

Desempenho de intervenções de saúde em países da América Latina: uma revisão sistemática

Performance evaluation of health interventions in Latin America: a systematic review

Juliana Martins Barbosa da Silva Costa¹, Vanessa de Lima Silva², Isabella Chagas Samico³, Eduarda Ângela Pessoa Cesse⁴

RESUMO Este estudo teve como objetivo investigar na literatura latino-americana os modelos teóricos de avaliação de desempenho de intervenções de saúde nessa Região. Para tanto, realizou-se revisão sistemática da literatura nas bases do Lilacs, Medline e Capes, com termos relacionados à avaliação e desempenho, durante o período 2008-2013. Utilizaram-se critérios da *Joint Committee on Standards for Educational Evaluation* para apreciação da qualidade bem como de matriz para análise dos atributos bibliométricos e dos modelos. Encontraram-se seis estudos explicitando o modelo de avaliação de desempenho de intervenções em saúde com destaque para modelos unidimensionais com abordagem sistêmica.

PALAVRAS-CHAVE Avaliação em saúde; Modelos teóricos; América Latina; Revisão.

ABSTRACT *This study aimed to investigate, in Latin American literature, theoretical models to assess performance evaluation of health interventions in the Region. To this end, a systematic review of literature on the Lilacs, Medline and Capes' databases viewing assessment and performance-related terms for the period 2008-2013 was conducted. Criteria based on the Joint Committee on Standards for Educational Evaluation were applied to assess the quality of articles as well as a matrix to analyze bibliometric attributes array and models. Six studies were found explaining the assessment model of health intervention performance, especially for one-dimensional models with systemic approach.*

KEYWORDS *Health evaluation; Models, theoretical; Latin America; Review.*

¹Universidade Federal de Pernambuco (UFPE) - Recife (PE), Brasil.
julimartins.costa@gmail.com

²Universidade Federal de Pernambuco (UFPE) - Recife (PE), Brasil.
vanelima@gmail.com

³Instituto de Medicina Integral Prof. Fernando Figueira (Imip) - Recife (PE), Brasil.
isabella.samico@gmail.com

⁴Fundação Oswaldo Cruz Pernambuco (Fiocruz - PE), Centro de Pesquisas Aggeu Magalhães - Recife (PE), Brasil.
educesse@cpqam.fiocruz.br

Introdução

O desempenho é uma preocupação crescente na gestão dos sistemas e serviços de saúde por todo o mundo (MARCHAL ET AL., 2014). O estabelecimento de mecanismos que auxiliem uma prática gerencial focada no aprimoramento da governança pública tem sido cada vez mais discutido na literatura (CARVALHO ET AL., 2012; HUNTER; NIELSEN, 2013; LAHREY; NIELSEN, 2013; MARCHAL ET AL., 2014).

A avaliação de desempenho surge nesse contexto, e objetiva melhorar a eficiência, eficácia e qualidade das intervenções (programas, serviços e sistemas) de saúde (CONTRANDIOPOULOS ET AL., 1997; CARVALHO ET AL., 2012; HATRY, 2013), bem como aumentar a responsabilização e a transparência da gestão pública (LAHREY; NIELSEN, 2013). Diversos Organismos Internacionais (WHO, 2000; OPAS, 2001; USAID, 2011) têm desenvolvido modelos teóricos para avaliar o desempenho de organizações de saúde, estimulando sua implantação e a instituição de sistemas de gerenciamento do desempenho.

Na América Latina, a avaliação de desempenho foi impulsionada pelo *'World Health Report 2000'* (WHO, 2000), pelo quadro de referência proposto pela Organização Pan-Americana de Saúde (OPAS, 2001) e por exigências de instituições financeiras, tais como o Banco Mundial e o Fundo Monetário Internacional. Algumas iniciativas foram desenvolvidas na região com o intuito de propor modelos que se aproximassem da realidade local, como o Proadess, que delinea uma proposta para avaliar o desempenho do sistema de serviços de saúde do Brasil (VIACAVA ET AL., 2004).

Apesar da importância crescente do tema, a delimitação do conceito de desempenho é uma das mais imprecisas em termos de teoria organizacional, surgindo, ao longo do tempo, inúmeras definições e modelos para avaliar o desempenho. Essa diversidade de abordagens gera confusão e dúvidas quanto à escolha do modelo teórico mais apropriado

à cada localidade e formas adequadas para sua operacionalização (VIACAVA ET AL., 2004; CHAMPAGNE ET AL., 2011).

A definição hegemônica de desempenho está ancorada na escola racional, que entende o desempenho como o grau de alcance dos objetivos e metas da organização (SICOTTE ET AL., 1998). Contudo, a teoria organizacional é vasta e apresenta uma profusão de modelos que carregam consigo visões diferenciadas sobre o desempenho. Alguns modelos valorizam a estabilidade e o controle; outros, a aquisição de recursos, o crescimento e a adaptação e, outros, a negociação e o compromisso (CHAMPAGNE ET AL., 2011; MARCHAL ET AL., 2014). Apesar dessa aparente discordância, Champagne et al. (2011) ponderam que todos os modelos são legítimos e focam dimensões diferentes do desempenho.

O aprofundamento no tema e a verificação da insuficiência das abordagens que valorizam uma única dimensão do desempenho, os denominados modelos unidimensionais, para avaliar organizações complexas como as do setor saúde levaram ao desenvolvimento de modelos multidimensionais (CHAMPAGNE ET AL., 2011; VIACAVA ET AL., 2004). Esses modelos permitem uma análise mais global e examinam o desempenho a partir da junção conceitual de dois ou mais modelos unidimensionais (CHAMPAGNE ET AL., 2011; TCHOUAKET ET AL., 2012).

Outra questão que vai além das concepções e abrangência dos modelos para avaliar o desempenho é o fato que muitos deles foram elaborados para avaliar o desempenho de organizações que estão fora do setor saúde ou, quando relacionados, focam serviços específicos (CHAMPAGNE ET AL., 2011; MARCHAL ET AL., 2014; TCHOUAKET ET AL., 2012) ou realidades bem diversas da encontrada na América Latina, o que traz a preocupação adicional no que tange à escolha daquele que seria o mais adequado para a realidade desses países.

