

Validação e uso de painéis de indicadores de desempenho no SUS aplicados aos resultados de unidades de Atenção Primária

Galba Freire Moita

RESUMO

Após 30 anos do SUS ainda não se registra nenhum sistema nacional institucionalizado e contínuo de avaliação de resultados e desempenho em saúde pública. A Nova Gestão Pública ou Gerencialismo, exige dos gestores esforços para monitorar resultados, controle e prestação de contas (accountability). Método: através de painéis colaborativos explora-se a translação de conhecimentos com 220 gestores e profissionais de saúde de 11 estados brasileiros e captou-se valorações de itens de avaliação no contexto do SUS. Resultados: A aplicação dos instrumentos de captação e validação das percepções obteve resultados das três fases iniciais (n=108) que consolidaram um instrumento aplicado para validação de campo (n=112), cuja análise descritiva validou que demonstrou elevados (> 4) graus de importâncias/relevâncias das subcategorias (20 e 24 domínios de avaliação) e das categorias (cinco e seis macrodomínios de avaliação). Discussão: A análise inferencial gerou um modelo final que garantiu confiabilidade e validade de 17 domínios (indicadores) e quatro macrodomínios (dimensões) de avaliação de desempenho conforme os conhecimentos prévios e valorações dos decisores e validou a usabilidade de um proxy de painel de indicadores e dimensões para unidades do SUS.

Palavras-chave: Direção e Governança do Setor de Saúde; Avaliação de Processos e Resultados (cuidado de saúde); Avaliação em Saúde; Qualidade, Acesso e Avaliação da Assistência à Saúde; Avaliação do Impacto na Saúde.

ABSTRACT

Introduction: After 30 years of SUS, there is still no institutionalized and continuous national system of evaluation of results and performance in public health. The New Public Management or Managerialism, managers are required to monitor results, control and accountability. Method: through collaborative panels, the translation of knowledge with 220 managers and health professionals from 11 Brazilian states is explored, and valuations of evaluation items were captured in the context of SUS. Results: The application of the instruments for capturing and validating perceptions obtained results from the three initial phases (n = 108) that consolidated an instrument applied for field validation (n = 112), whose descriptive analysis validated that it demonstrated high (> 4) degrees of importance / relevance of the subcategories (20 and 24 evaluation domains) and of the categories (five and six evaluation macro domains). Discussion: The inferential analysis generated a final model that guaranteed the reliability and validity of 17 domains (indicators) and four macrodomains (dimensions) of performance evaluation according to the prior knowledge and valuations of the decision makers and validated the usability of an indicator panel proxy and dimensions for SUS units.

Keywords: Health Sector Stewardship and Governance; Process and Results Evaluation (health care); Health Assessment; Quality, Access and Evaluation of Health Care; Health Impact Assessment.

Revista da Rede APS 2022

Publicada em: 23/09/2022

DOI:10.14295/aps.v4i2.250

Galba Freire Moita
(Universidade de Coimbra,
Faculdade de Economia,
Centre for Business and
Economics Research (CeBER),
Coimbra, Portugal.

Fundação Oswaldo Cruz,
Ministério da Saúde, Brasília,
Brasil.)

Correspondência para:

Galba Freire Moita
prgalba@gmail.com

INTRODUÇÃO

Desde a criação do Sistema Único de Saúde (SUS) no Brasil, muitos são os avanços do ponto de vista de garantias legais (Brasil, 1988, 1990a, 1990b) de atendimento, da organização sistêmica e da descentralização da gestão única para os níveis estadual e municipal da saúde. Na atenção primária da saúde (APS) do Brasil tem sido relacionados a expansão das equipes de saúde da família (ESF) e a redução de internações hospitalares por condições sensíveis à atenção primária em crianças (Victora, 2011) e em adultos (Barreto et al., 2011), a redução de consultas não-urgentes nas emergências (Stein et al., 2002) e a redução da mortalidade infantil (Dourado et al., 2011; Macinko et al., 2007; Rasella et al., 2010), além da ampliação de controle das doenças cardiovasculares (Barreto et al., 2014; Rasella et al., 2014).

Dados relativos a 2016 apontam que 35 países da OCDE gastam em torno de 9,0% do PIB no sistema de saúde, nos Estados Unidos cerca de 17,2%, em Portugal 8,9%, e no Brasil 6,2% (OECD, 2017). Além disso, os gastos com a saúde têm crescido nos últimos anos (2003-2016) em média de 3,6% nestes países (OECD, 2017). No Brasil, os gastos no consumo de bens e serviços de saúde atingiram R\$ 608,3 bilhões em 2017 (9,2% do PIB) comparados a 8,0% do PIB em 2010.

Apesar do aporte destes recursos, potenciais ineficiências do SUS precisam ser estudadas visto que repercute em filas nacionais de cirurgias eletivas de cerca de 667.014 pacientes (Frasão, 2017), além do que a mídia brasileira tem mostrado a ampliação das taxas de mortalidade (Barreto, 2016).

Na saúde, o aporte crescente de recursos aplicados, por si, exige avanços no monitoramento de resultados quanto à eficiência e efetividade das políticas e programas.

Por sua vez, uma tendência gerencial recente tem exigido dos gestores públicos esforços de monitoramento de resultados em diversas áreas para a controladoria e responsabilização (*accountability*) dos gastos e recursos públicos.

Neste contexto de elevação de gastos e conforme a visão da NGP ou Gerencialismo, surge a necessidade de avaliação e monitoramento de resultados do sistema de saúde.

A inclusão dos diversos atores na decisão da modelagem da avaliação se faz necessária também pela variedade de dimensões mensuráveis nos serviços de saúde, e pela necessidade de priorização pelos seus futuros utilizadores. Neste quesito, Uchimura e Bosi (2002) referem autores e listas das possíveis dimensões de avaliação. A multiplicidade de perspectivas e de complexidades envolvidas, além das diversidades de escopos e definições demandam a análise dos principais quadros teóricos e metodológicos de avaliação relevantes da literatura. Não há na literatura um consenso sobre escopo e amplitude da avaliação e do monitoramento em saúde, mas existem muitos possíveis quadros teóricos conceituais relevantes (G. F. Moita, V. M. R. Raposo, & A. C. Q. Barbosa, 2019b) pelo que Arah et al. (2003) apresentam as abordagens de quadro teórico-conceitual (*framework*) dos sistemas de medição de performance do Reino Unido (NHS), do Canadá, da Austrália e dos EUA.

