

Princípios e padrões em metaavaliação: diretrizes para os programas de saúde

Principles and standards in metaevaluation: guidelines for health programs

Zulmira Maria de Araújo Hartz ¹

Abstract *This article poses questions imperative for the use of principles and standards of evaluation allowing for pointing out directives for metaevaluation on the example of the Baseline Studies (BLS) of PROESF. These questions are important in countries like Brazil, where a National Evaluation Policy is still in its beginning, and where one observes absence of legal devices, directives or parameters clearly seeking to ensure the quality, credibility and recognition of the practice of evaluation. We present a brief review of the literature about principles and norms constituting the bases for the general directives formulated for a formative metaevaluation of the baseline studies, oriented by the four international standards for evaluation of social policies – utility, feasibility, propriety and accuracy. The same standards were used for guiding the design of studies for evaluating health promotion and protection programs, together with specificity standards defined by a mixed consensus technique based on the prescriptions of the Terms of Reference of the BLS. The main interest rests in encouraging a broad use of these evaluation tools in the field of health, as a crucial point for its improvement and the development of the Unified Health System.*

Key words *Metaevaluation, Health programs, Institutionalization of evaluation*

Resumo *Este artigo apresenta questões imperativas para o uso de princípios e padrões em avaliação, permitindo destacar diretrizes para estudos de metaavaliação, tendo como exemplo os Estudos de Linha de Base (ELB) do Projeto de Expansão e Consolidação do Saúde da Família (Proesf). Essas questões são importantes em países em que a Política Nacional de Avaliação ainda é incipiente, como é o caso do Brasil, onde se observa a ausência de dispositivos legais, diretrizes, ou parâmetros que busquem claramente garantir a qualidade, credibilidade e reconhecimento da prática da avaliação. Apresenta-se uma breve revisão da literatura sobre princípios e normas que fundamentam as diretrizes gerais formuladas para uma metaavaliação formativa dos estudos de linha de base, orientada pela aplicação dos quatro padrões internacionais para avaliação das políticas sociais: utilidade, exequibilidade, propriedade e precisão, que têm igualmente servido de guia para o desenho dos estudos de avaliação de programas de promoção e proteção à saúde, aos quais se associam padrões de especificidade definidos por uma técnica de consenso mista baseada nas prescrições do Termo de Referência dos ELB. O principal interesse é incentivar o amplo uso dessas ferramentas na avaliação em saúde como um ponto crucial para aperfeiçoamento e desenvolvimento do SUS.*

Palavras-chave *Metaavaliação, Programas de Saúde, Institucionalização da avaliação*

¹ Département de
Médecine Sociale
et Préventive,
Université de Montréal.
Pavillon 1420,
Mont-Royal, Montréal, CA.
zulmira.hartz@umontreal.ca

Este artigo coloca em destaque questões imperativas, éticas e profissionais que justificam a necessidade do uso de princípios e padrões em avaliação, permitindo nos posicionar em relação a algumas diretrizes para os estudos de metaavaliação dos programas de saúde, tomando como exemplo o termo de referência elaborado para os Estudos de Linha de Base do Proesf¹. Tais balizamentos nos parecem particularmente importantes nos países em que a Política Nacional de Avaliação ainda está em estágio incipiente, como é o caso do Brasil.

Apesar de os padrões avaliativos constituírem a base da metaavaliação, incluímos a questão dos “princípios e diretrizes”, considerando que desempenham papéis coadjuvantes ou similares na institucionalização da avaliação em alguns países, sobretudo na Comunidade Européia². Embora o sentido destas palavras possa ser intercambiável e/ou subentendido na expressão portuguesa *parâmetros*, concordamos com Picciotto³ que tais expressões possam ter significado distinto: *princípio* é uma proposição geralmente percebida como um anseio ou uma meta a ser alcançada, que serve de fundamento ou convicção para determinadas ações; *padrão* é o termo preferido para normas obrigatórias, ou algo que serve como exemplo reconhecido, no qual ações são julgadas ou acompanhadas, conformando mecanismos de certificação; e *diretrizes* são freqüentemente enunciadas como regra geral ou recomendações.

