

## A avaliação de programas de saúde: continuidades e mudanças

Health programs assessment: continuities  
and changes

La evaluación de programas de salud:  
continuidades y cambios

Lígia Maria Vieira-da-Silva <sup>1</sup>  
Juarez Pereira Furtado <sup>2</sup>

doi: 10.1590/0102-311X00237219

### Introdução

A avaliação de programas de saúde tem sido desenvolvida como um conjunto de saberes e práticas com influência de diversas disciplinas científicas e tradições teórico-metodológicas. Tendo em vista o envolvimento de agentes oriundos de diversos campos sociais <sup>1</sup>, pode ser considerada um espaço intercampos, no sentido de Bourdieu <sup>2</sup>. As trajetórias e posições ocupadas pelos autores têm produzido historicamente diferentes tipos de abordagem e polissemia conceitual <sup>3,4,5</sup>. Há ênfases distintas nos aspectos teórico-metodológicos, na prestação de contas e no compromissos com a gestão, dentre outras questões, gerando debates duradouros e mesmo polarizações, sobretudo na produção especializada dos Estados Unidos, do Canadá e da Inglaterra, que concentram certas tradições e significativa produção prática e teórica da área <sup>6</sup>. A delimitação de seu escopo comporta algum grau de arbitrariedade, tendo em vista que políticas, serviços, tecnologias e práticas de saúde <sup>7,8</sup>, sob a denominação de avaliação de programas, têm sido tomados por objeto.

No Brasil, o desenvolvimento da avaliação em saúde ocorreu no interior do campo da Saúde Coletiva <sup>9</sup> e ali se beneficiou de diversas influências <sup>1</sup>. Trabalhos anteriores buscaram identificar convergências e diferenças entre os desenhos da pesquisa avaliativa e da epidemiologia <sup>10</sup>, a avaliação de programas e tecnologias <sup>11</sup>, entre as chamadas abordagens estruturadas e não estruturadas <sup>12</sup>, e também entre a avaliação e a análise de políticas <sup>13</sup>. Análises comparativas a esse respeito, entre Brasil e Estados Unidos, mostram o expressivo crescimento das publicações sobre avaliação em saúde a partir dos anos 2000 <sup>14</sup>, voltadas principalmente para a avaliação de políticas e programas do Sistema Único de Saúde (SUS) <sup>15</sup>.

Contudo, um aspecto pouco explorado tem sido a superposição de objetos, conceitos e metodologias que por vezes não estabelecem diálogo entre si. Também a persistência da diversidade de abordagens e modelos tem sido registrada. Em que medida tem havido uma incorporação das proposições teórico-metodológicas dos autores fundadores da disciplina nos trabalhos subsequentes? Em que medida esses modelos têm sido contestados e substituídos ou aperfeiçoados? O presente ensaio pretende contribuir com essas questões e oferecer uma melhor compreensão acerca do desenvolvimento do espaço da avaliação de programas de saúde, buscando atualizar o debate na área, por meio da análise e da discussão de algumas abordagens presentes na produção científica relacionada a essa temática.

<sup>1</sup> Instituto de Saúde Coletiva, Universidade Federal da Bahia, Salvador, Brasil.

<sup>2</sup> Departamento de Políticas Públicas e Saúde Coletiva, Universidade Federal de São Paulo, São Paulo, Brasil.

### Correspondência

L. M. Vieira-da-Silva  
Instituto de Saúde Coletiva,  
Universidade Federal da Bahia.  
Rua Basílio da Gama s/n,  
Salvador, BA 40110-040,  
Brasil.  
ligiamvs@ufba.br



## Origens da avaliação em saúde como espaço de saberes e práticas (1960-1980)

Tendo em vista que a avaliação de programas de saúde tem um desenvolvimento muito ligado àquele da avaliação como disciplina acadêmica, uma análise acerca dessa problemática requer algumas considerações preliminares sobre a evolução das relações entre a disciplina da avaliação e os programas de saúde como um dos objetos possíveis.