No contexto brasileiro, procedeu-se uma revisão sistemática ampliada nas principais bases publicações de artigos científicos do Brasil (SciELO e Lilacs e Medline, via Biblioteca Virtual em Saúde - BVS) tendo localizado 112 publicações no período de 1990 a 2018 (G. F. Moita, V. M. d. R. Raposo, & A. C. Q. Barbosa, 2019), sendo mais da metade classificada como de análise estratégica e de avaliação de implementação de políticas e pouco mais de uma dúzia de estudos sobre eficiência e resultados em saúde, sendo um campo incipiente.

O modelo de financiamento em saúde ao longo dos anos tem sido utilizado como ferramentas capazes de auxiliar na melhoria de saúde das populações, uma vez que quanto mais investimento deveriam ser aplicados na melhoria dos níveis de atenção à saúde e para ajudar os gestores no planejamento e na avaliação em saúde (Arretche, 2003; Santana &

Costa, 2008; Santana, 2011; Vieira & Zucchi, 2013). E como resultado da obtenção desse cuidado, a representação e a implementação de um adequado sistema de financiamento que são essenciais na busca de cobertura universal (Brasil, 2009, 2016).

Assim, há um dilema recorrente no SUS: há insuficiência de recursos ou ineficiência na gestão? Surge a questão central: Os conhecimentos dos gestores e profissionais do SUS podem contribuir com a validação de uma proposta de indicadores de desempenho nas unidades da APS e no SUS? Essa questão demanda uma metodologia robusta de avaliação e monitoramento de resultados, visando estabelecer traçadores (dimensões e indicadores de desempenho) de resultados. Dessa forma, este estudo envolveu e mapeou os conhecimentos de uma amostra multicêntrica de profissionais e gestores do SUS para validação e legitimação do uso de um modelo aplicado de mensuração da eficiência e demais dimensões de resultados em unidades de saúde, para a gestão de performance de unidades do SUS.

Nesse escopo, pesquisas robustas apontam que são incipientes os instrumentos de avaliação de no âmbito do SUS, inclusive na APS (Albuquerque & Martins, 2017; Almeida & Giovanella, 2008a; Moita et al., 2019; Volpato, 2014; Volpato et al., 2010). Assim, justifica-se a importância deste estudo avaliativo, pois a avaliação é vista não só como instrumento para verificar a efetividade de programas, mas utilizada também com o intuito de melhorar a qualidade dos serviços por um processo de julgamento de valor orientado para a tomada de decisão e governança (Moita, 2021; Raposo, 2007).

No entanto, não há registro de avaliação de impactos resultantes nessas decisões. O objetivo desse estudo é propor um modelo avaliativo quanto ao impacto, eficiência, efetividade e resolutividade de alguns regimes de financiamento e modelos gestão da APS,

instrumentalizando a decisão nos três níveis de gestão do SUS. Para tal, pretende-se mapear e analisar indicadores explicativos, por exemplo quanto as condições de saúde sensíveis a atenção primária (Alfradique et al., 2009), dentre outros sugeridos nesse trabalho.

A AVALIAÇÃO DA APS NO SUS

No contexto de avaliação de resultados da APS, em 2011, no bojo da nova PNAB, o Ministério da Saúde lançou o Programa de Avaliação para a Qualificação do Sistema Único de Saúde (SUS) (Brasil, 2011, 2012, 2015), instituindo o instrumento de Autoavaliação para a Melhoria do Acesso e da Qualidade da Atenção Básica (AMAQ) (Brasil, 2012). Nesse escopo, o PMAQ, que teve três ciclos (2012, 2014 e 2015/2017), tem possibilitado um melhor entendimento da estrutura e forma de organização da ESF, sendo um banco de dados importante neste aspecto. O ciclo inicial do PMAQ voltou-se às unidades de atenção básica e equipes ESF. O 2º. ciclo do PMAQ (2014) ampliou o escopo para os centros de especialidades odontológicas. Em seguida, o 3º. ciclo do PMAQ (2015/2017), contou pela primeira vez com um quadro de indicadores de resultados das unidades (Brasil, 2017a, 2017b).

Em 2019, o MS estruturou o Programa Previne Brasil (Brasil, 2019). Harzheim (2020) defendeu a radicalização dos atributos da APS, o financiamento por capitação ponderada para induzir a vinculação dos usuários da APS para indução à responsabilização dos gestores e dos profissionais pelas pessoas que assistem, além da definição de carteira de serviços da APS. A operacionalização do Previne Brasil está focada em resultados na avaliação da APS por meio de três macro indicadores: o Net Promoter Score (NPS)¹ - mede afiliação, fidelização e confiança (longitudinalidade) entre as pessoas e as equipes de APS, O PDRQ-9 (Wollmann et al., 2018) - mede a qualidade e a confiança da relação entre as pessoas e o médico de família e comunidade (longitudinalidade). E o PCATool-Brasil - mede a presença e a extensão dos quatro atributos essenciais (acesso de primeiro

¹ <https://endeavor.org.br/estrategia-e-gestao/nps/>

contato, longitudinalidade, integralidade e coordenação), além de dois atributos derivados (orientação comunidade e centralidade na família) (Harzheim, 2020). Segundo (Lima, 2016) o Instrumento PCATool mede a presença e a extensão dos 4 atributos essenciais e dos 3 atributos derivados da APS, baseando-se em um modelo de avaliação da qualidade de serviços de saúde, de acordo com aspectos de estrutura, processo e resultados de Donabedian. As variáveis dos três ciclos do PMAQ também estão voltadas para esse tripé ou para algum aspecto normativo.