Se a demanda dos governos pela avaliação de programas em países como Estados Unidos, Canadá e Inglaterra vem de várias décadas, a propensão em adotar a gestão por resultados, como prática corrente, influencia mundialmente o crescimento significativo do interesse pela avaliação, ainda que apresente uma diversidade de abordagens no seu controle de qualidade. Assim, nas sociedades nacionais e governos, é sintomático o interesse crescente direcionado à avaliação de programas, com a formulação de dispositivos legais, diretrizes políticas, princípios, regras de deontologia ou parâmetros específicos que buscam claramente balizar a prática da avaliação. Para Marceau⁴, parece tratar-se de uma resposta à demanda amplamente difundida da prática da avaliação, garantindo credibilidade ao seu reconhecimento.

Para Picciotto³, a gênese de padrões nacionais de avaliação, da mesma maneira que na auditoria ou contabilidade, faz parte do contexto social da profissão. Historicamente, é no setor educacional dos Estados Unidos que os

parâmetros de avaliação de programas públicos têm origem, através do Joint Committee on Standards for Educational Evaluation (JCEE). A primeira versão data de 1981 e uma reedição completa, em vigor desde 1994, vem sendo aplicada ao conjunto de políticas e programas sociais, inclusive aos programas de saúde⁵.

Muitas associações profissionais americanas e canadenses aderem a esses parâmetros que, entre outros, inspiraram fortemente a Sociedade Suíça de Avaliação, a Associação Africana e a Sociedade Alemã de Avaliação. Quatro grandes princípios norteiam seus enunciados:

- 1) *utilidade* – atende às necessidades de informação dos usuários;
- 2) *exequibilidade ou factibilidade* – realista e moderada nos custos, de modo a justificar a sua realização;
- 3) *propriedade* – conduzida eticamente, com respeito ao bem-estar dos envolvidos;
- 4) *precisão ou acurácia* – divulga e transmite informações sobre o valor ou mérito dos programas avaliados com a devida validade.

Ainda que o JCEE constitua a referência primordial em matéria de parâmetros, a Sociedade Americana de Avaliação (AEA) desenvolveu durante as duas últimas décadas seus próprios princípios-diretrizes para os avaliadores: 1) a investigação sistemática e rigorosa; 2) a competência e, entre outras, a competência cultural; 3) a integridade/honestidade; 4) o respeito às pessoas; e 5) a responsabilidade em relação ao interesse público e coletivo, além do interesse das partes em questão. Esta última consideração foi registrada pela primeira vez como uma diretriz formal⁴.

A Sociedade Canadense de Avaliação (CES) elaborou também diretrizes sobre ética que visam à competência, à integridade e à imputabilidade dos avaliadores. Juntamente com os quatro padrões de avaliação originais, essas diretrizes foram incorporadas, ampliadas e operacionalizadas pelo Centro de Excelência em Avaliação do Secretariado do Tesouro federal (similar ao nosso Tribunal de Contas da União), atualmente disponíveis em relatórios recentes de avaliação dos programas governamentais^{6,7}.

A Sociedade Francesa de Avaliação empreendeu um processo regular de revisão de padrões internacionais visando construir o mais amplo consenso possível entre seus membros. Em 2003, aprovou sua Carta de Avaliação das Políticas e Programas Públicos, que consiste de um preâmbulo e seis princípios, ou diretrizes: pluralismo, independência, competência, res-

peito à integridade dos indivíduos, transparência e responsabilidade. Cada um destes princípios é brevemente descrito em documento da Societé Française de l'Évaluation, que pode ser consultado em seu sítio na internet ⁸.

Em contraste com a Suíça, Alemanha e França, o Reino Unido tem uma tradição mais longa de avaliação e pertence ao grupo europeu de países que atingiram um nível relativamente alto de institucionalização (inclusive com a criação da United Kingdom Evaluation Society – UKES, em 1994). Somente em 2003, porém, a UKES apresentou um “guia de boas práticas” (que inclui a auto-avaliação programática) cuja peculiaridade está na distinção de normas dirigidas aos avaliadores, profissionais e participantes. Os resultados de sua aplicação, entretanto, ainda não são conhecidos².

Picciotto³, ao se referir às semelhanças entre as normas nacionais de avaliação de sete países (Austrália e Nova Zelândia; Canadá, França, Alemanha, Suíça, Estados Unidos e Inglaterra), faz um alerta. Se, por um lado, todas se referem a uma dupla audiência (organizações que demandam, preparam, conduzem e usam as avaliações, bem como indivíduos que se interessam por elas ou que são por elas afetados), é igualmente patente a assimetria entre o elevado peso atribuído às responsabilidades dos avaliadores, comparando-as às obrigações que deveriam assumir gestores de programas, explicitado apenas pela Inglaterra.