Nessa perspectiva, pode-se dizer que, na área da saúde, a avaliação passou a ser considerada objeto de estudo e pesquisa a partir de três vertentes que emergiram nos anos 1960. A primeira, oriunda da clínica, estava representada pelas contribuições de Avedis Donabedian<sup>16,17</sup>, que sistematizou as diversas abordagens propostas para avaliação da qualidade do cuidado médico. A segunda surgiu no interior da epidemiologia e passou a incorporar como objeto de análise a investigação em serviços de saúde, da qual a avaliação fazia parte<sup>18,19</sup>. Já a terceira vertente sofreu a influência das pesquisas avaliativas em educação e em psicologia, iniciadas a partir dos estudos de Campbell & Stanley<sup>20</sup> e desenvolvidas posteriormente por grupos multidisciplinares reunidos em torno de associações, a exemplo da Fundação Russell Sage (Russell Sage Foundation), em 1967, e da Sociedade de Pesquisa em Avaliação (Evaluation Research Society), em 1978. Entre esses primeiros autores estavam Peter Rossi (Rossi & Freeman<sup>21</sup>), Thomas Cook (Cook & Campbell<sup>22</sup>), Michael Scriven<sup>5</sup> e Carol Weiss<sup>4</sup>. Havia também pessoas vinculadas à burocracia governamental, além dos pesquisadores. Inicialmente voltado para o desenvolvimento de uma teoria da avaliação, logo o grupo dividiu-se entre os que compreendiam a avaliação como um guia para a gestão do setor público e aqueles que a concebiam como um recurso para a prestação de contas (*accountability*). Também se verificou uma especialização com a divisão do espaço dos estudos avaliativos entre disciplinas: educação, políticas públicas e gestão, economia, medicina e saúde pública<sup>23</sup>. Esta última vertente foi aquela em que a avaliação foi teorizada de forma mais sistemática e se desenvolveu como disciplina relativamente autônoma. Em coletânea denominada de raízes da avaliação, além dos autores acima mencionados como pioneiros ou fundadores, Christie e Alkin acrescentam Daniel Stufflebeam e mais duas dezenas de autores que, segundo eles, teriam contribuído com teorias ou modelos originais para a disciplina da avaliação<sup>24</sup>.

## Algumas tendências na evolução das abordagens para avaliação de programas de saúde (1997-2019)

Entre 1997 e 2000, surgiram internacionalmente diversas propostas de esquemas ou quadros de referência (*frameworks*) voltadas para padronizar ou orientar a realização de avaliações de programas de saúde. Algumas foram desenvolvidas como proposições genéricas para a disciplina de avaliação e posteriormente incorporadas às avaliações de programas de saúde, e outras foram especificamente desenvolvidas para essas últimas. Essas propostas fundamentaram diversas avaliações realizadas nas primeiras décadas do século XXI<sup>25,26,27,28,29</sup>.

A avaliação de programas denominada realista data de 1997<sup>30</sup>, tendo posteriormente evoluído para se apresentar como ciência da avaliação<sup>25</sup>. Seu principal formulador, Ray Pawson<sup>25</sup>, situa essa abordagem como parte das ciências sociais, embora defenda a especificidade e autonomia da disciplina da avaliação que permitiriam considerá-la como ciência. Justifica sua formulação, em 1997, em parceria com Tilley, a partir de uma crítica às limitações dos desenhos então dominantes na avaliação que, na sua opinião, ou eram quase experimentais importados dos ensaios clínicos ou descrições formativas de programas locais. A avaliação realista então foi proposta como uma alternativa que buscava explicar a teoria dos programas. A esse respeito, explicita sua filiação à escola denominada de avaliação centrada na teoria (*theory-driven evaluation*), remetendo sua inspiração à ideia da teoria de médio alcance de Merton e aos trabalhos de Campbell & Popper<sup>25</sup>.