Apesar de suas particularidades, os países latino-americanos apresentam características semelhantes no que tange à

sua conformação histórica, política e social. De maneira geral, caracterizam-se por uma elevada concentração de renda, desigualdades sociais e instabilidade política e econômica (SOUZA, 2011; HOMEDES; UGALDE, 2011). Além disso, a estruturação dos processos gerenciais e a baixa qualificação da gestão pública impõem para a região desafios comuns, tais como aumento insustentável nos custos, ineficiência no gasto público, baixa cobertura, acesso desigual, ausência de mecanismos de avaliação do desempenho e pouca transparência na prestação de contas para a sociedade (HOMEDES; UGALDE, 2011; OPAS, 2012).

Dessa forma, surge a necessidade de identificar os modelos teóricos adotados por pesquisadores latino-americanos para avaliar o desempenho de intervenções de saúde como forma de adicionar luz ao debate sobre o tema. Assim, este estudo teve como objetivo identificar na literatura latino-americana os modelos teóricos para avaliar o desempenho de intervenções de saúde. A intenção é a de que sirva de subsídio às futuras pesquisas que pretendam analisar a pertinência e adequação dos modelos à realidade destes países.

Métodos

Realizou-se uma revisão sistemática da literatura, orientada pela seguinte pergunta de pesquisa: Quais os modelos utilizados para avaliar o desempenho de intervenções de saúde presentes na literatura de países da América Latina, no período de janeiro de 2008 a outubro de 2013?

Para tanto, procedeu-se pesquisa bibliográfica de textos científicos (artigos, teses, dissertações e monografias) publicados no período de janeiro de 2008 a outubro de 2013, contendo temas relacionados à avaliação do desempenho de intervenções (programas, serviços e sistemas) de saúde, publicados nos idiomas, português, inglês ou espanhol. O período de estudo foi selecionado com o

intuito de identificar a produção científica mais recente sobre tema.

Modelos de avaliação de desempenho são esquemas teórico-conceituais que especificam os elementos utilizados para medir, monitorar e promover processo sistemático de melhoria do desempenho das intervenções de saúde (KLASSEN ET AL., 2009; CHAMPAGNE ET AL., 2011).

A pesquisa guiada sobre publicações foi realizada nos meses de outubro e novembro de 2013 nas bases de dados Lilacs (Literatura Latino-Americana e do Caribe em Ciências da Saúde) e Medline (*Medical Literature Analysis and Retrieval System Online*) a partir dos termos relacionados à avaliação em saúde ou desempenho que fazem parte da lista de Descritores em Ciências da Saúde (DeCS), disponível no portal da Biblioteca Virtual em Saúde (<http://decs.bvs.br>).

Para a pesquisa da produção científica, selecionaram-se os países latino-americanos constantes da classificação da Comissão Econômica para a América Latina e o Caribe (Cepal), órgão da Organização das Nações Unidas, que distingue os países americanos de língua latina em latino-americanos e caribenhos. Dessa forma, foi pesquisada a produção científica dos seguintes países: Argentina, Bolívia, Brasil, Chile, Colômbia, Equador, México, Paraguai, Peru, Uruguai e Venezuela.

Utilizou-se a seguinte expressão de busca: 'avaliacao de desempenho' or 'avaliacao de impacto' or 'avaliacao de processos (cuidados de saúde)' or 'avaliacao de processos e resultados (cuidados de saude)' or 'avaliacao de programas' or 'avaliacao de programas e projetos de saude' or 'avaliacao de resultados (cuidados de saude)' or 'avaliacao do impacto na saude' or 'mecanismos de avaliacao em cuidados de saude' or 'benchmarking' or 'benchmarking em assistencia a saude' or 'avaliacao de servicos de saude' [descriptor de assunto] and '2008' or '2009' or '2010' or '2011' or '2012' or '2013' [ano de publicação] and 'argentina' or 'bolivia' or 'brasil' or

'chile' or 'colombia' or 'equador' or 'mexico' or 'paraguai' or 'peru' or 'uruguai' or 'venezuela' [país de afiliação].

Além dos artigos científicos, a *grey literature* foi pesquisada com o objetivo de incluir o maior número de modelos de avaliação de desempenho, principalmente a literatura científica ainda não publicada. Dessa forma, pesquisou-se o banco de teses da Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), agência de fomento à pesquisa ligada ao Ministério da Educação do Brasil que congrega todos os programas de pós-graduação deste País. A chave de busca utilizada foi: 'desempenho' or 'avaliacao

de desempenho' and 'saude' [descriptor de assunto] and '2008' or '2009' or '2010' or '2011' or '2012' or '2013' [ano de publicação].

A leitura do título e autor(es) permitiu a exclusão de duplicidades. A seleção dos textos foi realizada em três etapas. A primeira consistiu na leitura dos resumos de cada trabalho, realizada por dois examinadores independentes, com base em critérios de inclusão e exclusão (*quadro 1*). Aplicou-se o índice de *Kappa* para análise de concordância entre os examinadores. Nos casos de discordância, realizaram-se reuniões de consenso com a presença de um terceiro examinador.

Quadro 1. Critérios de inclusão e exclusão dos textos na revisão sistemática sobre modelos de avaliação do desempenho de intervenções de saúde em países da América Latina

| CRITÉRIOS DE INCLUSÃO | CRITÉRIOS DE EXCLUSÃO |
|---|--|
| Explicitar o modelo teórico para avaliar o desempenho de sistemas, serviços ou programas de saúde | Não explicitar o modelo utilizado para avaliar o desempenho de organizações de saúde Tratar do desempenho profissional, cuidados individuais ou do desempenho de tecnologias médicas e biomédicas Apenas apresentar o modelo teórico-lógico da intervenção (programa, serviço ou sistema de saúde) |
| Ter sido realizado no período de 1 de janeiro de 2008 até 31 de outubro de 2013 | Ter sido realizado antes de janeiro de 2008 |
| Estar publicado em português, inglês ou espanhol | Estar publicado em outras línguas que não o português, inglês ou espanhol |
| Ter como objeto de estudo a América Latina ou seus países componentes (Argentina, Bolívia, Brasil, Chile, Colômbia, Equador, México, Paraguai, Peru, Uruguai, Venezuela), de forma isolada ou em conjunto com países de outras regiões (Europa, África, Ásia, Oceania e os da América do Norte, Central ou do Sul que não sejam classificados como latino americanos) | Ter como objeto de estudo unicamente países europeus, africanos, asiáticos, da Oceania, caribenhos e americanos (exceto os latino-americanos: Argentina, Bolívia, Brasil, Chile, Colômbia, Equador, México, Paraguai, Peru, Uruguai, Venezuela) |

Fonte: Elaboração própria.

