

Meta-evaluation of baseline studies of the Brazilian Family Health Strategy Expansion Project: a participatory and formative approach

Meta-avaliação dos estudos da linha de base do Projeto de Expansão da Estratégia Saúde da Família no Brasil: uma abordagem formativa e participativa

Ana Cláudia Figueiró ¹
Zulmira Maria de Araújo Hartz ¹
Luiz Claudio Santos Thuler ²
André Luiz Freitas Dias ³

Abstract

A participatory, formative meta-evaluation of baseline studies in Brazil is presented. International standards recommended by associations of evaluators were used, along with "specificity" criteria built up using the terms of reference for proposals for the selection of studies. The methodological approach combined a "peer review" of baseline study reports, with a participatory (self) assessment for "primary" evaluators, the average of which provided the final score. Results revealed a classification of "good" and "very good" for the set of standards. The differences between the attribution of scores further highlight the importance of taking into account multiple points of view. Given the lack of pre-existing standards for the reports, the absence of standards and the incipient nature of evaluation focusing on utility, this meta-evaluation does not adequately reflect the quality or potential utility of the baseline studies, however, it will certainly contribute to overcoming these limitations and improving future impact studies of the Brazilian Family Health Strategy Expansion Project (PROESF).

Family Health Program; Health Evaluation; Evaluation Studies

Introduction

The Brazilian health system is currently facing the challenge of consolidating its guidelines without jeopardizing its feasibility, in a social and health context that generates increasing problems and needs for the population, and in light of scarce resources, despite the rapid and diverse technological progresses in the sector. A main focus of interest for managers in confronting these challenges lies in the field of primary health care, with the Family Health Program (FHP). As such, the Project for the Expansion of the Family Health Strategy (PROESF) has been developed with the World Bank's support, with the aim of achieving, by 2010, an overall coverage of 60% in the group of 231 municipalities with over 100,000 inhabitants (total population estimated at 90.1 million individuals), up from an average coverage of 22% in 2003. This effort, which was implemented in three stages over eight years, was intended to ensure the reorganization of local systems, including improvements in the work process and in the performance of services. Investments included institutional modernization activities, improvements to the health care network, human resource development, strengthening of information systems, and monitoring and evaluation ^{1,2}. In the implantation phase, PROESF resources corresponded to R\$ 147 million approximately, of which R\$ 13 million have been allocated in baseline studies,

¹ Instituto de Medicina Integral Prof. Fernando Figueira, Recife, Brasil.
² Coordenação de Ensino e Divulgação Científica, Instituto Nacional de Câncer, Rio de Janeiro, Brasil.

³ Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

Correspondence

A. C. Figueiró
Instituto de Medicina Integral Prof. Fernando Figueira,
Rua dos Coelhos 300, Recife,
PE 50070-550, Brasil.
anaclaudiafigueiro@yahoo.com.br

an amount that is close to 10% of investments recommended by international standards for evaluation^{3,4}.