Embora considere que o realismo seja mais uma estratégia de pesquisa geral do que um procedimento técnico, Pawson explicita seus procedimentos metodológicos. Tem por ambição ampliar o objeto da avaliação para além do programa, buscando incorporar o contexto, a história e desvendar os mecanismos das intervenções. O objetivo desse tipo de avaliação seria identificar em um programa “o que funciona, para quem, em quais circunstâncias, a respeito do que, em qual duração (...) e por que”<sup>25</sup> (p. 14). Para isso, propõe analisar o que ele denomina de configuração, contexto, mecanismo e resulta-

do (CMOc)<sup>25</sup>. Contexto seria o conjunto das características dos participantes individuais, suas inter-relações, a localização institucional e a infraestrutura ao redor. Mecanismos seriam os processos, e os resultados seriam os produtos esperados<sup>25</sup>. Propugna o uso de múltiplos métodos: evidências qualitativas para analisar os mecanismos, técnicas quantitativas para os resultados e análise histórica para os contextos<sup>25</sup>. Embora as semelhanças com a tríade donabediana de estrutura-processo-resultados sejam grandes, o autor não a inclui entre suas referências.

Uma segunda perspectiva corresponde ao esquema analítico proposto por Glasgow et al.<sup>31</sup>, em 1999, denominado RE-AIM para avaliação do impacto de intervenções e programas de promoção à saúde. O modelo corresponde à ideia de que o impacto de um programa de Saúde Pública seria uma função de cinco fatores: cobertura, eficácia, adoção, implementação e manutenção (*reach, efficacy, adoption, implementation and maintenance*). Tendo como objetivo superar as limitações dos ensaios clínicos para avaliação de eficácia e visando a subsidiar as avaliações de efetividade e a pesquisa de disseminação, o RE-AIM reproduz a tradição de alguns estudos precursores relacionada à identificação das principais dimensões para avaliação de um programa de saúde<sup>21</sup>. O pressuposto da avaliação dessas dimensões, de forma articulada, é que todas as cinco concorreriam para o impacto de uma política e seriam frequentemente ignoradas em avaliações de eficácia tradicional. No componente eficácia, o modelo inclui tanto efeitos positivos como negativos. Inclui também outros tipos de efeitos, além dos fatores de risco, tanto para os pacientes, a exemplo de comportamentos, como para a equipe de profissionais e sua organização do processo de trabalho. O conceito de adoção está ligado ao número de unidades de saúde que aderem ao programa. Já a efetividade seria um produto da eficácia e da implementação que também é aferida<sup>31</sup>. O balanço feito, segundo seus principais idealizadores vinte anos depois, concluiu que houve uma consolidação desse modelo, apoiado nas citações do artigo seminal e em revisão de trabalhos que utilizaram essa metodologia<sup>26</sup>.

Há uma terceira abordagem representada por pesquisadores vinculados aos MCR (Conselhos de Pesquisa Médica) da Inglaterra e do Canadá e a diversas universidades inglesas, a exemplo de Oxford e Cambridge. Os membros do conselho de pesquisa em serviços de saúde e saúde pública elaboraram um documento inicial, em 2000, que foi sumarizado em artigo de Campbell et al.<sup>32</sup>. Os principais aspectos do documento relacionam-se ao propósito de avaliação de intervenções complexas por meio de desenhos observacionais. As intervenções complexas, entre outras definições, dizem respeito a ações populacionais. O documento propõe cinco fases para o desenvolvimento de uma avaliação, inspiradas no modelo de estudos sobre medicamentos. Na atualização feita em 2006 a linearidade entre as fases foi criticada, e foi ressaltado pelo comitê que os desenhos experimentais, embora importantes, nem sempre são possíveis. Foi ainda conferida importância ao ajuste das intervenções complexas aos contextos locais<sup>27</sup>.

Numa quarta proposição originada na epidemiologia, o Centro de Controle e Prevenção de Doenças (CDC), vinculado ao Departamento de Saúde e Serviços Humanos do governo americano, publicou um modelo para avaliação por meio do seu programa de epidemiologia<sup>33</sup>. A justificativa para a elaboração dessa proposição estava na constatação da ausência de um *framework* para avaliação em saúde pública, a despeito de toda evolução da avaliação como disciplina. O *framework* foi elaborado a partir de uma oficina de trabalho envolvendo especialistas em avaliação, gestores em diversos níveis e pesquisadores. Voltado para a incorporação da avaliação na rotina dos serviços de saúde, o modelo do CDC incorpora as etapas de envolvimento dos profissionais responsáveis pelo programa, definição do foco da avaliação e do desenho, elaboração de modelo lógico, entre outras<sup>33</sup>.