Concluída essa fase, os textos com os resumos que preencheram os critérios de inclusão (*quadro 1*) foram selecionados para a segunda etapa, na qual se procedeu à leitura completa dos textos por dois examinadores independentes, incorporando nessa etapa mais um critério de inclusão: a identificação explícita no texto do modelo de avaliação do desempenho utilizado no estudo.

Ao término da leitura completa calculou-se o índice *Kappa*, após o que os textos incluídos passaram para a terceira e última etapa de seleção. Essa etapa consistiu na análise da qualidade dos textos selecionados. Para tanto, aplicaram-se os critérios de avaliação da utilidade e qualidade de estudos avaliativos propostos pela *Joint Committee on Standards for Educational Evaluation*

(*JCSEE*), traduzidos e adaptados à realidade brasileira por Santos e Natal (2006).

Os critérios propostos pela *JCSEE* estão divididos em quatro parâmetros: a utilidade (sete critérios) verifica se a avaliação atende às necessidades de informação dos usuários; a factibilidade ou viabilidade (três critérios) analisa se o estudo é realista e moderado nos custos, de modo a justificar a sua realização; a propriedade (oito critérios) verifica se o processo avaliativo é conduzido eticamente, com respeito ao bem-estar dos envolvidos; e a acurácia ou precisão (doze critérios) analisa aspectos relativos à validade do julgamento de valor ou mérito realizado (SANTOS; NATAL, 2006).

Cada um desses critérios recebeu uma pontuação que variou de 0 a 10. Para julgar a qualidade calculou-se a média dos critérios por parâmetro classificados conforme os escores obtidos, segundo a seguinte classificação: excelente de 9,0 – 10,0; bom de 7,0 – 8,9; regular de 5,0 a 6,9; fraco de 3,0 – 4,9 e crítico < 3,0. Adotou-se como critério de exclusão escore crítico em qualquer um dos critérios avaliados (SANTOS; NATAL, 2006).

Os textos incluídos tiveram seu conteúdo analisado e sistematizado por meio de uma matriz de extração de dados. Nela, foram descritos: título; ano de publicação; idioma; país de afiliação; objetivo do estudo; modelo teórico utilizado para avaliar o desempenho; abordagem conceitual; tipo de inovação, tipo de modelo, dimensões avaliadas, nível de complexidade e fonte.

Na abordagem conceitual, classificou-se os modelos segundo os objetivos e características em três categorias denominadas de determinação social, centrado nos objetivos e sistêmico. Os modelos com a abordagem da determinação social são aqueles que possuem uma compreensão hierarquizada e integrada dos vários fatores que influenciam a saúde como um todo. Os que possuem uma abordagem centrada nos objetivos, consideram que uma intervenção apresenta um bom desempenho quando consegue atingir seus objetivos e metas. Os modelos ancorados na

abordagem sistêmica analisam a intervenção como um conjunto de elementos interconectados, de modo a formar um todo indivisível, podendo ser compreendido enquanto um sistema racional, no qual o desempenho é avaliado por normas não apenas de resultados, mas também, de processo e/ou estrutura, ou como um sistema de ação social, em constante transformação (REIS, 2012).

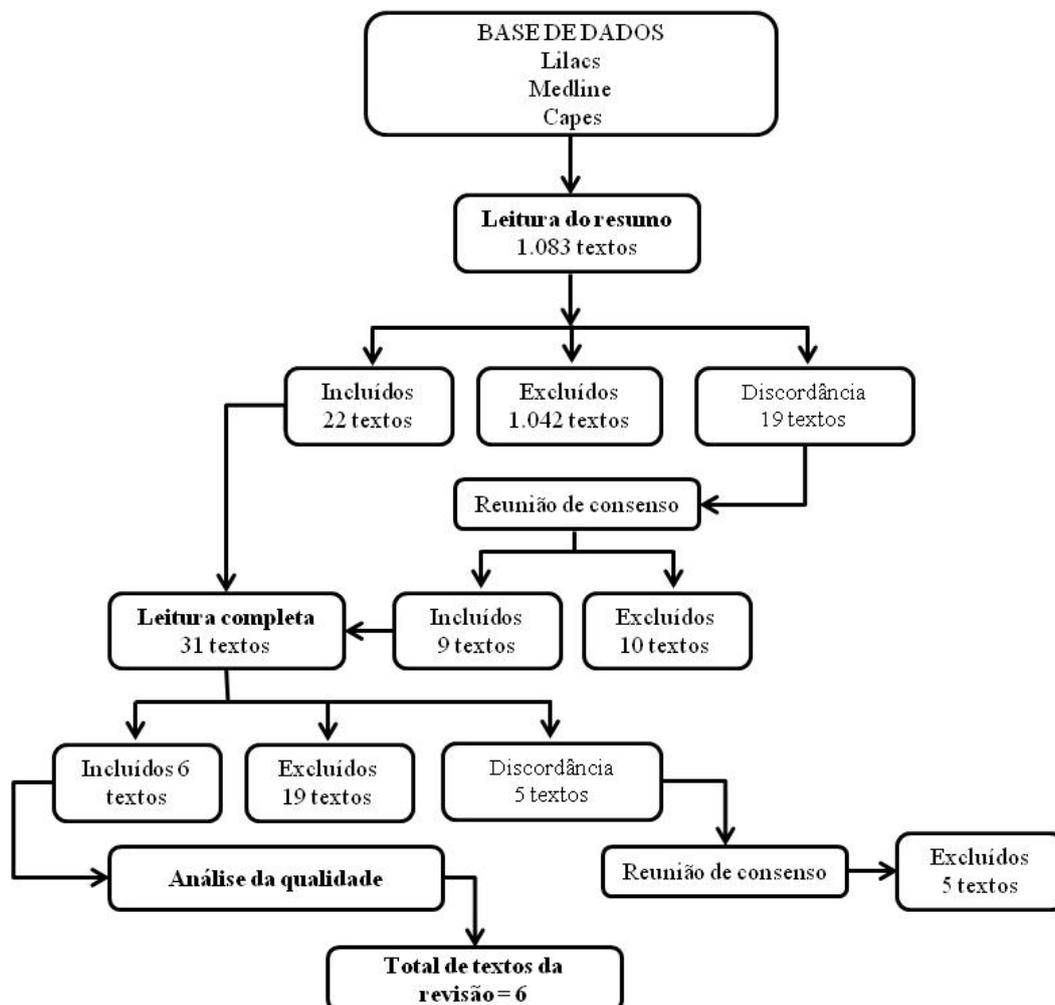
Em relação ao tipo de inovação, classificou-se os modelos quanto ao seu ineditismo, podendo ser um novo modelo ou aplicação de pré-existentes. O tipo de modelo considerou o quantitativo de dimensões envolvidas na avaliação de desempenho classificando-os em unidimensional, quando privilegia-se uma única dimensão do desempenho, ou multidimensional, quando há a junção conceitual de dois ou mais modelos unidimensionais (CHAMPAGNE, 2011; TCHOUAKET ET AL., 2012). O nível de complexidade da intervenção a ser avaliada considerou os recursos necessários para seu desenvolvimento classificando-os em programa, serviço ou sistema de saúde. A fonte de informação para a coleta de dados denominou-se de artigo científico ou *grey literature*.