O Previne Brasil instituiu apenas sete indicadores da clínica/epidemiologia para o “Pagamento por Desempenho”, relativos às ações estratégicas de saúde da mulher, pré-natal, saúde da criança e hipertensão arterial e diabetes mellitus para 2020/2021, projetando atingir 21 indicadores até 2022. Em sua concepção, pretendia medir aspectos de resultados e desempenho das equipes da APS. No entanto, nenhum desses três instrumentos citados no programa aborda os aspectos de mensuração de impactos e efeitos comparativos na população assistida, mostrando a necessidade de novos estudos nessa área.

No que se refere à efetividade, a *Joint Commission on Accreditation of Healthcare Organizations* (JCAHO, 1993) a define como o grau em que uma determinada intervenção ou tecnologia médica traz benefícios para indivíduos de uma população definida, sob condições regulares de uso. Assim, Ugá *et al.* (2013) afirmam que a efetividade pode ser mensurada a partir dos resultados da intervenção do sistema de serviços sobre a saúde da população.

Por sua vez, a resolutividade envolve aspectos relativos à demanda, à satisfação do cliente, às tecnologias dos serviços de saúde, à existência de um sistema de referência preestabelecido, à acessibilidade dos serviços, à formação dos recursos humanos, às necessidades de saúde da população, à adesão ao tratamento, aos aspectos culturais e socioeconômicos da clientela, entre outros (Turrini *et al.*, 2008).

Diante das limitações dos modelos de medidas unidimensionais da *eficácia* ou *performance*

organizacional (Brunet *et al.*, 1991 apud Carvalho, 2007: 164) prosperou variados modelos operacionais *multidimensionais*, sendo os principais caracterizados por Carvalho (2007). Assim, pode-se explorar os diversos indicadores socioeconômicos, os quais condicionam aspectos epidemiológicos da população, bem como direcionam a política pública em saúde.

Quanto à avaliação de impacto existem diversas possibilidades consolidadas, por exemplo, a regressão segmentada de dados de séries temporais, conhecida como *Interrupted Time Series Analysis* (ITSA) ou análise de *séries temporais interrompidas* (ITS) (Mcdowall *et al.*, 2019: 17). A ITS permitem avaliar, em termos estatísticos, quanto uma intervenção mudou um resultado de interesse, imediatamente e ao longo do tempo e se outros fatores além da intervenção poderiam explicar a mudança, tendo sido usado para avaliar diversos aspectos do programa Farmácia Popular (Silva *et al.*, 2017).

METODOLOGIA

A pesquisa é observacional, transversal, por método misto de avaliação. Apoiou-se em estudo de casos múltiplos, em uma amostra de informantes selecionados por conveniência, no âmbito do PhD da Universidade de Coimbra (CAAE: 54972816.9.0000.5051) (Moita, 2019).

A parte empírica do estudo foi desenvolvida em diversas fases subsequentes para estruturação, aprimoramento e validação incrementais dos instrumentos de coleta de dados (ciclo 1), no período de abril a setembro de 2017, com futura pesquisa de campo (ciclo 2). A partir de um modelo lógico (Souza, Vieira-Da-Silva, & Hartz, 2005), no qual se destacam: eficácia, eficiência, efetividade, impacto, qualidade técnico-científica, percepção do usuário (satisfação e aceitabilidade), dentre outros, consultou-se um painel de 25 especialistas estruturou e depois se aplicou questionários semiestruturados (TGN - 53 especialistas), seguidos pela validação (*Delphi* - 55 gestores) e aplicação de campo (112 gestores). Aplicou-se a análise descritiva e inferencial para as respostas obtidas em 12 estados do Brasil (Ceará - CE, Rio Grande do

Norte - RN, Piauí - PI, Maranhão - MA, Bahia - BA, Pará - PA, Amazonas – AM, Rondônia – RO, Mato Grosso – MT, Distrito Federal - DF, Goiás – GO e Paraná – PR).

Foi realizada consulta às preferências dos decisores-chave em duas rodadas: de validação de face e; de valoração de itens. Os informantes-chave foram instados a classificar o grau de usabilidade (importância/relevância), das macrodimensões e indicadores de performance pelos níveis governamentais/sistêmicos, organizacionais e gerenciais/chefia setoriais, com base na ponderação de itens por uma escala *Likert* de 1 (nada importante) até 5 (extremamente importante).

Aplicou-se análise descritiva dos dados integrados das fases 1 a 3 (n= 108) e análise descritiva e inferencial das respostas da fase 4 (n=112). Em cada fase, as respostas foram consolidadas com o software Excel 2016, traçados gráficos de valores e aplicadas as análises descritivas e inferenciais com o software R (versão 3.3.2), com versões de licenças livres.

RESULTADOS E DISCUSSÃO

Na exploração inicial dos conhecimentos prévios dos agentes de decisão, seguiu-se os princípios de Sampieri et al. (2006) que sugerem a amostragem por conveniência quando a investigação está centrada em indivíduos ou grupos portadores de um ou vários atributos que possam contribuir para o desenvolvimento de uma teoria, apesar da limitação de generalização de resultados.

Em virtude do desalinhamento de termos propostos pelos decisores consultados manteve-se nomenclaturas genéricas para as categoriais de avaliação (macrodomínios ≈ dimensão) e para as subcategorias de avaliação (domínios ≈ indicadores). Pode-se consultar os conceitos dos macrodomínios e domínios das figuras 1 a 4 no trabalho original de Moita (2019). Os resultados foram equalizados pela literatura vigente e geraram um novo instrumento contendo indicadores e dimensões de performance para serem utilizados em fases posteriores.

VALIDAÇÃO POR ANÁLISE DESCRITIVA

O perfil da amostra de profissionais e gestores foi considerado adequado, pois em todas as fases, a maioria dos indivíduos (38,46%, fase 1 a 3; 76,58%, fase 4) possuía pós-graduação e ainda 34,62% (fase 1 a 3) e 12,61% (fase 4) possuía nível superior. Respectivamente, nas fases 1 a 3 e na fase 4, a maioria dos indivíduos (49,07%; 70,27%) possuía mais de 60 meses de experiência; boa parte dos indivíduos (20,37%; 17,92%) trabalhava há mais de 60 meses em cargo/função de decisão, enquanto 42,59% e 13,21% não possuía experiência em cargo/função de decisão. A formação/profissão mais frequente entre os indivíduos foi a de Enfermeiro (23,36%; 40,18), seguida por Administrador Hospitalar (22,43%; 17,86%) e Assistente Social (10,28%; 9,82%).