Naquele mesmo ano, os epidemiologistas Habicht et al.<sup>34</sup> publicaram um artigo em que analisavam as possíveis contribuições de desenhos alternativos ao experimental para a avaliação de intervenções populacionais. Os autores propuseram um esquema analítico para a escolha da estratégia de investigação sintonizado com a pergunta da avaliação que pode ser considerada como quinta proposição. A abordagem metodológica é composta de dois eixos: o primeiro diz respeito aos indicadores de interesse, a exemplo da utilização de serviços, cobertura ou medidas de impacto. O segundo eixo refere-se ao tipo de inferência a ser feita. A esse respeito, os autores introduzem dois conceitos além da probabilidade para a inferência: adequação e plausibilidade.

Alguns autores que se propõem a realizar avaliações de programas de saúde, a exemplo de Brownson et al.<sup>35</sup>, situam-se nos desdobramentos da medicina e da saúde pública baseada em evidências que resultaram na denominada pesquisa de implementação e disseminação, da qual a avaliação é um

capítulo <sup>36</sup>. Essa vertente, que partiu da constatação, nos anos 1990, da pequena incorporação da pesquisa nas práticas, tem o propósito de desenvolver-se como uma ciência, da qual a avaliação faria parte. Nessa última perspectiva, há uma tentativa, ainda que retórica, de considerar pertinentes diversas outras abordagens como possibilidades metodológicas para a avaliação voltada para a disseminação e implementação. Com esse objetivo, Gaglio & Glasgow <sup>37</sup> sintetizaram 11 diferentes modelos para avaliação, entre os quais está a avaliação realista, buscando identificar convergências e especificidades. Ressaltam, contudo, a necessidade de criar pontes entre as abordagens formativas ou de processos e as somativas ou de resultados <sup>37</sup>.

### Continuidades e mudanças nas abordagens para avaliação de programas

Como aspecto em comum entre essas diversas proposições, tem-se a preocupação com a avaliação das intervenções complexas ou da complexidade social como um dos problemas a serem equacionados. Os ensaios randomizados persistem como o contraponto para as abordagens que se propõem a superá-los, a exemplo da avaliação realista e também dos quadros de referência (*frameworks*) elaborados por grupos de pesquisadores vinculados a instituições de pesquisa, como é o caso do MRC do Reino Unido <sup>38</sup>. Outro ponto em comum é o recurso às teorias do comportamento como quadro explicativo para as intervenções a serem avaliadas, bem como a incorporação da noção de contexto. Também é uma constante a articulação entre a avaliação e o momento da intervenção, com o planejamento e com a utilização dos resultados. Nesse sentido, pode-se considerar algumas dessas proposições como um desdobramento e atualização dos estudos clássicos, como os de Cook & Campbell <sup>22</sup> sobre os desenhos quase experimentais e seu potencial analítico. Destaca-se também a importância da teoria na avaliação desenvolvida por Chen <sup>39</sup>, da utilização dos resultados da avaliação proposta por Patton <sup>40</sup> e até mesmo a preocupação na incorporação do ponto de vista dos participantes (*stakeholders*) na avaliação, como analisaram Guba & Lincoln <sup>41</sup> e Fetterman & Wandersman <sup>42</sup>. Além desses aspectos, nos modelos revisados por Gaglio & Glasgow <sup>37</sup>, foram apontados como pontos convergentes, entre outros, que a avaliação precisa considerar fatores individuais, organizacionais, do staff e da situação, além de questões de custo e custo-efetividade. Contudo, um estudo empírico, a exemplo do que foi feito por Christie <sup>43</sup> sobre o papel de referência teórica na prática dos avaliadores, poderia identificar variações no interior das aparentes concordâncias em torno da concepção e uso de *frameworks*.