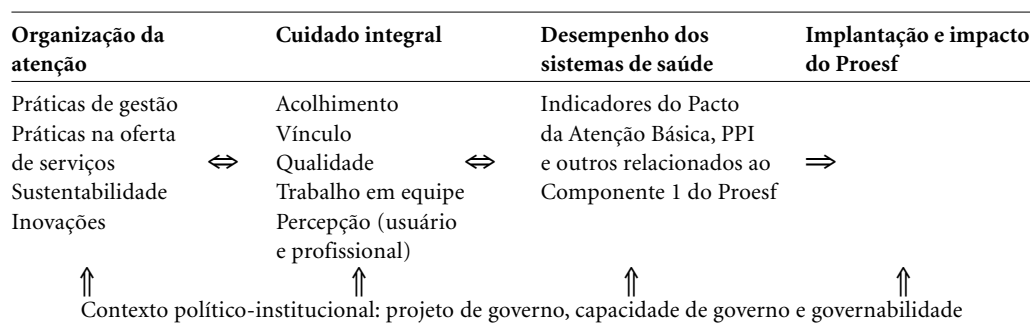
Em resumo, nesta breve revisão de normas e diretrizes nacionais de avaliação, gostaríamos de concordar com Widmer², que afirma que, se no conjunto, as distinções certamente resultam de processos culturais e participativos em cada comunidade de avaliação, as semelhanças dos princípios acordados sugere que a área atingiu

um alto grau de profissionalismo. A comparação de padrões entre as comunidades na União Européia, por exemplo, mostra que as semelhanças são bem maiores do que as pequenas diferenças de orientação (em nenhum caso os documentos se mostram inconsistentes ou contraditórios). Incentivar o mais amplo uso dessas ferramentas na prática da avaliação será um ponto crucial para seu aperfeiçoamento e desenvolvimento futuro.

Diretrizes para a metaavaliação do Proesf

A múltipla inserção institucional da avaliação dos Estudos de Linha de Base (ELB/Proesf), em se tratando de uma investigação demandada a grupos de pesquisa de instituições acadêmicas, selecionados por edital internacional elaborado em comum acordo com financiadores internos e externos (Ministério da Saúde e Banco Mundial), reforça a necessidade de se assegurar a qualidade e utilização do produto apresentado. O marco teórico do Termo de Referência (TR) proposto no edital (Figura 1) constitui claramente um avanço conceitual e um desafio para os metaavaliadores ao incorporar, em seu modelo, dimensões analíticas inter-relacionadas à implantação e ao impacto do programa, à organização e integralidade do cuidado, ao desempenho dos sistemas locais e ao contexto político. Além dessas características, os ELB são tratados como estratégia fundamental para a institucionalização do monitoramento e da avaliação municipal nas grandes cidades (mais de 100 mil habitantes), enfatizando-se a articulação e a indispensável colaboração entre o avaliador, o gestor, a academia e os serviços no

Figura 1
Marco analítico causal.



Fonte: Brasil, MS²².

apoio às atividades de gestão e de formação de avaliadores locais, através da participação, difusão e uso das investigações.

Se a complexidade da avaliação da gestão descentralizada em saúde, orientada pela atenção básica, tem se constituído em um objeto privilegiado da literatura nacional em avaliação, explorando diversos recortes e abordagens⁹, o mesmo não se observa no âmbito da metaavaliação¹⁰.

A metaavaliação – avaliação da avaliação – pode ser conceituada operacionalmente como um processo de descrição e julgamento de um estudo ou qualquer procedimento de avaliação, fundamentada em padrões e critérios propostos e validados pelas associações profissionais e outros organismos auditores. Ela consiste, portanto, na verificação da coerência teórica e prática da pesquisa com os padrões internacionais e/ou governamentais do controle de qualidade das avaliações de políticas públicas^{11, 12}.

Como qualquer avaliação de uma intervenção, a metaavaliação também deve ser introduzida preferencialmente com a pesquisa avaliada ainda em curso, de modo a poder contribuir para o aperfeiçoamento do próprio estudo (metaavaliação formativa) e não apenas como metaavaliação somativa, ao final da investigação do objeto de análise, quando poderá agregar maior credibilidade, mas não maior validade e utilidade aos resultados apresentados. Assim, tanto os avaliadores quanto os clientes se beneficiam desta dupla apreciação, favorecendo ainda o controle de vieses.