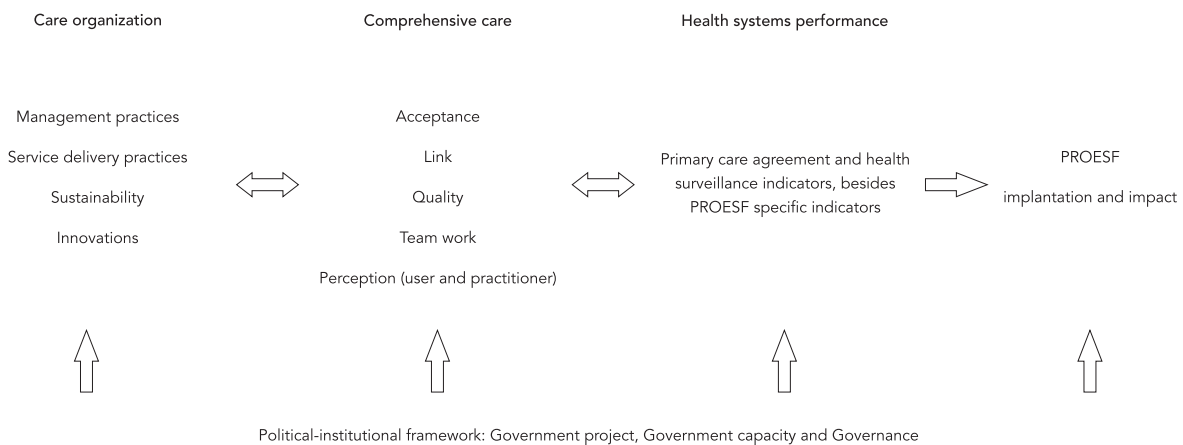
The agreement between the Brazilian government and the World Bank for the implementation of PROESF established a set of technical requirements and, among them, that baseline studies were carried out in the municipalities involved. These studies sought to characterize FHP in the initial phase of the project, for a future impact study, as well as to estimate the performance of primary health care for the group of municipalities and the observed differences in the areas with and without FHP. For primary health care policy makers, the process implemented in the evaluation of PROESF represented an opportunity to adopt innovations proposed in the policy being constructed, such as broad engagement of all stakeholders, capacity building in evaluation for individuals involved in PROESF implementation in the target municipalities, and the integration of academic work and service delivery. Baseline studies were carried out between 2005 and 2006 by eight research institutions, with recognized academic experience in the field of health management and evaluation, which were selected through an international bidding and selection process, grouping the municipalities into 14 contractual parts, according to population size, regional proximity, and financial threshold for the proportional allocation of resources. The

analytical plan for the baseline studies was to be organized by the institutional groups according to the theoretical model (Figure 1) and a criteria matrix of its various dimensions (political-institutional, care organization, comprehensive care, and health systems performance), as defined in the terms of reference of the call for proposals. The overall alignment of studies (including database review and outputs displayed) was made under the ongoing supervision of a multidisciplinary technical and scientific follow-up group, comprised of four Ministry of Health consultants and technical staff members responsible for monitoring and evaluating primary health care, with the participation of practitioners from the research institutions and municipal health care managers^{5,6}.

The multiple institutional creation of baseline studies – when dealing with an investigation commissioned to research groups from academic institutions, that have been selected by means of an international selection process organized in agreement with internal and external funders (the Ministry of Health and World Bank) – reinforced the interest to ensure the quality and utility of the output displayed, thus leading to the decision to carry out this meta-evaluation, “instituted” by the PROESF national manager. As such it is considered an *instituted evaluation*, meaning a “socially organized” evaluation that involves three components: use of a well-defined

Figure 1

Theoretical model as analytical landmark for the baseline studies.



PROESF: Brazilian Family Health Strategy Expansion Project. Source: Ministério da Saúde¹.

methodology and instruments; socially authorized persons who conduct it; and formalized results to be used ^{7,8,9}.

International standards for meta-evaluation studies

Operationally, meta-evaluation can be defined as a normative process of investigation, judgement and synthesis of a study or any evaluation procedure. It therefore differs from meta-analysis studies that seek, using statistical procedures, to summarize the results of different studies by testing the same hypothesis in one single estimation ¹⁰. Meta-evaluation therefore consists in checking the theoretical, practical and ethical consistency of research with international and/or governmental standards of quality control of public policy evaluations ^{11,12,13,14}. This apparently simple definition masks the complexity of meta-evaluation studies as the reflection and investigation of its academic and functional practices. As in the evaluation of any intervention, meta-evaluation should preferably be introduced when the evaluative research is still in course (formative meta-evaluation), so as to contribute to study improvement, and not as a summative meta-evaluation, at the end of the investigation, when it can add liability, but not more validity and utility, to the outcomes. In the case of baseline studies, as phase one of the PROESF evaluation process, the formative and participatory nature of the meta-evaluation is intended to contribute to its improvement, so that both evaluators and customers can take advantage of this double assessment, as the reflective and investigative action of academic and functional practices, also promoting bias control. The term “participatory” is used herein in the sense employed by Baron & Monier ¹⁵ of co-produced/pluralist fourth generation evaluation.