A elaboração de um corpo teórico e metodológico com pretensões à universalidade, revelada pelo uso da ciência como adjetivo conforme proposto por Pawson <sup>25</sup>, guardadas as devidas dimensões, pode ser considerada como um movimento na mesma direção daquele, iniciado nos anos 1970, de desenvolvimento de uma teoria da avaliação e de sua autonomização como disciplina específica. Esse esforço tem sido renovado por alguns dos fundadores ainda atuantes e/ou pelos seus seguidores <sup>3</sup>. Por outro lado, a progressiva incorporação da avaliação como componente do processo de implementação das intervenções no desenvolvimento da denominada “ciência” da implementação <sup>44</sup>, relacionada às outras áreas com grande intersecção com a pesquisa avaliativa, a exemplo da também rotulada “ciência” dos sistemas de saúde <sup>45</sup>, pode ser considerada como um movimento na direção contrária à autonomização desse espaço social.

As abordagens descritas oscilam entre a pesquisa avaliativa, em que os ensaios randomizados são considerados ainda como desenhos capazes de assegurar maior validade interna e externa, e proposições centradas na utilização e na avaliação formativa, com uso de desenhos pouco estruturados. Entre esses dois polos, haveria um gradiente de proposições que buscariam articular ambas as abordagens, usar métodos mistos, desenvolver desenhos quase experimentais e realizar avaliações centradas na teoria.

Cabe uma breve menção à relação entre alguns estudos brasileiros na área da avaliação e as tendências anteriormente mencionadas. À semelhança do ocorrido nos Estados Unidos e Reino Unido, durante os anos 1990, a avaliação no Brasil desenvolveu-se apoiada nos mesmos referenciais teóricos dos autores das três vertentes referidas na introdução, com uma influência adicional do grupo canadense da Universidade de Montreal <sup>46,47</sup>. Por seu turno, o framework desenvolvido pelo grupo canadense incorporava influências de Rossi & Freeman <sup>21</sup>, Chen <sup>39</sup>, Patton <sup>40</sup> e Guba & Lincoln <sup>41</sup> entre outros. Os quadros avaliativos desenvolvidos por autores brasileiros <sup>48,49,50</sup> incorporam alguns

elementos do modelo canadense, em particular a ênfase nas análises de implantação, mas também desenvolvem esquemas avaliativos semelhantes ao RE-AIM de Glasgow, concebidos praticamente ao mesmo tempo<sup>10,49,51</sup>. Essa concomitância também ocorreu no que diz respeito à incorporação das abordagens qualitativas e aos métodos mistos no Brasil<sup>52</sup> e abordagens construtivistas<sup>53</sup>.

## Comentários finais

A análise da evolução dessas abordagens selecionadas sobre a avaliação de programas de saúde revela uma tendência à proposição de novos quadros de referência (*frameworks*) para avaliação que, por vezes, não dialogam entre si e frequentemente recuperam elementos dos antecessores, sem estabelecer claramente a originalidade das proposições em relação às anteriores.

Dois movimentos aparentemente opostos foram identificados: por um lado, a busca de autonomização da avaliação como ciência; por outro, a sua incorporação em outras disciplinas. A avaliação de programas tanto aparece como uma área temática específica, com problemática e metodologia próprias, quanto como um dos objetos possíveis da temática da avaliação em saúde ao lado da avaliação de políticas, tecnologias e serviços de saúde. Todos esses objetos podem ser avaliados a partir da delimitação do objetivo da avaliação, que pode ser de eficácia, efetividade, qualidade, implantação, acesso entre outras<sup>3,7</sup>.

A diversidade verificada pode ser produto da constituição de um espaço social com pequena autonomia relativa, em que o que está em jogo é a autoridade sobre as questões teórico-metodológicas da avaliação de programas de saúde, não apenas do ponto de vista do conhecimento mas também como mercado de consultorias. É um espaço social que, embora tenha surgido no interior da Saúde Pública, no âmbito internacional, e no campo da Saúde Coletiva no Brasil, é, ao mesmo tempo, dominado pelo campo médico e pelo espaço mais geral da avaliação de programas.