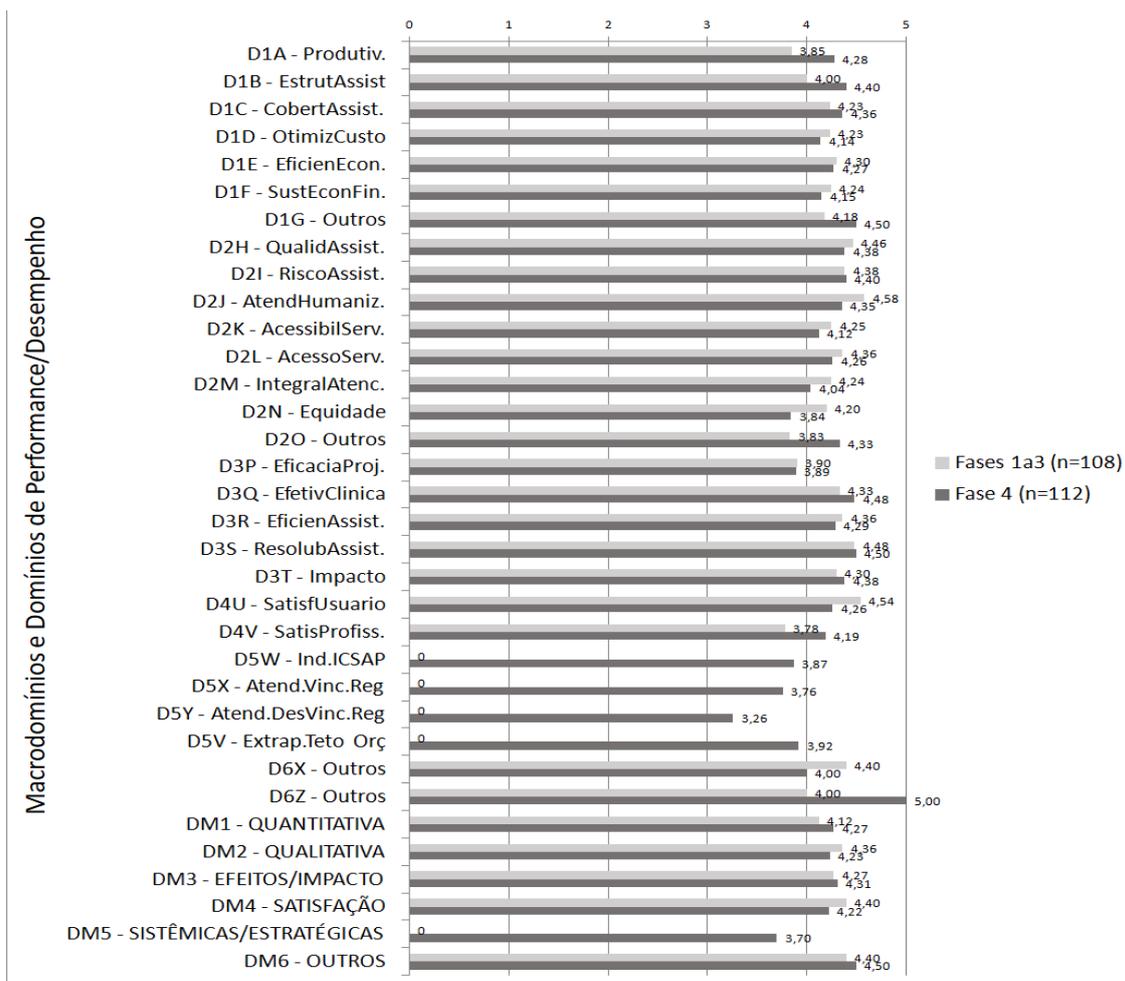
Os domínios e macrodomínios que obtiveram elevados graus de importâncias/relevâncias elevados (superior ou cerca de 3) nas respostas das fases 1 a 3, pautados exclusivamente nos conhecimentos prévios dos participantes, foram reavaliados a partir de consulta à literatura, depois categorizados e readequados conforme as macrodimensões e indicadores de performance definidos pela literatura mais relevante (Acurcio, Cherchiglia, & Santos, 1991; Akerman & Nadanovsky, 1992; Gattinara, Ibacache, Puente, Giaconi, & Caprara, 1995; Santos, 1995; F Viacava et al., 2004; Vieira-Da-Silva, 2005; Vuori, 1991b) para subsidiar o ajuste de um questionário semiestruturado a ser aplicado e validado na fase de pesquisa de campo (fase 4). Por fim, na fase 4, organizou-se outros três grupos pilotos de informantes chaves, gestores e profissionais de saúde, para a aplicação de um instrumento ajustado, em uma amostra de 112 gestores e profissionais, que emergiu dos resultados obtidos das fases anteriores.

Desta análise descritiva sintetizou-se que os especialistas e profissionais atribuíram elevados (superior ou cerca de 4) graus de importâncias/relevâncias das subcategorias (20 e 24 domínios de avaliação) e das categorias (cinco e seis macrodomínios de avaliação) dos itens de análise, sendo considerados

relevantes/importantes para a estruturação de uma proposta de um *proxy* de painel de indicadores e dimensões (figuras 1, 2 e 3). Além disso, em média, as valorações da ampla maioria destes itens de análise foram razoavelmente

dentro das mesmas margens de valores, uma vez que a maioria dos intervalos de confiança se sobrepuseram.

Figura 1 - Grau de importância/relevância de domínios e macrodomínios de avaliação (fases 1a 3 vs fase 4)



Fonte: elaborado pelos autores a partir de Moita, 2019.