Se os padrões originais de avaliação alcançaram um elevado consenso no conjunto das Américas para os diferentes setores de políticas sociais, inclusive no campo da saúde e das intervenções comunitárias¹³, eles não devem ser considerados uma receita no seu enunciado ou aplicação e ser alvo de negociações e escolhas ao serem adaptados pelo “metaavaliador”. Em uma perspectiva internacional, a noção da necessidade de “padrões abertos” está sendo desenvolvida não só por causa das dificuldades inerentes à transferência de parâmetros entre diferentes culturas e contextos¹⁴, mas também das políticas e programas avaliados.

Nesse sentido, julgamos que o desenvolvimento de padrões relacionados ao adequado tratamento da “especificidade das intervenções” nas pesquisas avaliativas (*specificity standards*), como proposto no estudo exploratório de metaavaliação das intervenções comunitárias de promoção da saúde nas Américas¹³, nos

parece uma pista promissora a ser explorada em outras ações setoriais.

Esta abordagem se inspirou em estudos de metaavaliação de natureza formativa, propostos pelas sociedades profissionais interessadas em aumentar a qualidade e credibilidade dos estudos de avaliação e que focalizam, em formato agregado, um conjunto de relatórios e/ou publicações de um determinado campo de ação¹⁵. Ela se constituiu, igualmente, pelo esforço colaborativo dos metaavaliadores na construção de um consenso (ainda que provisório) sobre terminologia e definições que se configuravam como áreas críticas de aplicação para se verificar o respeito à coerência teórica dos princípios da avaliação em promoção da saúde¹⁶.

Concordando com Stufflebeam¹² que o principal desafio de uma metaavaliação é verificar até que ponto os estudos analisados respondem aos pré-requisitos de uma adequada avaliação (mérito) e à necessidade de informações dos interessados (valor), e considerando-se a inexistência no Brasil de normas nacionais (associativas ou governamentais), torna-se fundamental explicitar quais os princípios que parecem facilitar a construção e uso dos padrões de qualidade para a metaavaliação dos ELB, a partir das referências aqui apresentadas.

Inicialmente é preciso destacar o caráter formativo dessa metaavaliação, levando em consideração a não completude peculiar aos estudos de “linha de base”, e nosso interesse em qualificar o trabalho realizado com vistas a sua melhor utilização e aos necessários ajustes para avaliação somativa posterior. A idéia central é trabalharmos de forma agregada com uma amostra representativa do conjunto de estudos, cujas diretrizes se estruturam em duas etapas.

1. *Construção dos instrumentos e plano de análise da metaavaliação, sob a responsabilidade dos avaliadores externos nacionais e internacionais, incluindo a participação do grupo de acompanhamento e coordenadores dos ELBs.*

O ponto de partida nos parece consensual, ou seja, a utilização dos quatro padrões tradicionais – utilidade, exequibilidade, propriedade, precisão ou acurácia – que têm servido de guia para o desenho dos estudos de avaliação de programas de prevenção e controle de agravos⁵. E também para programas comunitários¹⁷, já traduzidos e publicados no projeto nacional de institucionalização da avaliação do programa nacional de controle da Aids¹⁸. Para agregar uma maior validade de conteúdo aos diversos padrões, sugere-se a utilização do *checklist*

de verificação e valoração de escores propostos por Stufflebeam¹⁹, bem como a possibilidade de adaptação dos critérios utilizados para os mesmos padrões nas diretrizes de metaavaliação da análise de implantação da gestão descentralizada em saúde no estado da Bahia^{9,10}

(Quadro 1). Para a questão da “especificidade”, recomendaríamos o uso de uma técnica mista de consenso²⁰ orientada pelas prescrições do Termo de Referência dos ELB.

2. *Revisão sistemática e independente dos relatórios dos ELB com a pontuação dos critérios pa-*

Quadro 1

Critérios complementares para metaavaliação dos programas de saúde.