Embora a referência direta aos clássicos da avaliação esteja bastante reduzida em artigos dos últimos cinco anos, ela é mediada por esse conjunto de estudos do final dos anos 1990 e início do século XXI, acima sumarizados, cujos modelos teórico-metodológicos (*frameworks*) para avaliação de programas de saúde se apoiaram nos autores fundadores da avaliação como disciplina. Pode-se dizer que tanto a avaliação focada na utilização como as abordagens construtivistas foram englobadas nas duas vertentes que buscam a unificação da disciplina acadêmica. Assim, a *implementation science* incorpora o discurso da avaliação para a utilização e a proposta de uma ciência realista da avaliação, incorporando a teoria e as abordagens qualitativas.

Ainda que a relação da pesquisa avaliativa com as pesquisas em outras disciplinas, particularmente a pesquisa social, seja reconhecida pela maioria dos autores, há uma tendência ambivalente e recorrente de defini-la como uma disciplina autônoma e com especificidades. A proposição de Pawson de uma ciência da avaliação é relativamente recente e passível de críticas que fogem ao escopo desse artigo, porém podem ser indicadas. A principal crítica é a subordinação da avaliação às disciplinas das ciências humanas e sociais que, até o momento, não dispõem de uma teoria unificada do social, e sim, de tradições teóricas<sup>54</sup>. Outro aspecto não menos relevante é a ausência de diálogo com as teorias sociológicas das práticas sociais clássicas e contemporâneas para explicar os programas de saúde e sua implantação.

Dessa forma, a busca da consolidação da avaliação como campo de saberes e práticas relativamente autônomo, no sentido usado por Bourdieu<sup>55</sup>, tem ocorrido concomitantemente com a persistência da tendência histórica à fragmentação e pulverização das abordagens e *frameworks*, bem como de sua incorporação a outras disciplinas. A dinâmica entre esses dois movimentos pode ser produto não só da composição desse espaço social situado entre os campos científico e burocrático, entre a academia e os serviços de saúde, mas também da natureza do objeto da avaliação, que implica julgamento sobre políticas, governos e governantes<sup>56</sup>. A avaliação de programas de saúde como componente desse espaço social mais amplo sofre influência dessas mesmas tendências. A realização de revisões sistemáticas sobre a temática poderá dimensionar a magnitude dessas diferentes propostas teórico-metodológicas, bem como identificar outras. Também uma análise das trajetórias dos fundadores e autores mais influentes, das redes de relações e das condições históricas de possibilidade poderá contribuir para uma melhor compreensão desses processos.

## Colaboradores

L. M. Vieira-da-Silva concebeu a problemática do artigo, contribuiu para a revisão bibliográfica, analisou dados e redigiu o manuscrito. J. P. Furtado contribuiu com a construção da problemática, realizou a revisão bibliográfica, analisou dados e participou da redação.

## Informações adicionais

ORCID: Lígia Maria Vieira-da-Silva (0000-0003-2518-411X); Juarez Pereira Furtado (0000-0001-6605-1925).

## Agradecimentos

Os autores agradecem a Wagner Y. Oda, que contribuiu com a revisão preliminar e elaboração do banco de dados que subsidiou a análise.

