLIOGRAFIA CITADA²⁶

Aguilar, MS, Ander-Egg, E. *Avaliação de Serviços e Programas Sociais*. Petrópolis, Vozes, 1994.

*Ayres, JRCM. “Educação preventiva e vulnerabilidade as DST/AIDS e abuso de drogas entre escolares: como avaliar a intervenção?” In: Tozzi, D e cols. (orgs.) *Papel da educação na ação preventiva ao abuso de drogas e as DST/AIDS. Série Idéias, 29*. São Paulo, Fundação para o Desenvolvimento da Educação, 1996. [25-41].

Bruyne, P, Herman, J, Schoutheete, M. *Dinâmica da pesquisa em ciências sociais*. Francisco Alves, 3ª edição, s/d.

Campbell, D. *Assessing the impact of planned social change. Evaluation and Program Planning*, vol. 2, 1979. [67-90]

*Castanheira, ERL, Capozzolo, AA, Nemes, MIB. *Avaliação da aderência ao tratamento por antiretrovirais em usuários de ambulatórios do sistema público de assistência à AIDS no Estado de São Paulo, Brasil. Avaliação das características tecnológicas do processo de trabalho em serviços de saúde selecionados*, 1999. Arquivo disponível na Internet: www.aids.gov.br

*Castanheira, ERL, Nemes, MIB. *Avaliação da aderência ao tratamento por antiretrovirais em usuários de ambulatórios do sistema público de assistência à AIDS no Estado de São Paulo, Brasil. Avaliação das características de estrutura e processo dos serviços de saúde*, 1999. Arquivo disponível na Internet: www.aids.gov.br

Champagne, F, Contandriopoulos, AP, Pineault, RA. “Health care evaluation framework”, *Health Management Forum*, Summer, 1986.

Chen, HT. *Theory-driven evaluations*. Newbury Park, CA, Sage Publications, 1990.

perfil de qualidade do cuidado para todos os serviços públicos que atendem AIDS em nível ambulatorial no Brasil. Posteriormente, serviços representativos dos diversos níveis de qualidade terão avaliada a média de adesão dos pacientes ao tratamento ARV (Protocolo elaborado pelo Grupo de Pesquisa QualiAids. Informações sobre o protocolo e o andamento da pesquisa podem ser solicitadas à autora deste texto).

²⁶ As referências que vêm precedidas por asterisco referem-se diretamente a programas de HIV/AIDS.

Conrandiopoulos, A.P, Champagne F., Denis J.L., Pineault, R. “A avaliação na área de saúde. In: Hartz, ZMA (org.). *Avaliação de Saúde. Dos modelos conceituais à prática na análise da implantação de programas*. Rio de Janeiro, Editora Fiocruz, 1997. [29-47]

Donabedian, A. “Definition of quality and approaches to its assessment”. In: *Explorations in quality assessment and monitoring*, vol. I. Michigan, Health Administration Press, 1980.

Fernandes, F. *Fundamentos empíricos da explicação sociológica*, São Paulo, Editora Nacional, 1972 (2ª ed).

Guba, EG, Lincoln, YS. “The Methodology of Fourth Generation evaluation”. In: *Fourth Generation Evaluation*, Sage, 1989. [184-227]

Hartz, ZMA (org.). *Avaliação de Saúde. Dos modelos conceituais à prática na análise da implantação de programas*. Rio de Janeiro, Editora Fiocruz, 1997.

Malik, AM. “Qualidade em serviços de saúde nos setores público e privado”. *Cadernos Fundap*, 19, 1996. [7-24]

Minayo, MCS. *O desafio do conhecimento. Pesquisa qualitativa em saúde*. São Paulo, Rio de Janeiro, Hucitec-Abrasco, 1992.