Além disso, em média, as valorações da ampla maioria destes itens de análise (domínios e macrodomínios) foram razoavelmente dentro das mesmas margens de valores, uma vez que a maioria dos intervalos de confiança se sobrepuseram (gráfico 2, fases 1 a 3 e gráfico 3, fase 4). Isto sugere a expansão da pesquisa através da aplicação de um novo instrumento contendo indicadores e dimensões de performance, em fases posteriores da pesquisa.

A análise global permite afirmar que todos os quatro macrodomínios de avaliação apresentaram valoração média elevada (superior ou cerca de 4, muito importante/relevante), nas proxies A e B (n=108) e no *proxy* C (n = 112), além do quinto macrodomínio (Sistêmico/Estratégico) adicional, composto por agregações de quatro novos domínios de avaliação (D5W_IndICSAP,

D5X_AtendVincReg, D5Y_AtendDesVincReg, D5V_ExtrapTetoOrç).

Figura 2 - Grau de importância/relevância e intervalo de confiança de domínios e macrodomínios de avaliação (fases 1 a 3)

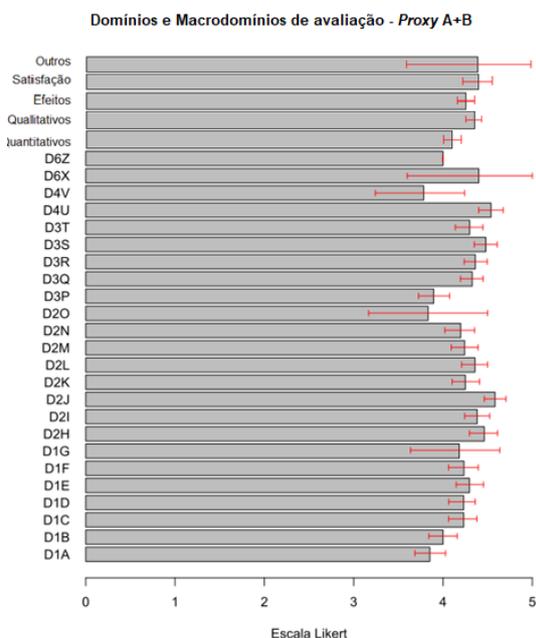
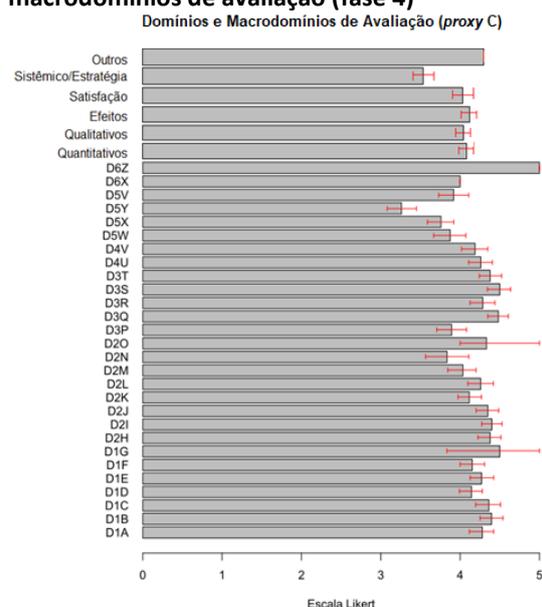


Figura 3 - Grau de importância/relevância e intervalo de confiança de domínios e macrodomínios de avaliação (fase 4)



Esses resultados têm validade interna. Há um nível razoável de importância relativa atribuída a todos os itens avaliados (variação de 3,5 a 4,0, ou seja, entre muito importante e

extremamente importante) em consonância com a literatura (Akerman & Nadanovsky, 1992; Gattinara et al., 1995; Vuori, 1991a).

VALIDAÇÃO POR ANÁLISE INFERENCIAL

Após a análise descritiva, iniciou-se a análise inferencial, aplicada exclusivamente nas respostas da fase 4 (n=112), que se dividiu em quatro análises de resultados: Estudo das cargas fatoriais dos domínios versus macrodomínios de avaliação; Validação por critérios de qualidade e validade dos macrodomínios de avaliação; Análise da qualidade (ajustamento) do modelo final resultante e Análise das correlações entre as variáveis e suas influências na Performance Global Percebida.