## Referências

1. Furtado JP, Vieira-da-Silva LM. A avaliação de programas e serviços de saúde no Brasil enquanto espaço de saberes e práticas. *Cad Saúde Pública* 2014; 30:2643-55.
2. Bourdieu P. *Sociologie generale*. Paris: Seuil; 2016.
3. Alkin MC. *Evaluation roots: a wider perspective of theorists' views and influences*. 2<sup>nd</sup> Ed. Los Angeles: Sage; 2013.
4. Weiss CH. *Evaluation research methods for assessing program effectiveness*. New York: Prentice Hall; 1972.
5. Scriven M. *Evaluation thesaurus*. 4<sup>th</sup> Ed. Thousand Oaks: Sage; 1991.
6. Shadish Jr. WR, Cook TD, Leviton LC. *Foundations of program evaluation: theories of practice*. Newbury Park: Sage; 1999.
7. Rossi PH. My views of evaluation and their origins. In: Alkin MC, editor. *Evaluation roots: a wider perspective of theorists' views and influences*. 2<sup>nd</sup> Ed. Los Angeles: Sage; 2013. p. 106-12.
8. Mertens DM, Wilson AT. *Program evaluation: theory and practice*. 2<sup>nd</sup> Ed. New York: The Guilford Press; 2019.
9. Vieira-da-Silva LM. *O campo da Saúde Coletiva: gênese, transformações e articulações com a reforma sanitária brasileira*. Rio de Janeiro: Editora Fiocruz/Salvador: EdUFBA; 2018.
10. Vieira-da-Silva LM, Formigli VLA. Avaliação em saúde: limites e perspectivas. *Cad Saúde Pública* 1994; 10:80-91.
11. Novaes HMD. Avaliação de programas, serviços e tecnologias em saúde. *Rev Saúde Pública* 2000; 34:547-9.
12. Hartz ZMA. Institutionalizing the evaluation of health programs and policies in France: cuisine internationale over fast food and sur mesure over ready-made. *Cad Saúde Pública* 1999; 15:229-60.
13. Furtado JP, Gasparini MFV. Há diferenças entre avaliar e analisar? *Ciênc Saúde Colet* 2019; 24:2933-8.
14. Crumpton CD, Medeiros JJ, Ferreira VRS, Sousa MM, Najberg E. Avaliação de políticas públicas no Brasil e nos Estados Unidos: análise da pesquisa nos últimos 10 anos. *Rev Adm Pública* 2016; 50:981-1001.
15. Furtado JP, Campos GWS, Oda WY, Campos RO. Planejamento e avaliação em saúde: entre antagonismo e colaboração. *Cad Saúde Pública* 2018; 34:e00087917.
16. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966; 44:166-206.
17. Donabedian A. *Explorations in quality assessment and monitoring volume 1: the definition of quality and approaches to its assessment*. Ann Arbor: Health Administration Press; 1980.
18. Goldbaum M. Epidemiologia e serviços de saúde. *Cad Saúde Pública* 1996; 12 Suppl 2:S95-8.
19. Novaes HMD. Pesquisa em, sobre e para os serviços de saúde: panorama internacional e questões para a pesquisa em saúde no Brasil. *Cad Saúde Pública* 2004; 20 Suppl 2:S147-57.
20. Campbell DT, Stanley JC. *Experimental and quasi experimental designs for research*. Boston: Houghton Mifflin Company; 1963.