Likewise evaluation, meta-evaluation is expected to specify its theoretical reference, allowing the various audiences to follow the steps of researchers, in order to discuss their findings and judgments. In this sense, the choice of evaluation standards from public programs of the 1970s and 1980s that were originally set in the US educational sector by the Joint Committee on Standards for Educational Evaluation (JCEE), is justified by the use of these standards to guide the actions of several professional communities of evaluators (in the United States, Canada, Europe, Africa, and so forth); the support to this theoretical reference by international agencies concerned with the financing and evaluation of development aid programs ¹⁶, such as PROESF; and, as mentioned by Guba & Lincoln ¹⁷, the fact that its application

would not go against the purposes of a fourth generation evaluation.

Four major principles or attributes guide its enunciations: utility, feasibility (practicability), propriety and accuracy (precision). Each one of these parameters is built upon the averages of their respective criteria (Table 1), ranging from 0 to 10, and ranked according to the scores they achieve, as follows ¹⁸: excellent (9.0-10.0); very good (7.0-8.9); good (5.0-6.9); weak (3.0-4.9) and poor (critical; < 3.0).

Given the lack of national rules (either associative or governmental), the above-mentioned principles and references – utility, feasibility, propriety and accuracy or precision – mark out, but do not restrain, the standards used in this meta-evaluation, whose “formative nature” is inescapable, given the incompleteness of baseline studies, assuming necessary adjustments for a (summative) impact analysis of PROESF, and the formal interest in deepening and qualifying the work developed so far, thus promoting its use by the organizations and institutions involved.

If the original evaluation standards have achieved a high consensus in the Americas, for the different sectors of social policies, including the fields of health and community interventions ^{8,19}, they should not be regarded as a blueprint in their enunciation and application, meaning that negotiations and decision-making are required in their adaptation by each “meta-evaluator”. Internationally, the idea is being developed that there is a need for “open standards” due to the difficulties inherent to the transfer of parameters among different cultures and backgrounds ²⁰, but also to the complexity of the policies and programs being evaluated. Despite the recognition of this necessary adaptation, the development of specific standards is not yet a practice. In this sense, Hartz et al. ²¹, in the meta-evaluation of community interventions for health promotion in the Americas, proposed a fifth guiding principle, related to the proper treatment of the “specificity” of interventions (*specificity standards*). This innovative nature of meta-evaluation studies was added to the present project, understanding that the specificity of an intervention is rooted in the theoretical grounding underlying its potential action, a condition that cannot be dissociated from the relevance and liability of answers given in the evaluation research, which would support the assumption, adopted herein, of more (or less) utility of its outcomes for decision-makers ^{22,23}.

Based on this assumption, the following questions have been defined as the object of this paper, regarding the meta-evaluation of the evaluative research called baseline studies of

Table 1

Meta-evaluation international standards.

Standards	Criteria
Utility	Stakeholder identification
	Evaluator liability
	Information scope and selection
	Report clearness
	Timely delivery and dissemination of the outcomes
	Evaluation impact
Feasibility or practicability	Cost-effectiveness
	Practical procedures
	Political feasibility
Propriety	Service oriented
	Formal agreements
	People's rights
	Human relationships
	Complete and fair evaluation
	Conflict of interests
Accuracy or precision	Fiscal accountability
	Program documentation
	Context analysis
	Description of purposes and procedures
	Reliable information sources
	Valid information
	Reliable information
	Systematic information
	Quantitative information analysis
	Qualitative information analysis
	Justifiable conclusions
	Report impartiality
Meta-evaluation	

PROESF: What was the classification obtained from the reports that have been analyzed in relation to meta-evaluation international standards? To what extent did baseline studies meet the specificity standards of the intervention, in terms of their theoretical and operational consistency with the terms of reference model? How did these results vary depending on the different positions of meta-evaluators?