Resultados

A busca nas bases de dados levou à identificação de um total de 1.262 textos científicos (674 no Lilacs, 525 no *Medline* e 64 no portal de teses da Capes). Desses, 179 foram excluídos por duplicidade, restando 1.083.

Na leitura dos resumos, 1.042 foram excluídos por não atenderem aos critérios de inclusão e 19 apresentaram discordância entre os examinadores. Para estes últimos, procedeu-se à leitura e análise por um terceiro examinador e uma reunião de consenso foi realizada, na qual foram incluídos nove e excluídos dez textos (*figura 1*). A concordância entre os examinadores foi substancial ($Kappa = 0,69$; $p > 0,05$). Assim, 31 textos foram selecionados para a etapa seguinte.

Figura 1. Fluxograma do processo de seleção de textos da revisão sistemática sobre modelos de avaliação do desempenho de intervenções de saúde em países da América Latina



Na segunda fase da seleção, os 31 textos foram lidos na íntegra. Desses, seis foram selecionados, 20 não atenderam aos critérios e foram excluídos do estudo e cinco apresentaram discordância entre os examinadores. A análise da concordância entre examinadores foi substancial ($Kappa = 0.60$; $p > 0,05$). Após a reunião de consenso, todos os textos discordantes foram excluídos (figura 1).

Portanto, ao término da segunda etapa, foram selecionados seis textos que tinham por objetivo avaliar o desempenho de programas, serviços ou sistemas de saúde e que

explicitavam o modelo teórico para esta avaliação.

No que concerne à análise da qualidade, os textos apresentaram critérios de qualidade entre excelente e regular, não havendo exclusão de nenhum texto nessa etapa do estudo. De maneira geral, os artigos de Yavich, Báscolo e Haggerty (2010) Brizola, Cordoni Júnior, (2011), Moraes e Andrade (2011) apresentaram uma melhor qualidade do que as teses de Battesini (2008) Duarte (2009), Ponce (2012). Dentre os parâmetros analisados, a utilidade apresentou o escore mais baixo e a propriedade o mais alto (tabela 1).

Tabela 1. Avaliação da qualidade dos textos da revisão sistemática sobre modelos de avaliação do desempenho de intervenções de saúde em países da América Latina

| Crítérios | Batesini, 2008 | Duarte, 2009 | Yavich et al., 2010 | Moraes & Andrade, 2011 | Brizola et al., 2011 | Ponce, 2012 |
|---|-----------------------|---------------------|----------------------------|-----------------------------------|-----------------------------|---------------------|
| Utilidade (U) | 8,6 **** | 7,0 *** | 7,9 *** | 7,3 *** | 8,7 **** | 6,4 *** |
| U1 - Identificação dos interessados | 10 | 3 | 5 | 3 | 5 | 5 |
| U2 - Credibilidade do avaliador | 10 | 10 | 10 | 10 | 10 | 10 |
| U3 - Escopo e seleção de informações | 10 | 9 | 10 | 10 | 10 | 10 |
| U4 - Identificação de valores | 10 | 3 | 3 | 3 | 10 | 3 |
| U5 - Clareza dos relatórios | 10 | 10 | 9 | 7 | 8 | 3 |
| U6 - Entrega a tempo e disseminação de relatórios | 3 | 7 | 10 | 10 | 10 | 7 |
| U7 - Impacto da avaliação | 7 | 7 | 8 | 8 | 8 | 7 |
| Factibilidade (F) | 9,7 ***** | 7,7 *** | 9,3 ***** | 8,3 **** | 7,7 *** | 8,0 **** |
| F1- Procedimentos práticos | 10 | 10 | 10 | 10 | 5 | 8 |
| F2 - Viabilidade política | 9 | 5 | 8 | 5 | 10 | 8 |
| F3 - Custo-efetividade | 10 | 8 | 10 | 10 | 8 | 8 |
| Propriedade (P) | 8,2 **** | 9,0 ***** | 10,0 ***** | 9,0 ***** | 9,6 ***** | 9,0 ***** |
| P1 - Orientação para o serviço | 10 | 9 | 10 | 10 | 10 | 10 |
| P2 - Acordos formais | 10 | 10 | 10 | 10 | 10 | 10 |
| P3 - Direitos dos indivíduos | 10 | 9 | 10 | 5 | 10 | 8 |
| P4 - Relações humanas | 3 | 10 | 10 | 10 | 10 | 7 |
| P5 - Avaliação completa e justa | 8 | 7 | 10 | 10 | 8 | 10 |
| P6 - Disseminação de resultados | 10 | 9 | 10 | 10 | 10 | 10 |
| P7 - Conflito de interesses | 10 | 10 | 10 | 10 | 10 | 10 |
| P8 - Responsabilidade fiscal | 10 | 9 | 10 | 5 | 10 | 8 |
| Acurácia (A) | 8,3 **** | 8,6 **** | 8,0 **** | 8,8 **** | 8,5 **** | 8,5 **** |
| A1 - Documentação do programa | 10 | 10 | 8 | 7 | 6 | 6 |
| A2 - Análise do contexto | 8 | 8 | 8 | 8 | 8 | 8 |
| A3- Descrição dos propósitos e procedimentos | 10 | 10 | 5 | 10 | 10 | 10 |
| A4 - Fontes confiáveis de informação | 8 | 9 | 8 | 10 | 8 | 8 |
| A5 - Informações válidas | 8 | 8 | 8 | 8 | 8 | 8 |
| A6 - Informações confiáveis | 6 | 8 | 8 | 8 | 8 | 6 |
| A7 - Informação sistemática | 10 | 10 | 8 | 10 | 10 | 10 |
| A8 - Análise quantitativa | 9 | 10 | | | 9 | 10 |
| A9 - Análise qualitativa | | | 7 | | | |
| A10 - Conclusões justificáveis | 5 | 5 | 5 | 9 | 9 | 9 |
| A11 - Imparcialidade dos relatórios | 9 | 9 | 5 | 9 | 9 | 10 |
| A12 - Meta-avaliação | 10 | 10 | 8 | 7 | 6 | 6 |
| Média | 8,7 **** | 8,1 **** | 8,8 **** | 8,3 **** | 8,6 **** | 8,0 **** |