Parâmetros da exeqüibilidade dos projetos de pesquisa	Atendimento aos preceitos da “ética” entre ciência e “boa governança”	Incorporação de princípios fundamentais em aconselhamento científicos	Adequada “análise de contribuição” das intervenções de saúde focalizadas	Propõe mecanismos que facilitem a apropriação das “lições aprendidas”
<ul style="list-style-type: none"> • É de interesse local, mas em coerência com prioridades nacionais. • As questões formuladas só podem ser respondidas por uma pesquisa. • Desenvolve modelos de avaliação para programas específicos incluindo suporte para elaboração de guias metodológicos. • Serve para implantar e/ou validar sistemas de informações. • Tem valor de demonstração, com poder de implicar a reestruturação organizacional de outros locais ou regiões. 	<ul style="list-style-type: none"> • A ciência inclui a totalidade do espectro de atividades, da geração do novo conhecimento às suas múltiplas aplicações. Inclui ainda o exame sistemático dos elementos naturais e sociais de nosso mundo, e a tradução de ganhos além de seus limites. Ela é crítica para um bom governo, mas raramente suficiente. A governança abarca todas as atividades governamentais que afetam a sociedade, como a aprovação de leis, formulação das políticas e regulações (por exemplo, avaliação), bem como o aconselhamento para cidadãos. 	<ul style="list-style-type: none"> • Antecipação de questões que surgem dos avanços do conhecimento. • Inclusão de várias fontes e disciplinas para capturar a diversidade de desafios, oportunidades e opiniões científicas. • Solidez para assegurar a qualidade, integridade e objetividade das recomendações. • Incerteza assumida pelo reconhecimento explícito e comunicação de limites. • Transparência e franqueza em discussões abertas sobre os problemas detectados, buscando opiniões de especialistas e acesso fácil ao processo de aconselhamento. • Revisão e reformulação de decisões baseadas em argumentos científicos para assegurar que estes reflitam os conhecimentos mais recentes. 	<ul style="list-style-type: none"> • Coleta de informações que agregam credibilidade sobre a contribuição da intervenção considerando as medidas dos resultados ou efeitos. • Inclui tanto medidas contínuas (monitoramento) como ocasionais, que respondam às questões da avaliação. • Reconhece que produtos ou efeitos observados são, por definição, influenciados não somente pela ação dos programas, mas também por fatores externos – outros programas, assim como fatores socioeconômicos. Desenvolve uma cadeia de resultados mais detalhada, que mostra como um resultado intermediário leva a um outro. • Localiza variações do programa e seus impactos, podendo-se identificar diferenças no programa (no tempo, entre localidades e entre diferentes grupos-alvo), recolhendo dados relativos às diferenças correspondentes. Realiza estudos de casos, examinando um ou alguns casos específicos dentro do programa que possam prover evidência confirmativa. 	<p>Nas conclusões da metaavaliação são apresentadas e discutidas condições facilitadoras para implementar a utilização dos resultados, que possam ser apropriados pelos atores das intervenções avaliadas. O pressuposto, em coerência com as diretrizes já abordadas de “ciência e boa governança”, é que um bom governo tenha a capacidade interna e colaborativa requerida para assimilar, interpretar e extrapolar o conhecimento obtido para suas necessidades.</p>

dronizados, complementando-se o aprofundamento da informação com estudos de casos exemplares²¹.

A avaliação, como qualquer outra ação social, deve primar pela coerência e consistência entre princípios, padrões e diretrizes profissionais e governamentais das políticas públicas. Assim a metaavaliação dos Estudos de Linha de Base do Proesf deveria ser aprofundada e requalificada com a adaptação e aplicação das diretrizes utilizadas pelo governo canadense⁷, abrindo-se também à possibilidade de observação do desejável diálogo com outras iniciati-

vas de institucionalização da avaliação da atenção básica de saúde nos municípios estudados e respectivas unidades federadas. Esta multiplicidade de focos, da maior importância ao se modelar a influência e utilidade das avaliações, nos obrigaria a ampliar o desenho metodológico, incorporando nos estudos de caso exemplares abordagens participativas e auto-avaliativas de gestores, avaliadores e outros atores. Essas abordagens, complementando o espectro de informações, iluminariam contribuições de maior complexidade e relevância das avaliações efetuadas, principal escopo de nossa proposição.