*Nemes, MIB. *Avaliação da aderência ao tratamento por antiretrovirais em usuários de ambulatórios do sistema público de assistência à AIDS no Estado de São Paulo, Brasil. Visão Geral dos Resultados*, 1999. Arquivo disponível na Internet: www.aids.gov.br

*Nemes, MIB, Souza, MFM, Kalichmam, AO, Grangeiro, A, Souza, RA, Lopes, J. *Avaliação da aderência ao tratamento por antiretrovirais em usuários de ambulatórios do sistema público de assistência à AIDS no Estado de São Paulo, Brasil. Prevalência da aderência e fatores associados*, 1999. Arquivo disponível na Internet: www.aids.gov.br

*Nemes, MIB, Melchior, R, Jordan, MS, Okazaki, E, Komatsu, CL. *Avaliação da aderência ao tratamento por antiretrovirais em usuários de ambulatórios do sistema público de assistência à AIDS no Estado de São Paulo, Brasil. Análise qualitativa de entrevistas semi-estruturadas com usuários*, 1999. Arquivo disponível na Internet: www.aids.gov.br

Nemes, MIB. *Avaliação do trabalho programático na aten-*

ção primária à saúde. Tese de Doutorado, Faculdade de Medicina da USP, São Paulo, 1996.

Novaes, HMD. *Mortalidade neonatal e avaliação da qualidade da atenção ao parto e ao recém-nascido no Município de São Paulo*. Tese de Livre Docência. Faculdade de Medicina da USP, Departamento de Medicina Preventiva, 1999.

Sala, A, Nemes, MIB, Cohen, DD. "Metodologia de avaliação do trabalho na atenção primária à saúde". *Cadernos de Saúde Pública*, 14(4), out/dez, 1998. [741-751]

Schillemans, L, De Grande, L, Remmen, R. "Using quality circles to evaluate the efficacy of primary care". In: Cornner, RF & Hendricks, M (eds.). *International Innovations in Evaluation Methodology. New Direction for Program Evaluation*, nº 42, San Francisco, Jossey-Bass, 1989.

Schraiber LB. *O Médico e seu Trabalho. Limites da Liberdade*, São Paulo, Hucitec, 1998.

Schraiber, LB, Nemes, MIB, Sala, A, Castanheira, ER, Kon, R. "Planejamento, gestão e avaliação em saúde: identificando problemas". *Revista Ciência e Saúde Coletiva* 4(2),1999. [221-242]

Schraiber, LB, Nemes, MIB. "Processo de Trabalho e avaliação de serviços de saúde". *Cadernos FUNDAP*, nº 19, janeiro/abril,1996. [106-121]

Schraiber, LB, Peduzzi, M, Sala, A, Nemes, MIB, Castanheira, ER, Kon, R. "Planejamento, gestão e avaliação em saúde: identificando problemas". *Ciência e Saúde Coletiva*, vol 4, nº 2, 1999. [221-242]

Yin, RK. *Case Study Research: Designs and Methods*, Sage, 1984.

BIBLIOGRAFIA CONSULTADA

*Basso, CR. *A qualidade da assistência aos portadores de HIV/AIDS nos serviços descentralizados do Estado de São Paulo*. Dissertação de Mestrado, Faculdade de Saúde Pública da USP, 1999.

*Bennett, FJ. "Qualitative and quantitative methods: In depth or rapid assessment?" *Social Science and Medicine*, vol. 40, nº 12, 1995. [589-90]

*Bindman, AB, Osmond, D, Hecht, FM, Vranzin, K & Reingold, A. "Multistate evaluation of anonymous testing and access to medical care". *JAMA*, 280:16, 1998 Oct. [1416-1420]

*CDC. Center For Disease Control And Prevention. *Guidelines for AIDS prevention program operations*, 1987.

*CDC. Center For Disease Control And Prevention. *HIV counseling, testing and referral standards and guidelines*, 1994.

CDC. Center For Disease Control And Prevention. "Framework For Program Evaluation". In: *Public Health*. MMWR:48 (11), 1999. [1-40]

*Chittick, JB. *Theoretical rationale and short guide for developing and implementing your own research designs (an outline of research methodologies)*, 1996. Arquivo disponível na Internet: www.tecnaids-peercorps.com

Donabedian, A. "Reflections on the Effectiveness of Quality Assurance". In: Palmer, RH, Donabedian, A, Povar, G. *Striving For Quality In Health Care. An Inquiry into Policy and Practice*, Ann Arbor/ Michigan, Health Administration Press, 1991.