Fonte: Elaboração própria.

Nota: Excelente: ***** Bom: **** Regular: *** Fraco: ** Crítico: *.

A maioria dos estudos incluídos na revisão foi desenvolvida no Brasil (n=5), publicada em 2011 (n=2), da qual a metade era composta

de artigos indexados (n=3) e os demais (n=3) procederam da *grey literature* e eram provenientes de teses de doutoramento (*quadro 2*).

Quadro 2. Textos selecionados na revisão sistemática sobre modelos de avaliação do desempenho de intervenções de saúde em países da América Latina

| AUTOR | TÍTULO | ANO | PAÍS | TIPO |
|-----------------------|--|------|-----------|------------------------|
| Battesini | Modelo multidimensional para avaliação de desempenho da vigilância sanitária: uma aplicação em nível municipal | 2008 | Brasil | Grey literature (Tese) |
| Duarte | Modelo sistêmico de avaliação de desempenho hospitalar por meio de indicadores assistenciais: o caso do indicador de dor aguda no pós-operatório | 2009 | Brasil | Grey literature (Tese) |
| Yavich <i>et al.</i> | Construyendo un marco de evaluación de la atención primaria de la salud para Latinoamérica | 2010 | Argentina | Artigo |
| Brizola <i>et al.</i> | Análise de desempenho de um hospital de ensino antes e após a contratualização com o Sistema Único de Saúde | 2011 | Brasil | Artigo |
| Moraes; Andrade | Indicadores de qualidade para o gerenciamento da disfagia em Unidades de Internação Hospitalar | 2011 | Brasil | Artigo |
| Ponce | Diagnóstico da tuberculose: desempenho do primeiro serviço de saúde procurado em São José do Rio Preto | 2012 | Brasil | Grey literature (Tese) |

Fonte: Elaboração própria.

No que tange à abordagem conceitual, os estudos trabalharam com a abordagem centrada nos objetivos (n=1) e sistêmica (n=5). Já nos estudos que trabalharam com a abordagem sistêmica, dois grupos distintos foram identificados. O primeiro (BRIZOLA; CORDONI JÚNIOR, 2011; MORAES; ANDRADE, 2011; PONCE, 2012) trabalhou com a concepção do sistema racional, no qual o desempenho é avaliado por normas e não apenas por resultados, mas também, por processo e/ou estrutura. E o segundo, (BATESINI, 2008; DUARTE, 2009) com a concepção de sistemas da ação social, na qual as intervenções são consideradas sistemas da ação social em constante transformação (*quadro 3*).

Quanto ao tipo de modelo os estudos se ancoraram em modelos unidimensionais

(n=4) e multidimensionais (n=2). As dimensões mais frequentemente avaliadas nos estudos unidimensionais foram estrutura, processo e resultado, dois estudos cada. Nos estudos com modelos multidimensionais, não houve predominância de nenhuma das dimensões avaliadas (*quadro 3*).

Quanto ao nível de complexidade, os estudos abarcaram intervenções com diversos graus de complexidade, desde serviços, como hospitais (n=2) e serviços públicos de saúde (atenção primária, pronto atendimento e serviço especializado) (n=1), passando por programas, como vigilância sanitária (n=1) e reabilitação da deglutição (n=1), até sistemas de saúde (n=1) (*quadro 3*).

Discussão

Apesar de o desempenho ser uma questão de grande importância para a gestão dos sistemas de serviços de saúde (VIACAVA ET AL., 2004; CHAMPAGNE ET AL., 2011; HATRY, 2013; MARCHAL ET AL., 2014), das tentativas da OMS (WHO, 200) e Opas (OPAS, 2001) em estimular o desenvolvimento de sistemas de gerenciamento do desempenho e da adoção do modelo de gestão por resultados em diversos países da América Latina, entre eles o Brasil, a Argentina e o México (DI GIÁCOMO, 2005; BASSIT; BLUM; MARTINS, 2009), poucos estudos explicitaram o modelo teórico para avaliar o desempenho de intervenções de saúde nesses países.

Diante dessa constatação, não se pode afirmar que pesquisas sobre a avaliação de desempenho não vêm sendo realizadas na região, configurando-se como um dos principais limites do estudo. Contudo, considerando a natureza complexa, paradoxal e contingente do desempenho de intervenções de saúde (SICOTTE ET AL., 1998; VIACAVA ET AL., 2004), bem como a profusão de modelos avaliativos disponíveis na literatura (VIACAVA ET AL., 2004; CHAMPAGNE ET AL., 2011), ressaltam-se a importância da delimitação teórica sobre o desempenho e a necessidade de explicitação das opções metodológicas adotadas para sua avaliação.