Methodology

Any activity involving reflective evaluation should be developed using a participatory approach oriented to foster the use of study findings^{15,24}. It allows the various stakeholders to (re) build the judgment of the (meta) "evaluand", that has been problematized and (re)qualified in its

construction, in accordance with the notice of selection of research groups, which advocated a participatory orientation to carry out baseline studies, whose reports are object of the present meta-evaluation.

The original study sample consisted of all municipal reports (24) concluded at the time the meta-evaluation begun and submitted by the eight research institutions. Given the homogeneity of evaluation approaches inside investigation groups and time constraints to carry out the study (3 months), the selection criterion used was to include the municipal report of higher population size by groups of investigators, amounting to eight reports.

The methodological approach combined a "peer review" procedure, in two independent readings (an external meta-evaluator and the other one belonging to the follow-up group),

with a process of (self) qualification by “primary” evaluators (research group coordinators). The average score gave the final classification of the eight reports, corresponding to a unitary sampling of each evaluation team selected for the study.

As mentioned before, the classification tool took into account the panel of checklist criteria and valuation proposed by Stufflebeam¹⁸ for the assessment of the four international quality standards of evaluation studies: utility, feasibility, propriety, and accuracy. This choice is sectorally supported by the fact that they have been widely disclosed (and in open access) over the last decade for the evaluation of disease control and prevention programs or in health promotion community programs^{19,25}, already translated and published as educational material for the formation of evaluators²⁶. The fifth standard, related to the project specificity, was assessed according to three criteria: consistency with the theoretical model; multidimensional and relational analysis of results, having the baseline studies terms of reference as “gold standard”¹; and the same scale of values of the remaining standards.

The testing of tools revealed that the following standards should be taken into account, with maximum scoring (10 points): evaluator liability; timely delivery and dissemination of the outcomes; cost-effectiveness; formal agreements; people’s rights; human relationships; conflict of interests; and fiscal accountability. This decision was made considering that these criteria were treated adequately in the international tender process, in the ethics committees of the selected research institutions, and checked by the external committee for follow-up of the work plans and project accomplishment⁶.

Main results and discussion

Table 2 shows the results of the meta-evaluation, grouped according to the five standards (average and differentiated values for the three groups of meta-evaluators), whose answers to the three questions can be summarized in the findings to follow. The highest scores given by external evaluators to parameters related to baseline studies being carried out within legal and ethical rules (propriety), and to the accountability towards the proper use of resources and meeting the deadlines (feasibility) highlight the liability and responsibility of the groups selected for the studies and the support given by experts in their follow-up. Studies oriented by terms of reference, expressed in the standard “specificity”, question two, was the only criterion rated as “poor” by the two groups of external auditors (3.3. and 4.5). Besides, (self) qualification scores were almost always higher than those given by the auditors, while specificity’ was rated as “excellent” (9.2) only once by the auditors of the research groups. The main purpose of the following discussion, interacting with the international literature, is to increase the understanding and contextualization of differentials observed in the composition of standard scores, detailing them in light of existing or inescapable constraints in the meta-evaluation.

Beginning with utility, even though local actors had been identified in a workshop that was a contractual requirement to start the research, we observed that it did not imply the concern with identifying their own information needs, nor the different interests and points of view of stakeholders; only the federal manager explained his needs through the terms of reference. Given that the participation of individuals and groups with interest in the evaluation process is essential to promote its utility, even justifying the study being carried out, the low scores for this parameter call

Table 2

Final scores of baseline studies – Brazilian Family Health Strategy Expansion Project (PROESF) meta-evaluation.

Parameters	Research groups	Follow-up group	External evaluators	Total average
Utility	8.9	5.9	6.3	7.1
Feasibility	8.5	6.3	7.1	7.3
Propriety	8.6	6.6	8.1	7.8
Accuracy	8.9	5.1	5.5	6.5
Specificity	9.2	3.3	4.5	5.7

Note: excellent (9.0-10.0); very good (7.0-8.9); good (5.0-6.9); weak (3.0-4.9); poor/critical (< 3.0).

the attention to a reduced potential use of baseline studies by municipal managers.