Donabedian,A. "Continuity And Change In The Quest For Quality". *Clinical Performance And Quality Health Care*, 1(1),1993. [9-16]

Donabedian, A. "The Quality Of Care. How Can It Be Assessed?" *JAMA*, 260 (12),1988. [1743-1748]

Donabedian, A. "The Role Of Outcomes In Quality Assessment And Assurance". *Qrb*, 18, 1992. [356-360]

Donabedian,A. "The Seven Pillars Of Quality". *Arch. Pathol. Lab.Med.*, 114, 1990. [1115-1118]

Guba, EG, Lincoln, YS. *Effective Evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*, San Francisco, Jossey-Bass, 1981.

Habicht, JP, Victora, CG & Vaughan, JP. "Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact". *International Journal of Epidemiology*, 28, 1999. [10-18]

*Laracque, F et al. "Effect of comprehensive

- intervention program on survival of patients with human immunodeficiency virus infection". *Arch. Inter. Med.*, 156, 1996.
- Lehoux, P, Levy, R, Rodrigues, J. "Conjuguer la modélisation systématique et l'évaluation de 4^{ème} génération". *Ruptures, revue transdisciplinaire en santé*, 2(1), 1995. [56-72]
- Mays, N & Pope, C. "Qualitative research in health care. Assessing quality in qualitative research". *MBJ*, Vol. 320, January 2000. [50-52]
- *Mehret, M et al. "Baseline for the evaluation of an AIDS programme using prevention indicators: a case study in Ethiopia". *Bulletin of the World Health Organization*, 74 (5), 1996. [509-516]
- *Mertens, TE & Carael, M. "Evaluation of HIV/AIDS prevention, care and support: An update on WHO's approaches". *AIDS Education and Prevention* 9 (2), 1997. [133-145]
- Miller, R, Fischet, A, Miller, K, Ndhlovu, L, Maggwa, BN, Sanogo, IAD & Tapsoba, P. "The situation analysis approach to assessing family planning and reproductive health services". *A handbook*, Population Council, 1997.
- *Minayo, MCS, Cruz Neto, O, Souza, ER, Assis SG e Deslandes, AF. *Avaliação qualitativa de programas ou serviços de saúde*. Dez, 1996. [mimeo]
- Minayo, MCS. "Abordagem antropológica para avaliação de políticas sociais". *Revista de Saúde Pública*, São Paulo, 25 (3), 1991. [233-238]
- Novaes, HMD. "Epidemiologia e avaliação em serviços de atenção médica: novas tendências de pesquisa". *Cadernos de Saúde Pública*, 12 (supl), 1996. [7-12]
- Novaes, RL. "Sobre a Técnica". *Manguinhos*, vol. III (1), mar/jun, 1996. [24-49]
- *Osmond, D, Bindman, AB, Vranzia, K, Hecht FM, Keane D & Reingold, A. "Name-based surveillance and public health interventions for persons with HIV infection". *Ann Intern Med*, 131:10, 1999 nov. [775-779]
- Palmer, RH, Hargraves, LJ. "Quality Measurement And Improvement Among Primary Care Practitioners. Results Of The Ambulatory Care Medical Audit Demonstration Project". *Medical Care*, 34(9): Supplement.
- Patton, MQ. *Practical Evaluation*. Beverly Hill, CA, Sage, 1982.
- Pedersen, D. "Evaluación de servicios de salud: el punto de vista de la gente". *Enfoques en Atención Primaria* 6(2), 1991. [12-22]
- Perry, PD & Backus, CA. "A different perspective on empowerment in evaluation: benefits and risks to the evaluation process". *Evaluation Practice*, 16(1), 1995. [37-46]
- Shalock, RL. *Outcome-Based Evaluation*. New York: Plenum Press, 1995.
- Starfield, B. *Primary Care. Concept, Evaluation and Policy*. Oxford, Oxford University Press, 1992.
- *UNICEF/UNAIDS. "Vertical transmission of HIV". *Rapid Assessment Guide*, 1999. Arquivo disponível na Internet: www.unaids.org
- Venton, JP. "A general theory of delegation, accountability and empowerment". *The Canadian Journal of Program Evaluation*, vol. 12, nº 2. [163-188]
- WHO. *Rapid Evaluation Method Guidelines for maternal and child health family planning and other health services*, 1999. Arquivo disponível na Internet: www.who.int
- Worthen, BR, Sanders, JR & Fitzpatrick, JL. *Program Evaluation. Alternative Approaches and Practical Guidelines*, Longman Publishers, 1997 (2^a ed.).