21. Rossi PH, Freeman HE. Evaluation: a systematic approach. 3<sup>rd</sup> Ed. Beverly Hills: Sage; 1985.
22. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Boston: Houghton Mifflin; 1979.
23. Mark MM, Caracelli V, McNall MA, Miller RL. The oral history of evaluation. *Am J Eval* 2018; 39:290-304.
24. Christie CA, Alkin MC. An evaluating theory tree. In: Alkin MC, editor. Evaluation roots: a wider perspective of theorists' views and influences. 2<sup>nd</sup> Ed. Los Angeles: Sage; 2013. p. 11-57.
25. Pawson R. The science of evaluation: a realist manifesto. London: Sage; 2013.
26. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. *Front Public Health* 2019; 7:64.
27. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008; 337:a1655.
28. Rivera Dommarco JA, González de Cosío T, García-Chávez CG, Colchero MA. The role of public nutrition research organizations in the construction, implementation and evaluation of evidence-based nutrition policy: two national experiences in Mexico. *Nutrients* 2019; 11:594.
29. Heitzman M, Weitzel J, Kroll S, Zabler B. Client experiences in a prenatal home visiting program: a prenatal care coordination program evaluation. *Public Health Nurs* 2019; 36:653-9.
30. Pawson R, Tilley N. Realistic evaluation. London: Sage; 1997.
31. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health* 1999; 89:1322-7.
32. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000; 321:694-6.
33. Centers for Disease Control and Prevention. Framework For Program Evaluation In Public Health. *MMWR* 1999; 48(RR-11):1-40.
34. Habicht J, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol* 1999; 28:10-8.
35. Brownson RC, Colditz CA, Proctor EK, editors. Dissemination and implementation research in health: translating science to practice. 2<sup>nd</sup> Ed. New York: Oxford University Press; 2018.
36. Chambers D. Foreword. In: Brownson RC, Colditz CA, Proctor EK, editors. Dissemination and implementation research in health: translating science to practice. 2<sup>nd</sup> Ed. New York: Oxford University Press; 2018. p. vii-ix.
37. Gaglio B, Glasgow RE. Evaluation approaches for dissemination and implementation research. In: Brownson RC, Colditz CA, Proctor EK, editors. Dissemination and implementation research in health: translating science to practice. 2<sup>nd</sup> Ed. New York: Oxford University Press; 2018. p. 327-56.
38. Medical Research Council. A framework for the development and evaluation of RCTs for complex interventions to improve health. London: Medical Research Council; 2000.
39. Chen HT. Theory-driven evaluations. Beverly Hills: Sage; 1990.
40. Patton MQ. Utilization-focused evaluation. 2<sup>nd</sup> Ed. Beverly Hills: Sage; 1986.
41. Guba EG, Lincoln YS. Fourth generation evaluation. Newbury Park: Sage; 1989.
42. Fetterman DM, Wandersman A. Empowerment evaluation: theories, principles, concepts and steps. In: Fetterman DM, Kaftarian S, Wandersman A, editors. Empowerment evaluation: knowledge and tools for self-assessment, evaluation capacity building, and accountability. Thousand Oaks: Sage; 2015. p. 20-42.
43. Christie CA. What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Dir Eval* 2003; 2003:7-36.
44. Madon T, Hofman KJ, Kupfer L, Glass RI. Public health: implementation science. *Science* 2007; 318:1728-9.
45. Skochelak SE, Hawkins RE, Lawson LE, Starr SR, Borkan JM, Gonzalo JD. Health systems science. St. Louis: Elsevier; 2017.
46. Brousselle A, Champagne F, Contandriopoulos A-P, Hartz ZMA, organizadores. Avaliação: conceitos e métodos. Rio de Janeiro: Editora Fiocruz; 2011.
47. Contandriopoulos A-P, Champagne F, Denis J-L, Avargues M. L'évaluation dans le domaine de la santé: concepts et méthodes. *Rev Epidemiol Santé Publique* 2000; (48):517-39.
48. Hartz ZMA, Champagne F, Contandriopoulos A-P, Leal MC. Avaliação do Programa Materno-infantil: análise de implantação em sistemas locais de saúde no Nordeste do Brasil. In: Hartz ZMA, organizadora. Avaliação em saúde: dos modelos conceituais à prática na análise da implantação de programas. Rio de Janeiro: Editora Fiocruz; 1997. p. 89-131.
49. Vieira-da-Silva LM, Hartz ZMA, Chave SCL, Silva GAP, Paim JS. Análise da implantação da gestão descentralizada em saúde: estudo comparado de cinco casos na Bahia, Brasil. *Cad Saúde Pública* 2007; 23:355-70.
50. Hartz ZMA, Vieira-da-Silva LM, organizadoras. Avaliação em saúde: dos modelos teóricos à prática na avaliação de programas e sistemas de saúde. Rio de Janeiro: Editora Fiocruz; 2005.
51. Vieira-da-Silva LM. Avaliação do processo de descentralização das ações de saúde. *Ciênc Saúde Colet* 1999; 4:331-9.

52. Minayo MCS, Cruz Neto O. Triangulación de métodos en la evaluación de programas y servicios de salud. In: Bronfman M, Castro R, organizadores. Salud, cambio social y política: perspectivas desde América Latina. Cor Del Valle: Edamex; 1999. p. 65-80.
53. Furtado JP. Um método construtivista para a avaliação em saúde. *Ciênc Saúde Colet* 2001; 6:165-81.
54. Bourdieu P, Chamboredon J-C, Passeron J-C, Teixeira GJF. A profissão do sociólogo: preliminares epistemológicas. Petrópolis: Editora Vozes; 1999.
55. Bourdieu P. As regras da arte: gênese e estrutura do campo literário. Lisboa: Presença; 1996.
56. Carol HW. Rooting for evaluation: digging into beliefs. In: Alkin MC, editor. *Evaluation roots: a wider perspective of theorists' views and influences*. 2<sup>nd</sup> Ed. Los Angeles: Sage; 2013. p. 130-43.

---

Recebido em 12/Dez/2019  
Versão final reapresentada em 26/Fev/2020  
Aprovado em 02/Mar/2020