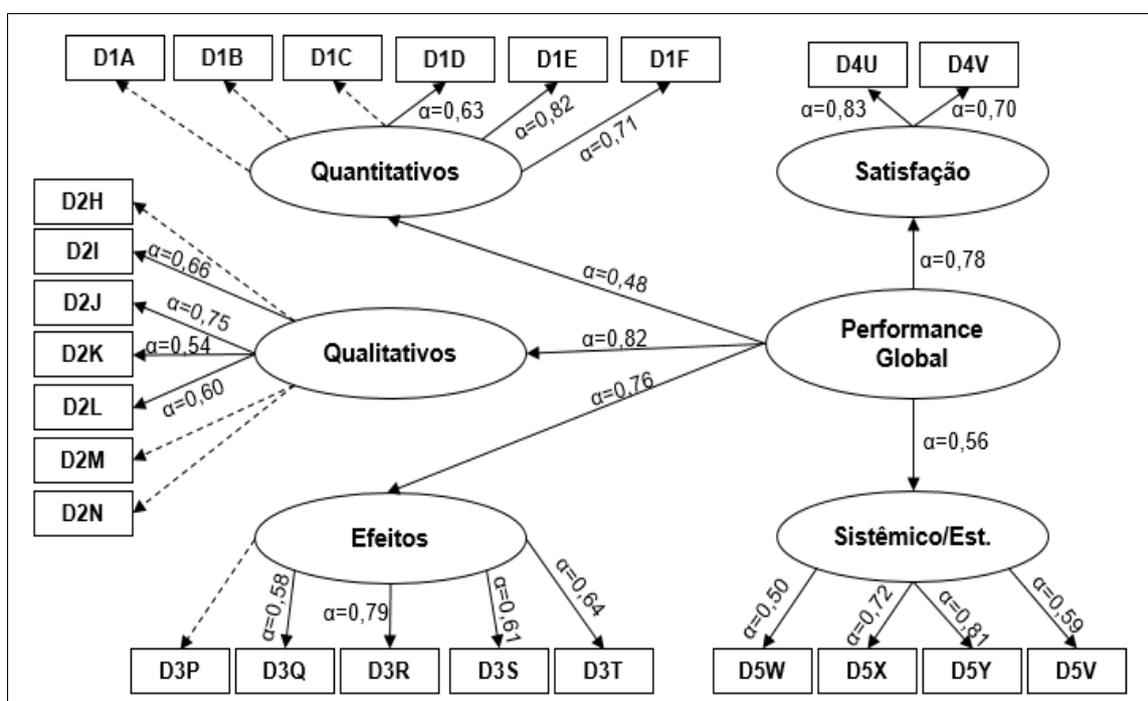
Para analisar a qualidade e validade dos macrodomínios de avaliação, foi verificada a dimensionalidade, confiabilidade e validade convergente. A dimensionalidade foi verificada pelo critério da Kaiser, que retorna a quantidade de dimensões de constructos, baseado na quantidade mínima de variância explicada por cada dimensão. Para a validade discriminante foi utilizado o critério de Fornell e Larcker (Fornell & Larcker, 1981), que garante a validade discriminante quando a Variância Média Extraída (AVE) de uma variável latente for maior que a Variância Compartilhada (VCM) com as demais. A confiabilidade foi mensurada pelo Alfa de Cronbach (A.C.) e Confiabilidade Composta (C.C.). De acordo com Tenenhaus et al. (Tenenhaus, Esposito Vinzi, Chatelin, & Lauro, 2005) os indicadores A.C. e C.C. devem ser maiores que 0,7 para garantir a confiabilidade, mas em pesquisas exploratórias valores acima de 0,6 são aceitos. Para verificar a validade convergente foi utilizado o critério proposto por de Fornell e Larcker (Fornell & Larcker, 1981). Ele garante tal validade caso a AVE sejam superiores a 50% (Henseler, Ringle, & Sinkovics, 2009), ou 40% no caso de pesquisas exploratórias (Nunnally & Bernstein, 1994; Nunnally & Berstein, 1994).

Dessa forma, a AFC da valoração média obtida para os macrodomínios de avaliação na proxy C (fase 4) destaca que no modelo final, após análise e exclusão de sete domínios de avaliação

(D1A_Produtiv, D1B_EstrutAssist, D1C_CobertAssit, D2H_QualidAssist, D2M_IntegralAtenc, D2N_Equidade e D3P_EficaciaProj), todos os demais 17 domínios que permaneceram na análise, apresentaram carga fatorial superior a 0,5 (ou suas permanências não impediram a convergência da AFC). O macrodomínio de maior peso foi Satisfação (27%) e o de menor peso foi

Quantitativa (15%), sendo que esse apresentou carga fatorial igual a 0,48. Todos os cinco macrodóminios (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico) permaneceram na análise por não impedir a convergência do modelo. Diante destes resultados, a figura 3 sintetiza o ajuste da AFC e a modelagem para a variável latente PGP.

Figura 4 - Grau Análise Fatorial Confirmatória do desempenho global percebido - Proxy C (fase 4)



Fonte: elaborado pelo autor.

No modelo final, cinco dos 24 domínios de avaliação (indicadores) propostos pelo painel de especialistas apresentaram cargas fatoriais validadas elevadas (0,76 a 1,00), que impactam fortemente seus respectivos macrodóminios, e outro grupo de 12 domínios com cargas fatoriais medianas (0,50 a 0,75) enquanto sete outros (D1A, D1B, D1C, D2H, D2M, D2N e D3P) foram excluídos do modelo final. Nenhum dos domínios vinculados aos macrodóminios Satisfação e Sistêmico/Estratégico foram eliminadas no modelo final, revelando ampla concordância entre as propostas do painel de especialistas e as valorações dos grupos pilotos. Em seguida a análise de carga fatorial, validou no modelo final os cinco macrodóminios de avaliação (Quantitativo quase no limite de 0,50),

sendo que três macrodóminios (Qualitativo, Efeitos e Satisfação) tiveram elevado nível de carga fatorial (0,76 a 1,00), tendo os outros dois (Quantitativo e Sistêmico/Estratégico) apenas carga fatorial mediana (0,50 a 0,75), com impactos na variável latente PGP.

De forma complementar, procedeu-se a análise das correlações entre as variáveis estudadas e suas possíveis influências na variável latente PGP, tendo ficado provado que a maioria dos domínios estavam fortemente correlacionados às respectivas macrodimensões de avaliação. Dessa forma, concluiu-se a análise inferencial das respostas da fase 4, proxy C (112 informantes).

CONSIDERAÇÕES FINAIS

Os resultados deste estudo, com viés de gerencialismo (accountability), além de enfatizar a avaliação dos resultados em saúde (Fletcher, 1993) oferece painéis de indicadores (F Viacava et al., 2004; Francisco Viacava et al., 2012) para monitoramento dos serviços do SUS (Silva & Formigli, 1994), tendo subsidiado a construção de um novo instrumento, alinhado pela literatura relevante de indicadores e dimensões de performance (Arah, Klazinga, et al., 2003), que será aplicado na fase posterior da investigação, em planejamento de pesquisa de campo.

O processo resultou em uma proposta de um proxy de painel de indicadores e dimensões. Destacam-se os aspectos inéditos na literatura de avaliação de performance, da colaboração com decisores-chaves na valoração da relevância/importância e estruturação de um proxy de painel de indicadores e dimensões de performance do SUS, além da validação multicêntrica de modelos finais através de análise de cargas fatoriais, validade e confiabilidade dos itens, e, ainda, dos testes de possíveis fatores mediadores na valoração. Garantiu-se a confiabilidade e a validade de 17 domínios (indicadores) e de quatro macrodomínios (dimensões) de performance, além de modelar os impactos de seus reflexos sobre a PGP no modelo final (gráfico 3).