Returning to our initial assumptions, criteria informing us about the relevance or not of the evaluation, that is, its ability to provide answers to problems facing decision-makers, were identified in the standard “utility”²². To meet the concrete needs of decision-makers and other stakeholders requires an evaluation practice that takes into account their backgrounds and points of view. As suggested by various studies^{27,28,29}, the utility of evaluations is largely determined by the perception of stakeholders about the study and its outcomes, and such perception derives from background factors and the nature of the relationship between evaluators and stakeholders. In this way, the stakeholders expect to recognize their ideas and experiences reflected in the evaluations, which supposedly they will be using in future.

Accuracy or precision represent a key standard for expressing the methodological rigor and, consequently, the liability of an evaluation. From this point of view, the design of studies in general, according to data available in the research reports, shared an inconsistency with the terms of reference and the very focus of interventions: the nature and governance of the municipality were pushed to the background whenever the site was repeatedly reduced as an observation unit, focusing the analysis on “operational lots”. Another aspect to be highlighted in this parameter is the omission and/or low importance regarding the description of the local program to be implemented by PROESE, even though numerous studies already published about FHP show a great diversity of organizational modalities in the different municipal backgrounds. Reports also indicated problems in handling data on beneficiaries and stakeholders, seen as “information sources” rather than as partners in the evaluation process, which is extremely important for the use of study outcomes. This constraint was reflected in the assessment of the criterion “justifiable conclusions”, which falls short not only from the lack of points of view from stakeholders, but also from partial approaches for the expected dimensions, sometimes revealing the failure of the methodological option to understand the program’s completeness. A prevalent risk in evaluation studies, as pointed out by other authors^{27,30}, refers to the choice of the study design by the evaluator first according to his (or her) expertise, giving less importance to issues raised by the intervention and by stakeholders, which should define the methodological options. On the other hand, other good quality evaluations did not succeed precisely because they neglected

to acknowledge the influences of interpersonal, ethical and political factors, all of which guide the work of the evaluator³¹.

The specificity issue, based on the terms of reference intervention logic model, characterizes theory-driven evaluations, which require, for the assessment of study results, an analysis of the program inferences in their causal interactions, in addition to ethical, political and methodological considerations observed in the remaining parameters. The program theory provides not only a guide to analyze the phenomenon, but also a framework to understand the meaning of the research findings³². Along with relevance and liability, this theoretical grounding, taken as one of the assumptions of our study, affects the use of evaluation results for decision-making²². If “specificity” had the lowest scoring in the independent readings of the reports, this was due to a low expression of the proposed model, which did not relate the findings of the investigated dimensions, their use resembling a “static” support, rather than presumably interdependent analytical categories. However, it is worth remembering that a similar result was observed in the meta-evaluation exploratory study of health promotion community interventions in the Americas²¹, in which only 52% of cases were classified as “good or very good”, in contrast to accuracy, which achieved this classification in 80% in the studies, suggesting that the concern with methodological rigor prevailed over the complexity of the treated objects.

The reason for the differences observed in the meta-evaluation between the participating groups was not investigated, but it can be suggested that one of the factors lies in one of the limits of this meta-evaluation study, which only used the final reports from baseline studies. The by-products of these baseline studies, such as journal articles and empowerment material, which might contain aspects missing in the final reports, could not be assessed in this first meta-evaluation. Another aspect has to do with the lack of previous orientation to research groups for the standardization of reports, with focus on the specificity and utility for decision-makers, especially at the local level. In this sense, the reports seem to serve much more to study funders and central level governance, explaining the conflict of interests between national versus local, a permanent challenge for any evaluator of social policies²⁸.