Esse aspecto é de grande importância diante das inúmeras definições e falta de consenso sobre o desempenho. Muitas vezes, o termo é utilizado como sinônimo de qualidade, efetividade ou eficiência, que não conseguem, sozinhos, representar todo o espectro de questões abarcadas pelo desempenho (SICOTTE ET AL., 1998; NAYLOR; IRON; HANDA, 2002; CHAMPAGNE ET AL., 2011), sendo insuficientes para promover o desenvolvimento de intervenções em saúde.

Segundo Champagne *et al.* (2011), o desempenho é um conceito genérico que envolve diversos aspectos relevantes do desenvolvimento organizacional, tais como a qualidade, a produtividade, a eficiência, sem,

contudo, se restringir a estes. Tais elementos são trabalhados como componentes ou subdimensões do desempenho, conforme observado nos modelos propostos pelo governo canadense (ARAH; WESTERT, 2005), pelo EGIPSS (SICOTTE ET AL., 1998; CHAMPAGNE ET AL., 2011) e pelo Proadess (VIACAVA ET AL., 2004).

Ademais, os modelos são representações da realidade e têm sido bastante utilizados na avaliação como forma de delimitação e descrição do objeto a ser avaliado e de obtenção de consensos e comunicação entre os envolvidos (BEZERRA; CAZARIN; ALVES, 2010; LAHREY; NIELSEN, 2013; REIS, 2012), além de ajudarem no esclarecimento das opções teóricas e na comparação entre estudos, apresentando-se como etapa fundamental na condução de avaliações orientadas pela teoria.

Outra questão de grande relevância, que pode se configurar como outro limite do estudo, é a ausência de termo específico que trate do desempenho de intervenções (programas, serviços e sistemas) de saúde nos DeCS, vocabulário estruturado utilizado para indexação e recuperação de textos científicos no Lilacs e Medline. Para superar tal dificuldade, utilizou-se uma ampla chave de busca envolvendo termos relacionados à avaliação em saúde e ao desempenho, e pesquisada a *grey literature*, como estratégia para aumentar a sensibilidade da pesquisa e captar o maior número de textos científicos (GALVÃO SAWADA; TREVIZAN, 2001), perdendo-se, contudo, a especificidade no processo de busca.

A inclusão do banco de teses da Capes como fonte de pesquisa (*grey literature*) revelou-se uma estratégia importante de captação de textos científicos (três dos seis selecionados foram provenientes dessa base de dados) por conseguir agregar em uma plataforma on-line informações bibliográficas das teses e dissertações defendidas nos programas de pós-graduação do Brasil. Sugere-se maior aproveitamento e exploração de tal fonte de dados para estudos e pesquisas, principalmente quando se procura

por informações atuais, dado o tempo necessário para a publicação de artigos científicos. Contudo, a ausência de bases de dados semelhantes nos demais países latino-americanos inviabilizou a busca desse tipo de literatura, constituindo-se uma das limitações deste estudo.

Todos os textos que preencheram os critérios de inclusão e exclusão explicitados no protocolo da pesquisa apresentaram uma boa qualidade segundo os critérios da *JCSEE* (SANTOS; NATAL, 2006). A qualidade das avaliações é uma preocupação crescente entre os estudiosos da avaliação de desempenho que vêm se dedicando à elaboração de critérios e ao desenvolvimento de meta-avaliações (ELLIOT, 2011; HARTZ, 2006). Essas meta-avaliações consistem na avaliação da qualidade das avaliações à luz de critérios bem definidos e vêm responder a uma necessidade fundamental das revisões sistemáticas, uma vez que seus resultados são dependentes da qualidade dos textos selecionados (GALVÃO SAWADA; TREVIZAN, 2001; HARTZ, 2006; ELLIOT, 2011).

Um aspecto que chama a atenção é a abrangência das intervenções em saúde (serviços, programas, sistemas) identificadas neste estudo, demonstrando a potencialidade da instituição de mecanismos para avaliar o desempenho em intervenções com diferentes graus de complexidade. Contudo, ressalta-se que a escolha das dimensões e indicadores mais apropriados, atores participantes, nível de agregação e análise dos dados modificam-se a depender do público alvo e das necessidades que a avaliação venha a suprir (COSTA ET AL., 2013; REIS, 2012; LAHREY; NIELSEN, 2013), e a escolha deve ser feita em consonância com o modelo e nível de análise desejado.

Em que pese os limites dos modelos unidimensionais para avaliar organizações de saúde (SICOTTE ET AL., 1998; CHAMPAGNE ET AL., 2011; MARCHAL., ET AL. 2014), este foi o tipo de modelo predominante encontrado neste estudo. Dentre os modelos unidimensionais, a abordagem conceitual baseada na tríade, estrutura, processo e resultados proposta

por Donabedian (2003) foi a mais frequentemente utilizada.

Revisão sistemática realizada por Klassen *et al.* (2009) estudou os modelos para medição e melhoria do desempenho de sistemas de saúde, educação e serviços sociais contidos nos artigos publicados em inglês nas bases de dados *Medline*, *CINAHL*, *Embase*, *Eric* e *PsycINFO* no período de 1986 a 2007. Esse estudo identificou o *Balanced Score Card (BSC)* como modelo multidimensional mais frequentemente utilizado, seguido pelo modelo unidimensional Donabedian. Apesar de o estudo não tratar especificamente da realidade latino-americana, identifica a preponderância do uso de modelos multidimensionais na literatura pesquisada (KLASSEN ET AL., 2009). Tais modelos têm se mostrado mais adequados para avaliar intervenções de saúde em um mundo globalizado, onde os governos necessitam lidar com realidades complexas e voláteis (BASSIT; BLUMM; MARTINS, 2009; MARCHAL ET AL., 2014) e com a dispersão do poder entre mercados e instituições internacionais (UUSIKYLÄ, 2013).

Os estudos desta revisão que utilizaram modelos multidimensionais (BATTESINI, 2008; DUARTE, 2009) adotaram a concepção de sistemas da ação social, modelos inspirados na Teoria Geral da Ação elaborada por Parsons (1977). Nessa perspectiva, o conceito de sistema se refere às estruturas sociais relativamente estáveis, dentro de um limite definido de variações compostas pelas relações entre o sistema e o meio exterior, os objetivos, a integração e a latência que corresponde à espécie de depósito de motivação acumulada e difusora de energia que assegura a fidelidade dos atores às normas e valores que o sistema inspira (PARSONS, 1977).