Referências

1. Brasil. Ministério da Saúde. Projeto de Expansão e Consolidação do Saúde da Família Proesf (Componente III) Termo de Referência para o sub componente D Linha de Ação 3 – Estudos de Linha de Base “Desenvolvimento de Estudos de Linha de Base nos Municípios Selecionados para o Componente 1” Disponível em www.saude.gov.br; 2004.
2. Widmer T. The development and status of evaluation standards in Western Europe. In Russon C, editor. International Perspective on evaluation standards. New Directions for Evaluation 2004; 104:31-54.
3. Picciotto R. The value of evaluation standards: a comparative assessment. JMDE 2005; 3:30-59.
4. Marceau R. Chartes, principes, lignes directrices, éthique ou standards; une réponse internationale diversifiée à des exigences de qualité des évaluations de programme. Bulletin SQEP 2005; 18:13-6.
5. Centers for Diseases Control. Framework for Program Evaluation in Public Health MMWR, 1999; 48 (RR-11).
6. Centre of Excellence for Evaluation, Treasury Board of Canada Secretariat. “Report on Effective Evaluation Practices”; 2004 [acesso em dezembro 2005]. Disponível em <http://www.tbs-sct.gc.ca>.
7. Centre of Excellence for Evaluation, Treasury Board of Canada Secretariat. “Review of the Quality of Evaluation across Departments and Agencies”; 2004 [acesso em dezembro 2005]. Disponível em <http://www.tbs-sct.gc.ca>.
8. Societé Française de l'Évaluation. Charte de l'évaluation des politiques publiques et des programmes publiques. 2003 [acesso em dezembro 2005]. Disponível em www.sfe.asso.fr
9. Vieira da Silva LM, Hartz ZMA, Chaves SCL, Silva GAP e Paim JS. Análise da implantação da gestão descentralizada em saúde: estudo comparado de cinco casos na Bahia-Brasil. Cad Saúde Pública 2006 (aceito para publicação).
10. Hartz ZMA. Significado, validade e limites do estudo de avaliação da descentralização da saúde na Bahia: uma metaavaliação. Anais Congresso da Abrasco; 2003.
11. Stufflebeam, DL. The metaevaluation imperative. AJE 2001; 2(2):183-209.
12. Whorthen, BR. Evaluating evaluations. In: Evaluation: alternative approaches and practical guidelines. New York: Longman Publishers; 1997.
13. Hartz ZMA, Goldberg C, Figueiro AC, Potvin L. Multi strategy evaluation in health promotion community interventions: an exploratory metaevaluation of American studies published from 2000 to 2005. (Aceito para publicação em 2006).
14. Love A, Russon C. Evaluation standards in an international context. New Directions for Evaluation 2004; 104(winter):5-14.
15. Stufflebeam DL. A note on purposes, development and applicability of the Joint Committee Evaluation Standards. AJE 2004; 25(1):99-102.
16. Goodstadt M, Hyndman B, McQueen DV, Potvin L, Rootman I, Springett J. Evaluation in health promotion: synthesis and recommendations. In: I Rootman, M Goodstadt, B Hyndman, D. McQueen, L Potvin, J Springett, E Ziglio, editores. Evaluation in health promotion. Principles and perspectives; 2001. Copenhagen: WHO Regional Publications. (European Series, 92).
17. The Center for the Advancement of Community Based Public Health (CBPH). An Evaluation framework for community health programs; 2000 [acesso em maio 2005]. Disponível em www.cbph.org.
18. Santos EM, Natal S. Série Ensinando Avaliação. Rio de Janeiro: Fiocruz; 2006. (No prelo)
19. Stufflebeam D. Metaevaluation checklist; 1999 [acesso em dezembro 2004]. Disponível em www.umich.edu.
20. Fernandes LEP, Vieira da Silva LM, Hartz ZMA. Conferência de Consenso sobre a Imagem-Objetivo da Descentralização da Atenção a Saúde no Brasil. In: Hartz ZMA, Vieira-da-Silva LM, organizadores. Avaliação em saúde: dos modelos teóricos à prática na avaliação de programas e sistemas de saúde. Salvador: EDUFBA; Rio de Janeiro: Fiocruz; 2005.
21. Yin, R. Estudos de caso (planejamento e métodos). Porto Alegre: Bookman; 2003.
22. Brasil. Ministério da Saúde/SAS/DAB/CAAB. Planos Estaduais para o fortalecimento das Ações de Monitoramento e Avaliação da Atenção Básica (diretrizes e orientações). Brasília; 2004.

Artigo apresentado em 27/02/2006

Aprovado em 31/03/2006

Versão final apresentada em 17/04/2006