Another constraint in the search for evidences for interventions regarding complex problems such as baseline studies, which might have also interfered in their accuracy or precision, is inherent to normative analyses, as is the case of the ac-

tion matrix foreseen in the terms of reference for the operationalization of analytical dimensions of the logic model. According to Potvin¹⁰, this occurred because of the large number of arbitrary decisions concerning comparable results to be summarized, and how to weigh these decisions in order to reach the conclusions. This represents another challenge for the (meta) evaluation, in settings where it has been carried out with different expectations, in which there is a counteraction between project components or dimensions that are difficult to grasp in terms of extension and depth.

On the other hand, if it is likely that the reports do not express all the richness of the material collected and analyzed for the qualification of studies, the meta-evaluation, understood not as an end in itself, but as part of an open dialogue between the many actors in this process, should contribute to organizational learning, thus overcoming a few communication barriers and promoting the use of baseline study results³³.

Final considerations

The complexity of the intervention, represented in the terms of reference of baseline studies, required that investigators adopted multiple evaluation perspectives and strategies to understand the various dimensions of the program. Facing this challenge with the necessary consistency perhaps required some more time, resources and support than those available. If the constraints for the studies being carried out did not jeopardized their liability, in the desirable prospect of going back to the municipalities to present baseline studies results and/or new evaluative research, they lack more adequacy of analytical procedures for an overall and multidimensional understanding of their object, as exemplified in the outspreads of this meta-evaluation of case studies carried out by Pontes-da-Silva & Figueiró³⁴ and in the recommendations for new terms of reference for evaluating PROESF impact, addressed by Felisberto et al.².

Reiterating the formative nature of this meta-evaluation, our major commitment has been with the learning process it provided and the use of these lessons. Therefore, in terms of the quality of this experiment we understand this to mean not only the use of normative parameters, but also the awareness of the learning deriving from it, as well as the resultant and effective changes implied in the common project of evaluation institutionalization, or rather, acculturation. Such awareness requires that we, (meta) evaluators, be the first users of its results, as apprentices that

we are of each study carried out. Thus, for example, Medina & Fernandes³⁵, while investigating the development of the field of evaluations in other countries, observed, in the case of Canada, progresses obtained from the identification and correction of the main problems in the development of studies (such as quality improvement of reports after standardization and definition of systematic criteria for their assessment) that jeopardized the "evaluation health" in governmental policies³⁶.

Regarding the institutional status of the evaluation, the sense employed herein, which also inspired and guided our study, is that we are aligned with the literature of the field of evaluation which moves towards increasingly participatory and democratic procedures. Differences found in the results of the different groups involved in baseline studies corroborate the importance of multiple points of view being taken into account, simultaneously allowing evaluators to understand and contribute to decision-making throughout the studies, in accordance with recognized quality parameters, from both an internal and external point of view. When we consider that the judgement of an intervention must not be the exclusive privilege of evaluators and their logics, but equally of other stakeholders, from and beyond the results of evaluations, this means that we are engaged with effective participation. It therefore ensures the primacy of our interests and needs in defining the problems and objectives that will guide the evaluations, the inclusion of our values in the judgement criteria that have been agreed upon, and consequently more useful findings, legitimizing and fulfilling our mission to have a positive influence on health policies.

Resumo

O trabalho apresenta os resultados da meta-avaliação formativa e participativa dos estudos de linha de base no Brasil. Utilizaram-se padrões internacionais preconizados pelas associações de avaliadores e critérios de especificidade construídos usando o termo de referência para seleção de propostas de estudo. A abordagem metodológica associou "revisão de pares" dos relatórios municipais dos estudos de linha de base, com um processo de participação e (auto) qualificação pelos avaliadores "primários", cuja média gerou a classificação final. Os resultados mostram uma classificação entre "bom" e "muito bom" para o conjunto dos parâmetros. As diferenças na atribuição de escores reforçam a importância de se garantirem múltiplos pontos de vista. Considerando a falta de padronização prévia dos relatórios, inexistência de padrões e incipiência das práticas avaliativas com foco na utilidade, provavelmente esta meta-avaliação não refletiu adequadamente a qualidade e uso potencial dos estudos, mas certamente contribuirá para aprimorar os estudos futuro de impacto do Projeto de Expansão da Estratégia Saúde da Família (PROESF).