Esses modelos avaliam o desempenho sob a perspectiva integradora e do equilíbrio entre as funções básicas do sistema que se encontra em constante transformação (CHAMPAGNE ET AL., 2011). Além disso, a teoria da ação social propicia um meio adequado para a convergência de diversas perspectivas de análise organizacional com os vários

conceitos e dimensões do desempenho (SICOTTE ET AL., 1998; GUISET ET AL., 2002).

Os modelos multidimensionais são definidos na literatura como os mais adequados para avaliar intervenções na área da saúde devido à sua maior capacidade avaliativa e pela incorporação de relações de complementaridade entre as diferentes dimensões avaliadas (MARCHAL ET AL., 2014; REIS, 2012), dialogando com os conceitos de complexidade proposto por Morin (2012). Dessa forma, tais modelos estão mais aptos a lidar com as incertezas, ambiguidades e pluralidades dos ambientes externo e interno das organizações, nos quais deixa de existir um único padrão ordenador, mas uma necessidade crescente de combinação de alguma forma integradora de distintos padrões para o alcance dos resultados (BASSIT; BLUMM; MARTINS, 2009; CHAMPGNE ET AL., 2011; MARCHAL ET AL., 2014).

Considerações finais

O debate em torno da necessidade e da pertinência da avaliação de desempenho de sistemas, serviços e programas de saúde é uma questão que encontra certo consenso na literatura. Contudo, o aprofundamento da discussão em torno das abordagens teórico-conceituais e dos modelos mais adequados para avaliar o desempenho de intervenções de saúde ainda é necessário.

A literatura internacional, principalmente norte americana e européia, apresenta

uma profusão de modelos teóricos para avaliar o desempenho de intervenções de saúde. Na América Latina, algumas iniciativas foram realizadas no intuito de elaborar modelos condizentes com a realidade e a necessidade dos sistemas e serviços de saúde de alguns de seus países.

Apesar de tais iniciativas, nesta revisão, apenas um pequeno número de estudos explicitou o modelo de avaliação de desempenho. Esse fato traz a necessidade de maior aprofundamento no que tange ao tipo de pesquisa aqui empregado, reforçando uma maior clareza das opções teórico-conceituais adotadas. Além disso, os achados da revisão demonstram o predomínio do uso de modelos unidimensionais, que, a despeito de sua aplicabilidade e importância, são restritos e não conseguem captar facetas importantes para a melhoria do desempenho de organizações de saúde.

Diante das restrições financeiras por que passam os países do estudo, a avaliação de desempenho deve ser concebida de forma a prover a gestão com informações estratégicas, oportunas e com níveis distintos de agregação. Para tanto, torna-se necessária a explicitação do significado de desempenho, suas dimensões e componentes, bem como o arcabouço metodológico utilizado para o julgamento de valor. Assim, os modelos de avaliação de desempenho ganhariam papel de destaque por permitirem maior delimitação teórica, transparência e facilidade na condução da avaliação. ■

Referências

ARAH, O. A.; WESTERT, G. P. Correlates of health and healthcare performance: applying the Canadian health indicators framework at the provincial-territorial level. *BMC Health Services Research*, Londres, v. 5, n. 76, p. 1-13, 2005.

BASSIT, M.; BLUMM, M.; MARTINS, H. F. Governança para resultados no Ministério da Saúde. In:

MINISTÉRIO DA SAÚDE (Org.). *Mais Gestão é Mais Saúde: governança para resultados no Ministério da Saúde*. Série B. Textos básicos de saúde. Brasília, DF: Ministério da Saúde, 2009, p. 27-43.

BATTESINI, M. *Método multidimensional para avaliação de desempenho da vigilância sanitária: uma aplicação*

- em nível municipal. 2008. 210 f. Tese (Doutorado em Engenharia de Produção) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.
- BEZERRA, L. C. A.; CAZARIN, G.; ALVES, C. K. A. Modelagem de programas: da teoria à operacionalização. In: SAMICO, I.; FELISBERTO, E.; FIGUEIRÓ, A. C. *et al.* (Org.). *Avaliação em saúde: bases conceituais e operacionais*. Rio de Janeiro: Medbook, 2011, p. 65-78.
- BRIZOLA, J. B.; GIL, C. R. R.; CORDONI JÚNIOR, R. Análise do desempenho de um hospital de ensino antes e após a contratualização com o Sistema Único de Saúde. *Rev Adm Saúde*, São Paulo, v. 13, n. 50, p. 7-22, 2011.
- CARVALHO, A. L. B. *et al.* A gestão do SUS e as práticas de monitoramento e avaliação: possibilidades e desafios para a construção de uma agenda estratégica. *Ciênc. saúde colet.*, Rio de Janeiro, v. 17, n. 4, p. 901-911, 2012.
- CHAMPAGNE, F. *et al.* A avaliação no campo da saúde: conceitos e métodos. In: BROUSSELE, A. *et al.* (Org.). *Avaliação em saúde conceitos e métodos*. Rio de Janeiro: Fiocruz, 2011, p. 19-40.
- CONTRANDIOPOULOS, A. P. *et al.* A avaliação na área da saúde: conceitos e métodos. In: HARTZ, Z. M. A. (Org.). *Avaliação em saúde: dos modelos conceituais à prática na análise da implantação de programas*. 2 ed. Rio de Janeiro: Fiocruz, 1997, p. 29-47.
- COSTA, J. M. B. S. *et al.* Monitoramento do desempenho da vigilância em saúde: instrumento e estratégias de uso. *Ciênc. saúde colet.*, Rio de Janeiro, v. 18, n. 5, p. 1201-1216, 2013.
- DI GIÁCOMO, W. A. O New Public Management no Canadá e a gestão pública contemporânea. *Interfaces Brasil Canadá*, Canoas, v. 5, p. 155-170, 2005.
- DONABEDIAN, A. An introduction to quality assurance. In: BASHUR, R. (Org.). *Health care selecting approaches to assessing performance*. United Kingdom: Oxford University Press, 2003, p. 46-57.
- DUARTE, E. M. R. *Modelo sistêmico de avaliação de desempenho hospitalar por meio de indicadores assistenciais: o caso do indicador de dor aguda no pós-operatório*. 2009. 201 f. Tese (Doutorado em Engenharia de Produção) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.
- ELLIOT, L. G. Meta-avaliação: das abordagens às possibilidades de aplicação. *Ensaio: aval pol públ Educ.*, Rio de Janeiro, v. 19, n. 73, p. 941-964, 2011.
- GALVÃO, C. M.; SAWADA, N. O.; TREVIZAN, M. A. Revisão sistemática: recurso que proporciona a incorporação das evidências na prática da enfermagem. *Rev Latino-am Enfermagem*, Ribeirão Preto, v. 12, n. 3, p. 549-456, 2001.
- GUISSET, A. L.; SICOTTE, C.; LECLERCQ, P. *et al.* Définition de La Performance Hospitalière: une enquête auprès des divers acteurs stratégiques au sein des Hôpitaux. *Sciences Sociales et Santé*, Montrouge, v. 20, n. 2, p. 65-104, 2002.
- HARTZ, Z. M. A. Princípios e padrões em metaavaliação: diretrizes para os programas de saúde. *Ciênc saúde colet.*, Rio de Janeiro, v. 11, n. 3, p. 733-738, 2006.
- HATRY, H. P. Sorting the Relationships Among Performance Measurement, Program Evaluation, and Performance Management. *New Directions for Evaluation*, Honolulu, v. 137, p. 19-32, 2013.
- HOMEDES, N.; UGALDE, A. Reformas de Salud y equidad en América Latina. *Ciênc saúde colet.*, Rio de Janeiro, v. 16, n. 6, p. 2686-2696, 2011.
- HUNTER, D. E. K.; NIELSEN, S. B. Performance management and evaluation: exploring complementarities. *New Directions for Evaluation*, Honolulu, v. 137, p. 7-18, 2013.
- KLASSEN, A. *et al.* Performance measurement and improvement frameworks in health, education and social services systems: a Systematic review. *International Journal for Quality in Health Care*, Oxford, v. 22, n. 1, p. 4-69, 2009.
- LARHEY, R.; NIELSEN, S. B. Rethinking the relationship among monitoring, evaluation and results based management: observations from Canadá. *New Directions for Evaluation*, Honolulu, v. 137, p. 45-56, 2013.

- MARCHAL, B. *et al.* Building on the EGIPSS performance assessment: the multipolar framework as a heuristic to tackle the complexity of performance of public service oriented health care organizations. *BMC Public Health*, Londres, v. 14, p. 378-391, 2014.
- MORAES, D. P.; ANDRADE, C. R. F. Indicadores de qualidade para o gerenciamento da disfagia em Unidades de Internação Hospitalar. *J Soc Bras Fonoaudiol.*, São Paulo, v. 23, n. 1, p. 89-94, 2011.
- MORIN, E. *Meu Caminho*: entrevistas com Djenane Kareh Tager. São Paulo: Bertrand Brasil, 2012.
- NAYLOR, C. D.; IRON, K.; HANDA, K. Measuring health system performance: problems and opportunities in the era of assessment and accountability. In: ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OCDE). *Measuring up: improving health system performance in OCDE countries*. Paris: OCDE, 2002, p. 13-34.
- ORGANIZACIÓN PANAMERICANA DE LA SALUD (OPAS). *Health systems performance assesment and improvmnt in the region of Americas*. Washginton, DC: OPAS, 2001.
- ORGANIZACIÓN PANAMERICANA DE LA SALUD (OPAS). *Salud en las Américas: panorama regional y perfíles de país*. Washington, DC: OPAS, 2012.
- PARSONS, T. *Social Systems and the evolution of action theory*. New York: Free Press, 1977, 420 p.
- PONCE, M. A. Z. *Diagnóstico da tuberculose: desempenho do primeiro serviço de saúde procurado em São José do Rio Preto*. 2012. 97 f. Tese (Doutorado em Ciências) – Universidade de São Paulo, Ribeirão Preto, 2012.
- REIS, A. C. G. V. *A noção de equilíbrio como Proxy da avaliação de desempenho de sistemas de saúde*. 2012. 103 f. Tese (Doutorado em Saúde Pública) – Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo cruz, Rio de Janeiro, 2012.
- SANTOS, E. M.; NATAL, S. *Um guia para conhecer modelos teóricos de avaliação*. Rio de Janeiro: Fiocruz, 2006. (Série Ensinando Avaliação).
- SICOTTE, C.; CHAMPAGNE, F.; CONTANDRIOPOULOS, A. P. *et al.* A conceptual framework for health care organizations Performance. *Health Serv. Manage. Res.*, London, v. 11, p. 24-38, 1998.
- SOUZA, A. América Latina, conceito e identidade: algumas reflexões da história. *PRACS: revista de Humanidades do Curso de Ciências Sociais da UNIFAP*, Macapá, v. 4, p. 29-39, 2011.
- TCHOUAKET, E. N. *et al.* Health Care System Performance of 27 OCDE Countries. *Int. J. Health Plann. Mgmt.*, Kerman, v. 14, n. 2, p. 104 -129, 2012.
- UNITED STATES AGENCY FOR INTERNATIONAL DEVELOPMENT (USAID). *Evaluation Policy*. United States: United States Agency for International Development, 2011, 14 p.
- UUSIKYLÄ, P. Transforming silo-steering into a performance governance system: The case of the Finnish central government. *Performance management and evaluation. New directions for evaluation*, Honolulu, v. 137, p. 7-17, 2013.
- VIACAVAL, F. *et al.* Uma metodologia de avaliação do desempenho do sistema de saúde brasileiro. *Ciênc. saúde colet.*, Rio de Janeiro, v. 9, n. 3, p. 711-724, 2004.
- YAVICH, N.; BÁSCOLO, E. P.; HAGGERTY, G. Construyendo un marco de evaluación de la atención primaria de la salud para Latinoamérica. *Salud Pública de México*, Morelos, v. 52, n. 1, p. 39-45, 2010.
- WORLD HEALTH ORGANIZATION (WHO). *The world health report 2000: health system: Improving performance*. Genebra: WHO, 2000.

Recebido para publicação em fevereiro de 2015
Versão final em outubro de 2015
Conflito de interesses: inexistente
Suporte financeiro: não houve