Uma dificuldade destacada foi a incipiência de publicações em avaliação e monitoramento de sistemas e serviços de saúde no Brasil, que localizou apenas 112 publicações no período de 1986 a 2018 (G. F. Moita, V. M. R. Raposo, & A. C. Q. Barbosa, 2019a), além da incipiência de modelos estruturados de avaliação de desempenho limitando-se a algumas experiências isoladas, por exemplo, PROADESS, IDB, RIPSA, PNASS, IDSS, PROHASA/CQH e SIPAGEH, sintetizadas por Moita (2019). Estas incipiências de publicações resultaram em dificuldades de comparação dos resultados obtidos nesta investigação.

Para a continuidade desta investigação pode-se sugerir a análise da generalização dos resultados

obtidos pela aplicação dos artefatos gerados e validados em outros extratos amostrais, preferencialmente por amostragem aleatória de gestores, profissionais e usuários. Propõem-se a ampliação da amostra em todos os extratos amostrais, a fim de permitir outros processos analíticos e comparações entre diferentes classes de gestores, profissionais e usuários, bem como de diferentes tipos de unidades de saúde, contribuindo para a mensuração da performance em serviços de saúde por um método de seleção de itens de avaliação a partir das valorações de especialistas, profissionais e gestores.

REFERÊNCIAS

- Acurcio, F. A., Cherchiglia, M. L., & Santos, M. A. (1991). Avaliação de qualidade de serviços de saúde. *Saúde em debate*, 33, 4.
- Akerman, M., & Nadanovsky, P. (1992). Evaluation of Health Services - What to Evaluate? *Caderno de Saúde pública*, 8(4), 5.
- Alkin, M. C., & Christie, C. A. (2004). *Evaluation Roots: Tracing Theorist's Views and Influences*. Sage publications.
- Arah, O. A., Custers, T., & Klazinga, N. S. (2003). *Updating the Key Dimensions of Hospital Performance: The Move Towards a Theoretical Framework* Paper presented at the 3rd Workshop on Hospital Performance Measurement Barcelona.
- Arah, O. A., Klazinga, N. S., Delnoij, D. M. J., Asbroek, A. H. A. T., & Custers, T. (2003). Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement *International Journal for Quality in Health Care*, 15(5), 22. doi:10.1093/intqhc/mzg049
- Arah, O. A., Westert, G. P., Hurst, J., Niek, S., & Klazinga, N. S. (2006). A conceptual framework for the OECD Health Care Quality Indicators Project *International Journal for Quality in Health Care*, 9. doi:10.1093/intqhc/mzl024
- Barbosa, L., & Neto, A. P. (2017). Ludwik Fleck (1896-1961) e translação do conhecimento: considerações sobre a genealogia de um conceito. *Saúde Debate*, 41(Especial), 13.
- Barreto, E. (2016). Rio tem a maior taxa de mortalidade no SUS em três décadas. *O Globo - Caderno Brasil*. Retrieved from <http://oglobo.globo.com/brasil/rio-tem-maior-taxa-de-mortalidade-no-sus-em-tres-decadas-18770526#ixzz4OrM31ntY> [29 de fevereiro de 2016]
- Bosi, M. L., & Mercado, F. J. (2006). *Pesquisa qualitativa de serviços de saúde* (Vozes Ed.).
- Brasil. (1988). *Constituição da República Federativa do Brasil de 1988*. Distrito Federal: Brasília: Congresso Nacional
- Brasil. (1990a). *Lei 8.080, de 19 de setembro de 1990*. Distrito Federal: Brasília: Congresso Nacional
- Brasil. (1990b). *Lei 8.142, de 12 de novembro de 1990*. Distrito Federal: Brasília: Congresso Nacional
- Bryson, J. (2004). What to do when stakeholders matter: Stakeholder identification and analysis techniques. *Public Management Review*, 6(1), 33.
- Carvalho, C. (2007). *Organizações, actores envolvidos e partes interessadas: Determinantes da saliência dos stakeholders e sua relação com o desempenho organizacional*. (Doctoral). Universidade de Coimbra, Coimbra, Portugal.
- Chianca, T., & Youker, B. (2004). *La evaluación em América Latina y El Caribe. Visión general de los desarrollos recientes*.
- CIHR, C. I. o. H. R. (2004). *Knowledge Translation Strategy 2004-2009: Innovation in Action*. Retrieved from Ottawa: <http://www.cihr-irsc.gc.ca/e/26574.html>
- Clavier, C., Sénéchal, Y., & Potvin, L. (2011). A theory - based model of translation practices in

public health participatory research. *Sociology of Health and Illness*, 34(5), 16.

Fals Borda, O., & Mora-Osejo, L. E. (2004). La superación del Eurocentrismo. Enriquecimiento del saber sistémico y endógeno sobre nuestro contexto tropical. *Polis, Revista de la Universidad Bolivariana*, 2(7).

Fernandes, F. M. B., Ribeiro, J. M., & Moreira, M. R. (2011). Reflexões sobre Avaliação de Políticas de Saúde no Brasil. *Cad. Saúde Pública*, 27(9), 11.

Fletcher, A., et al. (1993). Implications for trials in progress of publication of positive results. *Lancet* 342, 5.

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39-50.

Frasão, G. (2017). Em oito meses, número de cirurgias eletivas cresceu 39,1% no Brasil. *Portal Ministério da Saúde*. Retrieved from <http://portalms.saude.gov.br/noticias/agencia-saude/42101-em-oito-meses-numero-de-cirurgias-eletivas-cresceu-39-1-no-brasil>

Gattinara, B. C., Ibacache, J., Puente, C., Giaconi, J., & Caprara, A. (1995). Percepcion de la comunidad acerca de la calidad de los servicios de salud públicos en los distritos Norte e Ichilo, Bolívia. *Caderno de Saúde Publica*, 11, 14.

Guba, E. G., & Lincoln, Y. S. (2011). *Avaliação de quarta geração* (B. Honorato, Trans. E. Unicamp Ed.). São Paulo: Editora Unicamp.

Hart, S., & Sharma, S. (2004). Engaging fringe stakeholders for competitive imagination. *Academy of Management Executive*, 18(1), 12.