Programa Saúde da Família; Avaliação em Saúde; Estudos de Avaliação

Contributors

A. C. Figueiró and Z. M. A. Hartz participated in all stages of producing the article. L. C. S. Thuler contributed to the data processing and analysis. A. L. F. Dias contributed towards the article's conception, and also participated in the revision of the parameters of the meta-evaluation.

Acknowledgments

To the Brazilian Ministry of Health, the Pan-American Health Organization and the Brazilian Post-Graduate Association in Collective Health (ABRASCO) for financial support.

References

1. Ministério da Saúde. Programa de Expansão e Consolidação da 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 – 2004. http://dtr2004.saude.gov.br/dab/docs/publicacoes/geral/tdr_linhas_de_base.pdf (acessado em 12/Mai/2009).
2. Felisberto E, Benevides IA, Hartz ZMA, Matos K. Estudos de impacto do PROESF: considerações e diretrizes preliminares necessárias à sua proposição. In: Hartz ZMA, Vieira-da-Silva LM, Felisberto E, organizadores. Meta-avaliação da atenção básica em saúde: teoria e prática. Rio de Janeiro: Editora Fiocruz; 2008. p. 369-98.
3. Scriven M. Meta-evaluation revisited. *Journal of MultiDisciplinary Evaluation* 2009; 6:iii-viii.
4. World Health Organization. National AIDS programmes: a guide to monitoring and evaluating HIV/AIDS care and support. Geneva: World Health Organization; 2004.
5. Felisberto E. Da teoria à formulação de uma Política Nacional de Avaliação em Saúde: reabrindo o debate. *Ciênc Saúde Coletiva* 2006; 11:553-63.
6. Bodstein R, Feliciano K, Hortale VA, Leal MC. Estudos de Linha de Base do Projeto de Expansão e Consolidação do Saúde da Família (ELB/PROESF): considerações sobre seu acompanhamento. *Ciênc Saúde Coletiva* 2006; 11:725-31.
7. Demarteau M. Theoretical framework and grid for analysis of programme – evaluation practices. *American Journal of Evaluation* 2002; 8:454-73.
8. Hartz ZMA. Princípios e padrões em meta-avaliação: diretrizes para os programas de Saúde. *Ciênc Saúde Coletiva* 2006; 11:733-8.

9. Hartz ZMA, Contandriopoulos AP. Do quê ao pra quê da meta-avaliação em saúde. In: Hartz ZMA, Vieira-da-Silva LM, Felisberto E, organizadores. Meta-avaliação da atenção básica em saúde: teoria e prática. Rio de Janeiro: Editora Fiocruz; 2008. p. 27-45.
10. Potvin L. Why we should be worried about evidence-based practice in health promotion. *Rev Bras Saúde Matern Infant* 2005; 5 Suppl 1:S93-7.
11. Stufflebeam DL. The metaevaluation imperative. *American Journal of Evaluation* 2001; 2:183-209.
12. Worthen BR, Sanders JR, Fitzpatrick JL. Program evaluation: alternative approaches and practical guidelines. 2th Ed. New York: Longman; 1997.
13. Worthen BR, Sanders JR, Fitzpatrick JL. Avaliação de programas: concepções e práticas. São Paulo: Editora Gente; 2004.
14. Scriven M. Evaluation thesaurus. 4th Ed. Newbury Park: Sage Publications; 1991.
15. Baron G, Monnier E. Une approche pluraliste et participative: coproduire l'évaluation avec la société civile. *Informations Sociales* 2003; 110:120-9.
16. Independent Evaluation Group. Sourcebook for evaluating global and regional partnership programs: indicative principles and standards – 2007. <http://www.worldbank.org/ieg/grpp> (accessed on 02/Dec/2008).
17. Guba EG, Lincoln YS. Judging the quality of fourth generation evaluation. In: Guba EG, Lincoln YS, editors. Fourth generation evaluation. Newbury Park: Sage Publications; 1989. p. 228-51.
18. Stufflebeam DL. Program evaluations metaevaluation checklist. http://www.wmich.edu/evalctr/checklists/program_metaeval.pdf (accessed on 15/Nov/2008).
19. Framework for program evaluation in public health. *MMWR Recomm Rep* 1999; 48(RR-11):1-40.
20. Love A, Russon C. Evaluation standards in an international context. In: Russon C, Russon G, editors. International perspectives on evaluation standards: new directions for evaluation. San Francisco: Jossey-Bass; 2004. p. 5-14.
21. Hartz ZMA, Goldberg C, Figueiró AC, Potvin L. Multi-strategy in the evaluation of health promotion community interventions: an indicator of quality. In: McQueen DV, Potvin L, editors. Health promotion evaluation practices in the Americas: values and research. New York: Springer; 2008. p. 253-67.
22. Contandriopoulos AP. Avaliando a institucionalização da avaliação. *Ciênc Saúde Coletiva* 2006; 11:705-12.
23. Contandriopoulos AP. Évaluer l'évaluation. In: Brousselle A, Champagne F, Contandriopoulos AP, Hartz ZMA, editors. L'évaluation: concepts et méthodes. Montréal: Les Presses de l'Université de Montréal; 2009. p. 273-82.
24. Cousins JB, Whitmore E. Framing participatory evaluation. *New Directions for Evaluation* 1998; (80):5-23.
25. Center for the Advancement of Community Based Public Health; Baker QE, Davis DA, Gallerani R, Sanchez V, Viadro C. An evaluation framework of community health programs. Durham: Center for Advancement of Community Based Public Health; 2000.
26. Santos EM, Natal S, Cruz MM. Dimensão técnico-operacional. Unidade didático-pedagógica: modelo teórico de avaliação. Rio de Janeiro: ABRASCO; 2006. (Série Ensinando Avaliação, 3).
27. Patton MQ. Utilization-focused evaluation: process and premises. In: Patton MQ, editor. Utilization focused evaluation: the news century text. 3th Ed. Thousand Oaks: Sage Publications; 1997. p. 371-85.
28. Van Der Meer FB, Edelenbos J. Evaluation in multi-actor policy processes. *Evaluation* 2006; 12:201-18.
29. Oliver ML. Metaevaluation as a means of examining evaluation influence. *Journal of MultiDisciplinary Evaluation* 2009; 6:32-7.
30. Weiss CH. The interface between evaluation and public policy. *Evaluation* 1999; 5:468-86.
31. Worthen BR, Sanders, JR. Educational evaluation: alternative approaches and practical guidelines. New York: Longman; 1987.
32. Chen HT. Method and theory in program evaluation: a question of balance. In: Chen HT, editor. Theory-driven evaluation. Thousand Oaks: Sage Publications; 1990. p. 17-35.
33. Uusikylä P, Virtanen P. Meta-evaluation as a tool for learning. *Evaluation* 2000; 6:50-65.
34. Pontes-da-Silva GA, Figueiró AC. Linha de base matricial da implementação do PROESE. In: Hartz ZMA, Vieira-da-Silva LM, Felisberto E, organizadores. Meta-avaliação da atenção básica em saúde: teoria e prática. Rio de Janeiro: Editora Fiocruz; 2008. p. 71-99.
35. Medina MG, Fernandes AS. Utilidade da avaliação para gestores de saúde: uma contribuição com base na análise dos relatórios dos Estudos de Linha de Base. In: Hartz ZMA, Vieira-da-Silva LM, Felisberto E, organizadores. Meta-avaliação da atenção básica em saúde: teoria e prática. Rio de Janeiro: Editora Fiocruz; 2008. p. 101-16.
36. Treasury Board of Canada Secretariat. Report on improving the professionalism of evaluation 2005. http://www.tbs-sct.gc.ca/eval/dev/Professionalism/profession_e.asp#tphp (accessed on 15/Nov/2008).

Submitted on 16/Dec/2009

Final version resubmitted on 18/Jun/2010

Approved on 29/Jun/2010