Hartz, S., & John, J. (2008). Contribution of economic evaluation to decision making in early phases of product development: A methodological and empirical review *International Journal of Technology Assessment in Health Care*, 24(4), 9. doi:10.1017/S0266462308080616

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In E. G. P. Limited (Ed.), *New Challenges to International Marketing (Advances in International Marketing (Vol. 20, pp. 277–319)*.

Machado, J. P., Martins, A. C. M., & Martins, M. S. (2013). Avaliação da qualidade do cuidado hospitalar no Brasil: uma revisão sistemática. *Cad. Saúde Pública*, 29(6), 20.

Moita, G. F. (2019). *Avaliação Integrativa de Performance Multidimensional e Decisão Multicritério: Um Proxy de Painel de Indicadores de Eficiência, Efetividade e Qualidade para Governança de Organizações Hospitalares e Serviços de Saúde no Brasil [Integrating Evaluation of Multidimensional Performance and Multicriterary Decision: A Panel Proxy of Efficiency, Effectiveness and Quality Indicators for Governance of Hospital Organizations and Health Services in Brazil]*. (Doutoramento Doctoral). Universidade de Coimbra, Coimbra.

Moita, G. F., Barbosa, A. C. Q., & Raposo, V. M. R. (2019). *The frontier of the evaluation of results of the Unified Health System: Reliability and validity of the evaluation items for the measurement of health performance in Brazil*. Paper presented at the X Congreso Internacional en Gobierno, Administración y Políticas Públicas #GIGAPP 2019, Madrid.

Moita, G. F., Raposo, V. M. d. R., & Barbosa, A. C. Q. (2019). *Confiabilidade e validade dos itens de avaliação no apoio à tomada de decisão e mensuração de performance na saúde – Uma análise do*

Sistema Único de Saúde. Paper presented at the 18o. Seminário de Diamantina, Diamantina - Minas Gerais - Brasil. <https://diamantina.cedeplar.ufmg.br/2019/>

Moita, G. F., Raposo, V. M. R., & Barbosa, A. C. Q. (2019a). Validação colaborativa de macrodimensões e indicadores-chave para avaliação de performance de serviços de saúde no Brasil. *Saúde em Debate*, 43(5), 232-247. Retrieved from revista.saudeemdebate.org.br/sed/issue/view/30

Moita, G. F., Raposo, V. M. R., & Barbosa, A. C. Q. (2019b). Validação colaborativa de macrodimensões e indicadores-chave para avaliação de performance de serviços de saúde no Brasil. *Revista Saúde em Debate*, 43(5), 232-247. Retrieved from revista.saudeemdebate.org.br/sed/issue/view/30

Murray, C. J. L., & Frenk, J. (2000). A framework for assessing the performance of health systems *Bulletin of the World Health Organization*, 78(6), 15.

Nunnally, J. C., & Bernstein, I. H. (1994). The Assessment of Reliability. *Psychometric Theory*, 3, 248-292.

Nunnally, J. C., & Berstein, I. H. (1994). *Psychometric theory* (McGraw-Hill Ed. 3a ed.). New York:OECD. (2017). *Health at a Glance 2017: OECD Indicators*. Retrieved from Paris: http://dx.doi.org/10.1787/health_glance-2017-en

Patton, M. Q. (1997). *Utilization-Focused Evaluation: The New Century Text* (S. Publications Ed.). London: Sage Publications.

Samico, I., Felisberto, E., Figueiró, A. C., & Frias, P. G. (2010). *Avaliação em saúde: bases conceituais e operacionais* (Medbook Ed.). Rio de Janeiro: Medbook.

Sampieri, R. H., Collado, C. F., & Lúcio, P. B. (2006). *Metodologia de pesquisa* (M. Hill Ed.). São Paulo: McGraw Hill.

Santos, M. P. (1995). Avaliação da qualidade dos serviços públicos de atenção à saúde da criança sob a ótica do usuário. *Revista Brasileira de Enfermagem*, 48, 11.

Silva, L. M. V., & Formigli, V. L. A. (1994). Avaliação em Saúde: Limites e Perspectivas. *Cad. Saúde Pública*, 10(1), 12.

Souza, L. E. P. F., Vieira-Da-Silva, L. M., & Hartz, Z. M. A. (2005). *Avaliação em Saúde: dos modelos teóricos à prática na avaliação de programas e sistemas de saúde*. Paper presented at the Conferência de consenso sobre a imagem-objetivo da descentralização da atenção à saúde no Brasil, Rio de Janeiro.

Tenenhous, M., Esposito Vinzi, V., Chatelin, Y., & Lauro, C. (2005). PLS path modeling. *Computational Statistics and Data Analysis*, 48, 159–205.

Uchimura, K. Y., & Bosi, M. L. M. (2002). Qualidade e subjetividade na avaliação de programas e serviços em saúde *Cad. Saúde Pública*, 18(6), 9.

Viacava, F., Almeida, C., Caetano, R., Fausto, M., Macinko, J., Martins, M., . . . Szwarcwald, C. L. (2004). Uma metodologia de avaliação do desempenho do sistema de saúde brasileiro. *Ciência & Saúde Coletiva*, 9(3), 14.

Viacava, F., Ugá, M. A. D., Porto, S., Laguardia, J., & Moreira, R. d. S. (2012). Avaliação de

Desempenho de Sistemas de Saúde: um modelo de análise *Ciência & Saúde Coletiva*, 17(4), 23.

Vieira-Da-Silva, L. M. (2005). Conceitos, abordagens e estratégias para a avaliação em saúde. In *Avaliação em saúde* (pp. 15-39). Salvador: EDUFBA.

Vuori, H. (1991a). A qualidade da saúde. *Divulg. saúde para debate*, 3(9), 8. Retrieved from <http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=LILACS&nextAction=Ink&exprSearch=223179&indexSearch=ID>

Vuori, H. (1991b). A qualidade da saúde. *Divulgação em saúde para debate*, 1, 9. WHO, W. H. O. (2006). *Bridging the "Know-Do" Gap: Meeting on Knowledge Translation in Global Health*. Retrieved from Geneva: