

Qualidade da residência médica na perspectiva das partes interessadas: revisão de escopo e painel Delphi

Quality of medical residency from the stakeholders' perspective: a scoping review and Delphi panel

Marcos Lopes de Miranda¹  marcoslmiranda@gmail.com
Sérgio da Cunha¹  sergio.d.cunha@gmail.com
Leila Barbosa²  leilafmr@yahoo.com.br
Michelle de Moura Balarini³  mibalarini@gmail.com
Suzane Almeida dos Santos Pinto⁴  suzaneasp@gmail.com
Denise Herdy Afonso¹  deniseherdy@gmail.com

RESUMO

Introdução: Embora na literatura médica existam diversas métricas para avaliar a qualidade dos programas de residência, os instrumentos avaliativos tendem a focar as dimensões dos residentes e preceptores. Isso negligencia uma ampla gama de partes interessadas (*stakeholders*), cujas perspectivas são fundamentais para uma compreensão holística da qualidade dos programas.

Objetivo: Diante desse contexto, esta revisão teve como objetivos mapear os principais *stakeholders* envolvidos na residência médica, identificar e categorizar as métricas de avaliação da qualidade dos programas mais prevalentes na literatura, e analisar a sua relevância em relação às perspectivas das partes interessadas.

Método: Inicialmente, foi realizada uma revisão de escopo da literatura para identificar e categorizar os *stakeholders*, além de mapear as métricas de qualidade. Posteriormente, um painel Delphi foi conduzido para analisar a relevância dessas métricas em relação às perspectivas das partes interessadas identificadas.

Resultado: Foram mapeados 14 *stakeholders* e identificadas 17 métricas, posteriormente divididas em quatro categorias principais. As métricas "adaptabilidade" e "bem-estar" se destacaram, sendo unanimemente reconhecidas por todos os *stakeholders* como "favorável" ou "altamente favorável". Por sua vez, "autoavaliação" e "satisfação do paciente" receberam avaliações mais cautelosas ou negativas. Os painelistas enfatizaram que "nenhuma métrica é capaz de fornecer individualmente uma avaliação precisa da qualidade de um programa de residência médica".

Conclusão: Ao mapear os *stakeholders* da residência médica, bem como identificar, categorizar e analisar as métricas de avaliação da qualidade mais prevalentes, este estudo ampliou o debate sobre a complexidade das perspectivas em torno da formação médica. A diversidade de atores envolvidos justifica valorizações distintas das várias dimensões da qualidade, reforçando a conclusão de que métricas isoladas não capturam integralmente a qualidade dos programas. Na prática, os resultados sublinham a importância da implementação de sistemas de avaliação da qualidade que sejam equilibrados e alinhados com as expectativas e necessidades dos principais *stakeholders*.

Palavras-chave: Estudos de Avaliação; Avaliação Educacional; Residência Médica; Indicadores; Participação dos Interessados.

ABSTRACT

Introduction: Medical literature abounds with metrics to assess the quality of residency programs, yet evaluative tools predominantly focus on the dimensions relevant to residents and preceptors. This overlooks a wide range of stakeholders, whose perspectives are essential for a holistic understanding of program quality. Given this context, this review aimed to map key stakeholders in medical residency, identify and categorize the most prevalent quality assessment metrics found in the literature, and analyze their relevance in relation to the stakeholders' perspectives.

Method: Initially, a scoping review of the literature was conducted to identify and categorize stakeholders, as well as to map quality metrics. This was followed by a Delphi panel to assess the relevance of these metrics in relation to the perspectives of the identified stakeholders.

Results: Fourteen stakeholders were mapped, and seventeen metrics were identified, subsequently divided into four main categories. The metrics "Adaptability" and "Well-being" stood out, being unanimously recognized by all stakeholders as "Favorable" or "Highly Favorable". Conversely, "Self-assessment" and "Patient Satisfaction" elicited more cautious or negative evaluations. Panelists emphasized that "no single metric can provide an accurate assessment of the quality of a medical residency program on its own".

Conclusion: By mapping the stakeholders of medical residency, as well as identifying, categorizing, and analyzing the most prevalent quality assessment metrics, this study broadened the debate on the complexity of perspectives surrounding medical education. The diversity of actors observed justifies distinct evaluations across the various dimensions of quality, reinforcing the conclusion that isolated metrics cannot fully capture the program quality. In practice, the findings underscore the importance of implementing quality assessment frameworks that are balanced and aligned with the expectations and needs of the main stakeholders.

Keywords: Evaluation Study; Educational Measurement; Medical Residency; Indicators; Stakeholders Participation.

¹ Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brasil.

² Instituto Brasileiro de Mercado de Capitais, Rio de Janeiro, Rio de Janeiro, Brasil.

³ Hospital Universitário Pedro Ernesto, Rio de Janeiro, Rio de Janeiro, Brasil.

⁴ Hospital Federal da Lagoa, Rio de Janeiro, Rio de Janeiro, Brasil.

Editora-chefe: Rosiane Viana Zuza Diniz.

Editor associado: Kristopheron Lustosa Augusto.

Recebido em 02/12/23; Aceito em 14/05/24.

Avaliado pelo processo de double blind review.

INTRODUÇÃO

Entre 2018 e 2021, o número de médicos cursando residência no Brasil cresceu de 38.681 para 41.853, representando cerca de 8% do total de médicos ativos no país¹. Esse aumento reflete as políticas públicas voltadas para a expansão dos programas de residência médica, estando em consonância com as demandas do Sistema Único de Saúde e com o Programa Pró-Residência do governo federal. Contudo, esse crescimento impõe o desafio de garantir a qualidade desses programas.

Embora na literatura médica existam diversas métricas para avaliar a qualidade dos programas de residência, a maioria dos instrumentos avaliativos foca as dimensões dos residentes e preceptores. Isso negligencia as perspectivas de outras partes interessadas (*stakeholders*), obstaculizando uma compreensão holística da qualidade.

Do ponto de vista gerencial, a residência médica é mais do que um programa de formação; é um ecossistema complexo com múltiplos atores, cada um com suas expectativas e necessidades. Nesse cenário, a premência de uma avaliação de qualidade abrangente que considere todas as partes envolvidas é uma demanda estratégica para assegurar que os programas não sejam apreciados apenas por suas entregas diretas, como a formação de especialistas. A verdadeira qualidade de um programa de residência médica residiria na sua capacidade de atender às expectativas de todos os interessados e superá-las.

Diante desse contexto, esta revisão teve como objetivos mapear os principais *stakeholders* envolvidos na residência médica, identificar e categorizar as métricas de avaliação da qualidade dos programas mais prevalentes na literatura, e

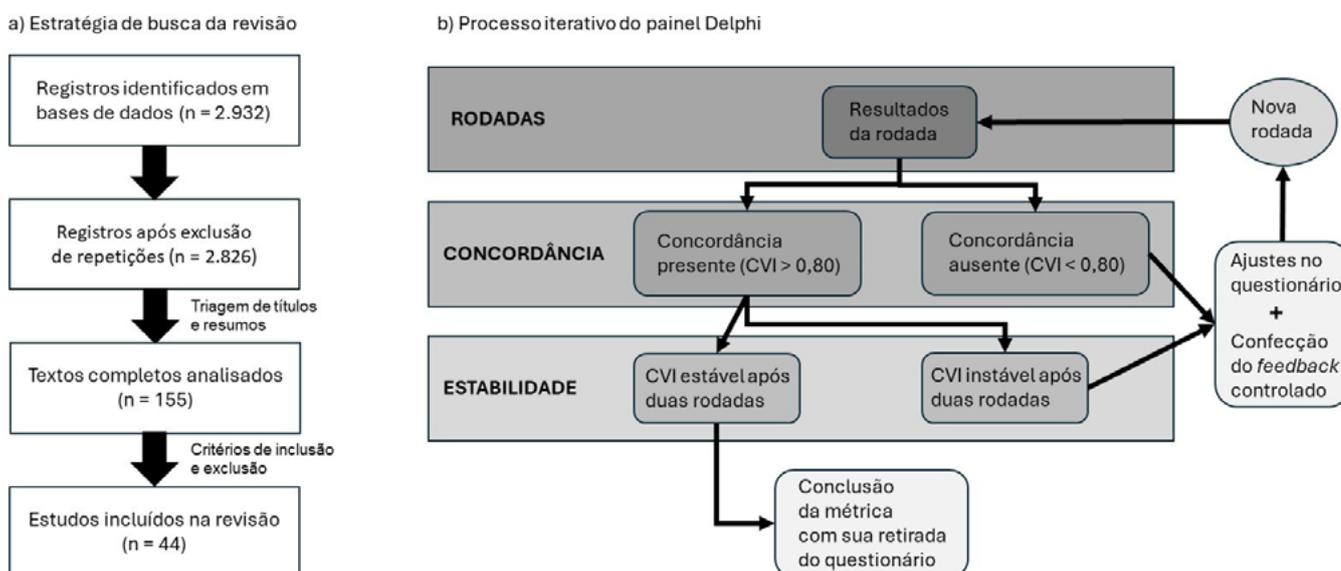
analisar a sua relevância em relação às perspectivas das partes interessadas.

MÉTODO

Seguindo as Recomendações PRISMA-ScR² e o método proposto pelo Joanna Briggs Institute³ para a condução de uma revisão de escopo, as bases de dados Medline (PubMed), Google Scholar, Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS) e Scientific Electronic Library Online (SciELO) foram pesquisadas empregando termos como *residency, postgraduate medical education, program evaluation, quality metrics, assessment, standards, achievement, benchmarking, stakeholder, partnership* e *customer*. Citações relevantes foram identificadas até a busca final em 1º de agosto de 2023. Para as estratégias de busca projetadas para o PubMed e LILACS, foram utilizados termos de vocabulário controlado para cada conceito e combinações booleanas com sinônimos de palavras-chave. Nas bases Google Scholar e SciELO usaram-se apenas combinações booleanas de termos-chave em inglês e seus equivalentes em português e espanhol, e os resultados foram limitados a revistas acadêmicas. Também se procedeu à busca manual nas referências de todos os artigos incluídos. A Figura 1 ilustra a estratégia de busca utilizada.

Títulos e resumos de artigos foram exportados para o programa Rayyan⁴ para exclusão de duplicatas e organização, e para o ASReview⁵ para triagem final, utilizando um processo semiautomatizado. Dois membros da equipe de pesquisa (MLM e MMB) revisaram de forma independente todos os resumos de artigos para determinar sua relevância. Discordâncias foram resolvidas por consenso ou com a intervenção de um terceiro

Figura 1. Estratégia de busca utilizada para condução da revisão de escopo e ilustração esquemática do processo Delphi.



CVI: *content validity index*.
 Fonte: Elaborada pelos autores.

revisor (SASP). Em seguida, dois membros da equipe (MLM e MMB) revisaram, também de forma independente, as versões completas dos artigos relevantes. Novamente, discordâncias foram resolvidas por consenso ou com a intervenção de um membro da equipe (SASP). Os critérios de inclusão abrangeram artigos em inglês, português e espanhol, sem limite de data de publicação, que abordavam a avaliação de qualidade e/ou os *stakeholders* dos programas de residência médica. Embora editoriais, cartas ao editor, comentários e opiniões tenham sido inicialmente excluídos, os critérios de exclusão foram ajustados para incorporar esses materiais. Essa mudança visou tornar a pesquisa ainda mais ampla, para abordar e preencher lacunas de conhecimento relacionadas aos temas investigados. Dessa forma, apenas artigos sem resumos disponíveis foram excluídos. Dois membros da equipe de pesquisa (MLM e MMB) extraíram, de forma independente, dados relevantes do conjunto final de artigos incluídos no estudo.

Após a revisão de escopo, percebeu-se a ausência de estudos que tratassem, direta ou indiretamente, da interação entre métricas e as perspectivas dos *stakeholders* no âmbito da avaliação da qualidade dos programas de residência. Diante dessa falta de dados, optou-se por implementar um painel utilizando a metodologia Delphi^{6,7}.

Seleção dos painelistas

A seleção dos painelistas foi realizada por meio de um método de amostragem intencional não probabilística. Os especialistas foram escolhidos por sua vinculação direta com os grupos de *stakeholders* identificados na revisão de escopo, destacando-se por conhecimento notório, experiência significativa, liderança reconhecida e/ou influência substancial em suas áreas, com preferência para aqueles com experiência na formação de residentes (médicos ou não). Visando a uma análise abrangente e equitativa da qualidade dos programas de residência médica, buscou-se uma composição equilibrada do painel, incluindo profissionais da academia e da assistência, representantes dos setores público e privado, além de garantir diversidade geográfica, de gênero e étnico-racial. Convites foram enviados por *e-mail* a 45 especialistas, dos quais 30 aceitaram participar. Cada grupo de interesse foi representado por, no mínimo, dois especialistas.

O processo Delphi

O painel Delphi foi estruturado em várias rodadas, durante as quais especialistas avaliaram métricas de qualidade dos programas, considerando a perspectiva de todos os grupos de *stakeholders*. Utilizou-se uma escala Likert de quatro pontos (“contrário”, “neutro”, “favorável”, “altamente favorável”) para avaliar a relevância de cada métrica, expandida

após o primeiro *round* para incluir “favorável com ressalvas”, baseado em sugestões dos painelistas. Após cada rodada, as respostas eram analisadas e os dados apresentados aos especialistas em formato facilmente interpretável, incluindo gráficos e estatísticas que mostravam medidas de tendência central, dispersão e porcentagem de distribuição. Comentários anônimos dos painelistas também foram incorporados como parte da devolutiva. Esse *feedback* controlado permitia aos membros do painel refletir sobre as tendências, possibilitando ajustes nas suas respostas. A concordância entre os participantes foi avaliada usando uma modificação do *content validity index* (CVI)⁸, expresso em porcentagem e calculado pela proporção de votos que cada métrica recebeu em um determinado ponto da escala Likert em relação ao número total de votantes. Quando o maior CVI de uma métrica era inferior a 0,80 (80%), a concordância era considerada insuficiente. Nesses casos, a descrição da métrica era revisada de acordo com as sugestões dos painelistas para melhorar a clareza e facilitar o consenso. Os itens revisados, juntamente com os comentários dos participantes, eram então reapresentados na próxima rodada Delphi. Métricas com CVI igual ou superior a 0,80 eram consideradas de concordância aceitável. O encerramento de cada métrica dependia da estabilidade do CVI acima de 0,80, com uma variação inferior a 5% após duas rodadas consecutivas. Métricas que atendessem a esse critério eram retiradas dos questionários subsequentes para focar as que ainda não haviam alcançado consenso. Esse processo iterativo continuou até que consenso sobre a pertinência e aplicabilidade de todas as métricas fosse alcançado. A Figura 1 ilustra esse processo esquematicamente.

RESULTADOS

Na busca inicial nas bases de dados, foram identificadas 2.826 entradas únicas. A triagem de títulos e resumos descartou 2.671 artigos, restando 155 para revisão completa do texto. Destes, 44 artigos atenderam aos critérios de inclusão. A busca manual nas referências não identificou artigos adicionais para revisão.

Mapeamento dos principais *stakeholders* dos programas de residência médica

Diversos artigos analisados durante a triagem focaram exclusivamente três grupos principais de *stakeholders*: residentes, preceptores e pacientes. Essa abordagem limitada não capturou a amplitude e a complexidade do tema. Dessa forma, esses artigos não foram selecionados para revisão completa do texto.

A maioria dos artigos triados para revisão completa não detalhou ou identificou claramente os “clientes” dos programas

de residência médica. Somente dois artigos^{9,10} explicitaram e organizaram os principais *stakeholders* dos programas:

- 1) Residentes: Buscam um processo de seleção justo, treinamento alinhado com demandas profissionais futuras e uma avaliação criteriosa de suas competências. Também esperam, ao concluírem a formação, oportunidades de prática em ambientes que valorizem sua especialização e ofereçam remuneração justa.
- 2) Consumidores diretos (como pacientes e familiares): Procuram cuidados de saúde competentes, compassivos e acessíveis. Demandam um número adequado de profissionais de saúde, especialmente em regiões desassistidas.
- 3) Órgãos de financiamento: Desejam que os programas de residência satisfaçam as necessidades e expectativas dos demais *stakeholders* de maneira custo-efetiva e alinhada com as finalidades do financiamento.
- 4) Organizações da saúde suplementar: Buscam médicos com uma compreensão clara das perspectivas dos pagadores e que pratiquem cuidados centrados no paciente, baseados em práticas assistenciais e evidências atualizadas. Além disso, valorizam a capacidade de trabalho multidisciplinar e o envolvimento em atividades gerenciais.
- 5) Centros de treinamento (como escolas médicas, hospitais de ensino e instituições da rede de saúde escola): Almejam que os programas de residência contribuam para sua missão de formar profissionais qualificados, líderes, pesquisadores e educadores. Apoiam a inovação e melhoria contínua nos processos educacionais para atingir a excelência.
- 6) Organizações profissionais (como conselhos de medicina, associações de especialidade e de educação médica): Apoiam a formação de profissionais éticos e competentes, capazes de representar sua classe com integridade, promover sua especialidade de forma positiva e impulsionar avanços significativos em suas áreas de atuação.
- 7) Órgãos reguladores e/ou de acreditação (como a Comissão Nacional de Residência Médica – CNRM): Exigem que os programas de residência cumpram padrões de conformidade rigorosos e proporcionem treinamento de alta qualidade, em linha com as melhores práticas e diretrizes. Buscam assegurar essa excelência por meio de avaliações periódicas dos programas.

Embora não explicitados diretamente, certos *stakeholders* foram identificados indiretamente (inferidos) durante a revisão de escopo:

- 1) Governo e entidades governamentais: Defendem o alinhamento dos programas de residência com as políticas de saúde pública. Além disso, incentivam uma distribuição equitativa das oportunidades de treinamento em todo o território, principalmente em regiões desassistidas.
- 2) Outros profissionais de saúde (equipe multidisciplinar): Valorizam a colaboração efetiva com os programas de residência, esperando que a formação médica de qualidade promova uma assistência à saúde integrada e centrada no paciente.
- 3) Pesquisadores e acadêmicos: Esperam que os programas de residência incorporem as últimas descobertas e inovações, preparando médicos para contribuir com pesquisas de ponta. Também almejam que a formação proporcione uma base sólida em metodologia científica, capacitando os residentes para que possam ser colaboradores em projetos acadêmicos.
- 4) Equipe de apoio ao ensino (incluindo coordenadores de programa e membros administrativos): Aspiram a programas que sejam bem organizados, eficientes e alinhados aos objetivos educacionais, o que facilita suas funções operacionais diárias.
- 5) Indústria farmacêutica e de dispositivos médicos: Apesar de seu papel ser controverso como parte interessada, essas entidades frequentemente buscam parcerias com programas de residência, oferecendo recursos educacionais, patrocínios e oportunidades de pesquisa.
- 6) *Alumni* (ex-residentes): Valorizam programas que mantêm ou melhoram seus padrões de qualidade ao longo do tempo, refletindo positivamente em suas carreiras. Estão abertos a oportunidades de contribuir para o desenvolvimento contínuo da residência, por meio de *feedback*, mentoria ou outras formas de apoio.
- 7) Estudantes de Medicina: Ambicionam ingressar em programas de residência de alta qualidade e, portanto, demonstram interesse por decisões e processos que influenciam a formação dos residentes.

Identificação e categorização das métricas de avaliação da qualidade dos programas de residência médica¹¹⁻⁵²

A revisão de escopo possibilitou a organização das métricas de avaliação da qualidade em quatro categorias principais:

“avaliações formativas e somativas”, “questões acadêmicas”, “aspectos clínicos e sociais da formação” e “bem-estar”.

As métricas mais frequentemente identificadas referem-se às “avaliações formativas e somativas”:

- 1) Testes escritos ou orais para avaliação de progresso ou provas de título: São ferramentas de avaliação tradicionais, frequentemente utilizadas por sociedades de especialistas. Visam avaliar o conhecimento teórico e prático dos residentes, podendo auxiliar no *benchmarking* da qualidade dos programas: comparações dos resultados dos residentes com padrões nacionais ou internacionais permitem identificar pontos fortes e oportunidades de melhoria.
- 2) Marcos de competência e atividades profissionais confiabilizadoras: Fornecem uma estrutura para avaliar o progresso dos residentes ao longo de sua formação. São geralmente elaborados por sociedades de especialistas, sendo específicos para cada especialidade ou subespecialidade. A padronização das avaliações entre os programas permite a realização de *benchmarking*.
- 3) Avaliação por docentes/preceptores: Baseada na observação direta e interação com os residentes, essa métrica permite abordar diversos aspectos, como habilidades clínicas, comunicação e profissionalismo. Pode incluir o uso de simulação, portfólio, OSCE etc.
- 4) Avaliação por pares: Oferece *insights* sobre a capacidade de trabalho em equipe, comunicação efetiva e outras habilidades interpessoais do residente. Adiciona uma nova perspectiva sobre o desempenho dos avaliados, podendo ser útil para identificar áreas de melhoria menos evidentes para os docentes/preceptores.
- 5) Autoavaliação: Permite que os residentes avaliem o próprio desempenho, promovendo reflexão sobre suas experiências e autoconhecimento, com identificação de pontos fortes e reconhecimento de áreas que carecem de desenvolvimento.

Também foram frequentemente identificadas métricas de avaliação da qualidade voltadas para “questões acadêmicas”:

- 1) Aprovação de ex-residentes em programas de especialização: Refere-se à aprovação de *alumni* em programas com pré-requisito, destinados a aprofundar competências em outra especialidade ou área de atuação. Uma elevada taxa de aprovação refletiria a solidez da formação inicial e a preparação adequada dos residentes para avanços acadêmicos e profissionais.

- 2) Papéis de liderança acadêmica: Refere-se às posições de liderança ocupadas por residentes durante ou após sua formação, como participação em comitês acadêmicos, chefia de equipes assistenciais ou coordenação de projetos. A presença significativa de residentes e ex-residentes em funções de destaque indicaria uma alta qualidade do programa.
- 3) Envolvimento dos residentes em atividades de pesquisa e reconhecimento acadêmico: Relaciona-se ao engajamento dos residentes em iniciativas científicas, como publicações, apresentações em conferências e desenvolvimento de projetos de pesquisa. Também envolve a obtenção de bolsas de estudo, prêmios e outros reconhecimentos no âmbito acadêmico. Ambos seriam indicativos da alta qualidade do programa.
- 4) Reconhecimento e prestígio do programa: Métrica frequentemente determinada pela demanda do programa entre os candidatos (relação candidato/vaga). Uma relação elevada indicaria o reconhecimento e a valorização do programa nos meios acadêmico e profissional, funcionando como um indicador da sua alta qualidade.

A saúde física e mental dos residentes (“bem-estar”) emergiu na revisão como uma métrica fundamental para aferir a excelência dos programas de residência:

- Bem-estar: O esgotamento profissional pode comprometer a qualidade do cuidado destinado ao paciente. Reconhecendo isso, muitos programas de residência implementaram estratégias para monitorar o bem-estar dos residentes ao longo da sua formação. Iniciativas incluem currículos voltados para o desenvolvimento de habilidades para a manutenção da saúde física e do equilíbrio emocional, não só durante a residência, mas também no transcurso de toda a vida profissional. A avaliação contínua do bem-estar dos residentes em nível programático é fundamental para utilização dessa métrica como um indicador-chave da qualidade dos programas.

Concluindo a identificação e categorização das métricas de avaliação da qualidade, algumas oferecem *insights* sobre os “aspectos clínicos e sociais da formação”, avaliando competências técnicas e o engajamento social dos residentes:

- 1) Métricas de desempenho clínico: Definidas por cada programa, essas métricas avaliam a eficiência e qualidade do atendimento utilizando indicadores como número de pacientes atendidos por hora ou taxa de retorno não planejado ao departamento

de emergência. O maior desafio no seu uso reside no desenvolvimento de métricas que reflitam adequadamente o desempenho e as habilidades do residente, sem serem influenciadas por fatores externos, como limitações de recursos.

- 2) **Competência procedural:** Foca a capacidade de o residente realizar procedimentos com sucesso e sem complicações, essencial para independência na prática médica. Procedimentos documentados em portfólios ou diários pelos próprios residentes são a base para essa avaliação. Contudo, a falta de padronização na mensuração do desempenho procedural entre diferentes programas representa um desafio para sua aplicação no *benchmarking*.
- 3) **Adaptabilidade:** Em um cenário médico que evolui cada vez mais rápido, a capacidade de adaptação dos programas a novas práticas, como consultas virtuais, destaca-se como um marcador de excelência. Adicionalmente, a qualidade de um programa poderia ser medida pela sua habilidade de promover um ambiente propício à aprendizagem adaptativa e pela avaliação contínua dessa competência. Programas de excelência fomentam o desenvolvimento de habilidades metacognitivas nos residentes, capacitando-os a autoavaliar-se, adaptar suas abordagens e assimilar novos conhecimentos eficientemente.
- 4) **Satisfação do paciente:** Apesar de sua natureza controversa devido à alta subjetividade, a satisfação do paciente é amplamente adotada como indicador da qualidade do atendimento, do desempenho dos médicos residentes e, conseqüentemente, da excelência do programa.
- 5) **Empatia:** A capacidade de compreender e conectar-se aos sentimentos dos pacientes é fundamental na medicina. Dessa forma, a preservação e o cultivo da empatia ao longo da residência, apesar dos desafios emocionais inerentes ao treinamento médico, são descritos como essenciais para avaliação da excelência dos programas.
- 6) **Serviço comunitário:** O envolvimento dos residentes em serviço comunitário, especialmente em determinadas especialidades, é visto como uma métrica valiosa, refletindo o compromisso do programa com o bem-estar comunitário. Ao priorizarem essa medida, os programas fomentam nos residentes uma conexão mais estreita com a comunidade, encorajando uma prática médica mais integral e centrada no paciente. O sucesso

da avaliação depende da definição de objetivos claros e critérios específicos para participação dos residentes em atividades comunitárias, possibilitando uma mensuração objetiva do impacto e alcance desse engajamento.

- 7) **Ativismo em saúde:** Avalia o engajamento dos residentes em causas sociais e políticas que impactam a saúde dos indivíduos e das comunidades, abrangendo ações voltadas à defesa e promoção dos direitos dos pacientes, à melhoria do sistema de saúde, à promoção da saúde pública etc. Embora ainda pouco explorados, indicadores como número de residentes envolvidos em ativismo em saúde, quantidade de projetos bem-sucedidos implementados por eles ou impacto dessas iniciativas na comunidade já foram considerados métricas potenciais da qualidade dos programas de residência.

Análise da relevância das métricas de qualidade em relação às perspectivas dos stakeholders

Entre setembro e novembro de 2023, realizaram-se quatro rodadas do painel Delphi até que um consenso fosse alcançado sobre todas as métricas propostas. Nesse período, a taxa média de resposta/devolução dos questionários foi de 87%. Os resultados das discussões entre os painelistas são apresentados nos quadros 1, 2 e 3. O Quadro 4 exibe trechos selecionados dos comentários dos painelistas, escolhidos para ilustrar as diversas perspectivas e discussões emergentes do processo Delphi.

As métricas “marcos de competência e atividades profissionais confiabilizadoras” e “avaliação por docentes/preceptores” receberam somente avaliações favoráveis, sendo reconhecidas como “altamente favorável” pela maioria dos *stakeholders* (Quadro 1). Por sua vez, “testes escritos ou orais”, “avaliação por pares” e “autoavaliação” mostraram-se métricas bastante polarizadas, com opiniões variando de “altamente favorável” a “contrário” (Quadro 1).

Os painelistas manifestaram oposição à adoção de todas as métricas voltadas para “questões acadêmicas” quando avaliadas sob a perspectiva dos consumidores diretos (Quadro 2). A métrica “envolvimento dos residentes em atividades de pesquisa e reconhecimento acadêmico” foi valorizada pela maioria dos *stakeholders*. As demais métricas mostraram-se bastante polarizadas, com opiniões variando de “altamente favorável” a “contrário” (Quadro 2).

Todos os *stakeholders* se mostraram favoráveis ou altamente favoráveis em relação ao uso da métrica “bem-estar” para avaliação da qualidade dos programas de residência (Quadro 2).

Quadro 1. Matriz de relação entre métricas de avaliação da qualidade (“avaliações formativas e somativas”) e perspectivas dos *stakeholders*.

Stakeholders	Testes escritos ou orais	Marcos de competência e atividades profissionais confiabilizadoras	Avaliação por docentes/preceptores	Avaliação por pares	Autoavaliação
Residentes	FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL
Consumidores diretos	CONTRÁRIO	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL, COM RESSALVAS	CONTRÁRIO
Órgãos de financiamento	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL, COM RESSALVAS	CONTRÁRIO
Organizações da saúde suplementar	NEUTRO	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	CONTRÁRIO
Centros de treinamento	FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL
Organizações profissionais	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	CONTRÁRIO
Órgãos reguladores e/ou de acreditação	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL, COM RESSALVAS
Governo e entidades governamentais	CONTRÁRIO	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	CONTRÁRIO
Outros profissionais de saúde	CONTRÁRIO	ALTAMENTE FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL	NEUTRO
Pesquisadores e acadêmicos	CONTRÁRIO	FAVORÁVEL, COM RESSALVAS	ALTAMENTE FAVORÁVEL	CONTRÁRIO	FAVORÁVEL, COM RESSALVAS
Equipe de apoio ao ensino	FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL, COM RESSALVAS	ALTAMENTE FAVORÁVEL
Indústria farmacêutica e de dispositivos médicos	CONTRÁRIO	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL, COM RESSALVAS	CONTRÁRIO	CONTRÁRIO
<i>Alumni</i>	FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	NEUTRO
Estudantes de Medicina	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL

Fonte: Elaborado pelos autores.

Quadro 2. Matriz de relação entre métricas de avaliação da qualidade (“questões acadêmicas” e “bem-estar”) e perspectivas dos *stakeholders*.

Stakeholders	Aprovação de ex-residentes em programas de especialização	Papéis de liderança acadêmica	Envolvimento dos residentes em atividades de pesquisa e reconhecimento acadêmico	Reconhecimento e prestígio do programa	Bem-estar
Residentes	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL
Consumidores diretos	CONTRÁRIO	CONTRÁRIO	CONTRÁRIO	CONTRÁRIO	FAVORÁVEL
Órgãos de financiamento	ALTAMENTE FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL
Organizações da saúde suplementar	FAVORÁVEL, COM RESSALVAS	CONTRÁRIO	FAVORÁVEL, COM RESSALVAS	CONTRÁRIO	FAVORÁVEL
Centros de treinamento	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL
Organizações profissionais	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL

Continua...

Quadro 2. Continuação.

Stakeholders	Aprovação de ex-residentes em programas de especialização	Papéis de liderança acadêmica	Envolvimento dos residentes em atividades de pesquisa e reconhecimento acadêmico	Reconhecimento e prestígio do programa	Bem-estar
Órgãos reguladores e/ou de acreditação	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL
Governo e entidades governamentais	NEUTRO	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL
Outros profissionais de saúde	NEUTRO	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL
Pesquisadores e acadêmicos	NEUTRO	NEUTRO	ALTAMENTE FAVORÁVEL	NEUTRO	ALTAMENTE FAVORÁVEL
Equipe de apoio ao ensino	FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL
Indústria farmacêutica e de dispositivos médicos	NEUTRO	CONTRÁRIO	FAVORÁVEL	CONTRÁRIO	FAVORÁVEL
Alumni	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	FAVORÁVEL
Estudantes de Medicina	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL

Fonte: Elaborado pelos autores.

Quadro 3. Matriz de relação entre métricas de avaliação da qualidade (“aspectos clínicos e sociais da formação”) e perspectivas dos stakeholders

Stakeholders	Métricas de desempenho clínico	Competência procedural	Adaptabilidade	Satisfação do paciente	Empatia	Serviço comunitário	Ativismo em saúde
Residentes	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL	ALTAMENTE FAVORÁVEL	NEUTRO	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL
Consumidores diretos	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL
Órgãos de financiamento	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL	NEUTRO
Organizações da saúde suplementar	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	NEUTRO	CONTRÁRIO
Centros de treinamento	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	NEUTRO	FAVORÁVEL	FAVORÁVEL	CONTRÁRIO
Organizações profissionais	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	NEUTRO	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL	FAVORÁVEL, COM RESSALVAS
Órgãos reguladores e/ou de acreditação	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	FAVORÁVEL
Governo e entidades governamentais	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL	NEUTRO
Outros profissionais de saúde	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL	FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL, COM RESSALVAS
Pesquisadores e acadêmicos	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL	ALTAMENTE FAVORÁVEL	CONTRÁRIO	CONTRÁRIO	NEUTRO	CONTRÁRIO
Equipe de apoio ao ensino	FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	NEUTRO	FAVORÁVEL	FAVORÁVEL	NEUTRO

Continua...

Quadro 3. Continuação

Stakeholders	Métricas de desempenho clínico	Competência procedural	Adaptabilidade	Satisfação do paciente	Empatia	Serviço comunitário	Ativismo em saúde
Indústria farmacêutica e de dispositivos médicos	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL, COM RESSALVAS	ALTAMENTE FAVORÁVEL	CONTRÁRIO	CONTRÁRIO	CONTRÁRIO	CONTRÁRIO
<i>Alumni</i>	FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	NEUTRO	FAVORÁVEL, COM RESSALVAS	ALTAMENTE FAVORÁVEL	FAVORÁVEL, COM RESSALVAS
Estudantes de Medicina	FAVORÁVEL	ALTAMENTE FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL, COM RESSALVAS	FAVORÁVEL	ALTAMENTE FAVORÁVEL	FAVORÁVEL

Fonte: Elaborado pelos autores.

Quadro 4. Trechos selecionados dos comentários e das discussões dos painelistas.

Métrica	Comentários dos painelistas
Testes escritos ou orais	<p>“Esses testes são arcaicos e falham em avaliar habilidades críticas como tomada de decisão sob pressão e empatia. Eles perpetuam um modelo obsoleto de educação médica focado em memorização, em vez de habilidades práticas e julgamento clínico.”</p> <ul style="list-style-type: none"> • Esse comentário gerou um debate sobre a necessidade de reformular ou atualizar esses testes para que reflitam mais adequadamente as exigências da medicina moderna, em contraposição à defesa de sua manutenção como métodos eficazes de avaliação dos conhecimentos teóricos fundamentais.
Avaliação por pares	<p>“Embora possa fornecer uma perspectiva única sobre o desempenho interpessoal, essa avaliação frequentemente se transforma em uma competição de popularidade ou simpatia, comprometendo sua objetividade. Além disso, pode criar um ambiente de trabalho hostil, onde os residentes se sentem pressionados a competir em vez de colaborar.”</p> <ul style="list-style-type: none"> • Esse argumento provocou divergências, destacando-se a apreciação da avaliação por pares como um reflexo autêntico do ambiente de trabalho contra as preocupações sobre seu potencial para criar tensões e impactar negativamente a dinâmica de equipe.
Autoavaliação	<p>“A autoavaliação frequentemente falha em proporcionar uma visão realista das habilidades e competências do residente. Pode levar a uma falsa sensação de competência ou, inversamente, a uma autocritica excessiva, sem fornecer um caminho claro para o aprimoramento.”</p> <ul style="list-style-type: none"> • Esse comentário provocou opiniões divididas entre o reconhecimento da autoavaliação como ferramenta essencial para o autoconhecimento e crescimento pessoal e profissional e questionamentos sobre sua eficácia e precisão, sugerindo a necessidade de complementá-la com métodos avaliativos mais objetivos.
Aprovação de ex-residentes em programas de especialização	<p>“A aprovação em programas de especialização não reflete necessariamente a qualidade da formação inicial. Pode ser mais um indicativo de esforços individuais pós-residência do que da qualidade do programa em si.”</p> <ul style="list-style-type: none"> • Houve grande controvérsia entre a valorização da métrica como um indicativo de sucesso contínuo e as dúvidas sobre sua correlação direta com a qualidade intrínseca do programa de residência.
Papéis de liderança acadêmica	<p>“Focar papéis de liderança acadêmica pode inadvertidamente promover uma cultura de elitismo e não necessariamente traduzir a competência clínica ou a qualidade do ensino.”</p> <ul style="list-style-type: none"> • Esse comentário provocou uma discussão sobre a importância de valorizar as conquistas/ realizações acadêmicas sem que elas ofusquem outros aspectos importantes da formação médica.
Reconhecimento e prestígio do programa	<p>“O prestígio de um programa pode ser baseado em percepções históricas ou em <i>marketing</i>, em vez de refletir a qualidade atual do ensino e do treinamento.”</p> <ul style="list-style-type: none"> • Esse <i>insight</i> desencadeou um debate sobre as formas de mensuração do reconhecimento e prestígio de um programa e sua confiabilidade como indicadores da qualidade oferecida.

Continua...

Quadro 4. Continuação.

Métrica	Comentários dos painelistas
Satisfação do paciente	<p>“Esta métrica é uma armadilha! Ela pode levar os residentes a priorizarem a satisfação superficial do paciente em detrimento de decisões médicas mais rigorosas e necessárias. Estamos formando médicos ou profissionais de relações públicas?”</p> <ul style="list-style-type: none"> Houve um debate em torno da valorização da satisfação do paciente como reflexo da qualidade do atendimento e cuidado <i>versus</i> a preocupação de que uma ênfase excessiva nessa métrica pudesse “reduzir a medicina a um espetáculo”, no qual agradar ao paciente se sobrepõe à prática médica baseada em evidências.
Empatia	<p>“Embora seja um aspecto crucial da prática médica, sua avaliação é complexa e desafiadora. Como medimos genuinamente a empatia sem cair em respostas ensaiadas ou comportamento performático?”</p> <ul style="list-style-type: none"> Esse ponto de vista suscitou discussões sobre a incorporação eficaz da empatia na avaliação dos residentes, ponderando entre a necessidade de autenticidade e os desafios de sua mensuração.
Serviço comunitário	<p>“Estamos correndo o risco de transformar futuros médicos em ativistas sociais em vez de clínicos competentes. O serviço comunitário é valioso, mas não deve ser o foco central da residência médica.”</p> <ul style="list-style-type: none"> Esse comentário evidenciou a divisão de opiniões entre a valorização do envolvimento comunitário como fundamental para a formação de médicos socialmente conscientes e as preocupações de que tal ênfase possa desviar tempo e recursos do desenvolvimento de habilidades clínicas essenciais.
Ativismo em saúde	<p>“Incorporar o ativismo em saúde na formação médica é uma distração perigosa. Estamos aqui para formar médicos, não ativistas políticos. O excesso de ênfase no ativismo pode corroer o foco na ciência médica e na prática baseada em evidências.”</p> <ul style="list-style-type: none"> Esse comentário destacou a tensão entre a defesa do ativismo em saúde como ferramenta para abordar questões sociais importantes e a preocupação de que enfatizar demais essa métrica possa desviar o foco da educação médica <i>per se</i> e comprometer a objetividade científica.

Fonte: Elaborado pelos autores.

A métrica “adaptabilidade” foi altamente valorizada por todos os *stakeholders* (Quadro 3). As métricas “métricas de desempenho clínico” e “competência procedural” receberam apenas avaliações favoráveis, apesar de algumas ressalvas por determinados grupos (Quadro 3). As métricas “empatia”, “serviço comunitário” e “ativismo em saúde” revelaram-se bastante polarizadas, com avaliações que variaram de “altamente favorável” a “contrário”. A métrica “satisfação do paciente” não alcançou avaliações “altamente favorável”, apresentando várias indicações de neutralidade e rejeição (Quadro 3).

Na análise de todas as métricas, “adaptabilidade” e “bem-estar” se destacaram, sendo unanimemente reconhecidas por todos os *stakeholders* como “favorável” ou “altamente favorável”. Por sua vez, “autoavaliação” e “satisfação do paciente” receberam avaliações mais cautelosas ou negativas. Em seu relatório final, os painelistas enfatizaram que “nenhuma métrica é capaz de fornecer individualmente uma avaliação precisa da qualidade de um programa de residência médica”.

DISCUSSÃO

No intrincado ecossistema da residência médica, em que diversos *stakeholders* interagem com perspectivas distintas, torna-se imperativo que a avaliação de qualidade dos programas transcenda as métricas tradicionais, que muitas vezes se limitam

a focar atores mais óbvios, como residentes e preceptores. Elliott et al.⁹ iniciaram essa discussão ao proporem uma definição multidimensional mais abrangente da qualidade do treinamento em residência, enfatizando a importância de identificar as principais partes interessadas e entender suas necessidades específicas. Essa abordagem foi aprofundada por Simpson et al.¹⁰, que enriqueceram a compreensão sobre as perspectivas dos “clientes” dos programas de residência. A presente revisão avançou no mapeamento dos *stakeholders*, revelando uma gama mais ampla de atores envolvidos, direta ou indiretamente, nos programas. Paralelamente, ao identificar e categorizar as métricas de qualidade dos programas de residência médica, esta revisão de escopo não só consolidou como também expandiu os achados de Jewell et al.²⁷, que, por meio de uma revisão narrativa, buscaram esboçar um panorama das métricas de avaliação de residentes e sua aplicabilidade na avaliação dos programas. Contudo, o maior pioneirismo deste estudo residiu na sua abordagem original de analisar a relevância de cada métrica identificada em relação às perspectivas dos *stakeholders*. Explorando a questão central “Qualidade para quem?”, este trabalho destaca a importância de considerar diversas perspectivas e múltiplas dimensões da qualidade na avaliação dos programas de residência médica.

As matrizes resultantes do consenso entre os painelistas ilustram a complexidade dessa avaliação.

Enquanto algumas métricas são amplamente valorizadas, outras geram controvérsia ou se mostram relevantes apenas para determinados grupos. Essa diversidade nas percepções dos *stakeholders* reflete as distintas perspectivas e prioridades de cada grupo em relação à formação médica, ressaltando a importância de adotar uma abordagem equilibrada que incorpore todas essas vozes. Esse sistema de avaliação holístico busca satisfazer as expectativas e necessidades das diversas partes interessadas da residência médica.

A valorização dos “marcos de competência e atividades profissionais confiabilizadoras” evidencia uma crescente preferência por avaliações mais práticas e orientadas para o desempenho clínico. Por exemplo, no Brasil, a colaboração entre a CNRM e associações de especialidades resultou na atualização dos requisitos mínimos de todos os programas de residência médica com a introdução de Matrizes de Competências. Quanto à “avaliação por docentes/preceptores”, apesar das avaliações positivas dessa métrica, surgiram questionamentos quanto à habilidade de profissionais sem formação específica em preceptoria para avaliar residentes de maneira objetiva e imparcial. Os painelistas ressaltaram a importância do treinamento e da capacitação dos instrutores para assegurar processos de avaliação justos e consistentes. Por sua vez, “testes escritos ou orais” enfrentam resistência por dúvidas sobre sua capacidade de avaliar competências do “mundo real”, sugerindo uma revisão para que reflitam melhor as demandas da prática médica. Esse debate é particularmente pertinente nas discussões sobre o “Exame de Ordem na Medicina” e o “Teste de Progresso”, ambos representando diretrizes avaliativas com perspectivas divergentes entre diferentes setores da sociedade. Da mesma forma, a “avaliação por pares” levantou questões sobre potenciais vieses e o risco de transformá-la em uma competição de popularidade que pode prejudicar a dinâmica da equipe. Finalmente, “autoavaliação” dividiu significativamente opiniões nas rodadas do painel: enquanto alguns a consideram essencial para a reflexão e o desenvolvimento pessoal e profissional, outros questionam sua objetividade e precisão, levantando a proposta de que seja complementada por avaliações mais objetivas.

A alta valorização da métrica “envolvimento dos residentes em atividades de pesquisa e reconhecimento acadêmico” sugere que os *stakeholders* prezam uma educação que transcende o ensino de habilidades clínicas, preparando os residentes para contribuições acadêmicas significativas. Em contraste, a rejeição unânime à adoção de métricas acadêmicas como representativas dos interesses dos consumidores diretos pode indicar uma preferência de pacientes e familiares pela qualidade do cuidado clínico direto, em detrimento de indicadores de “sucesso na carreira médica”. Tal divergência

aponta para um possível descompasso entre as expectativas dos profissionais de saúde em formação e as necessidades percebidas pelos “clientes” mais diretos, sublinhando a importância de melhorar a comunicação entre os programas de residência e seus diversos públicos, para alinhar perspectivas. Por fim, os *insights* que surgiram no painel em torno da categoria “questões acadêmicas” ampliaram a discussão existente na literatura sobre potenciais vieses nas métricas de “aprovação de ex-residentes em programas de especialização”, “papéis de liderança acadêmica” e “reconhecimento e prestígio do programa”, sublinhando a importância do desenvolvimento de indicadores mais precisos e objetivos que reflitam de forma fidedigna a qualidade da residência médica.

A ênfase dada à categoria “bem-estar” evidencia a consciência dos *stakeholders* sobre as pressões e o estresse inerentes à formação e à prática médica. Além disso, reflete a compreensão do bem-estar físico e mental como algo fundamental tanto para os profissionais de saúde quanto para a qualidade do atendimento aos pacientes. Os painelistas destacaram a falta de um consenso claro sobre o significado de “bem-estar” e a necessidade de explorar como diferentes grupos definem, percebem e valorizam esse conceito. Todavia, houve concordância sobre a centralidade dessa métrica na estruturação de um sistema de avaliação da qualidade dos programas de residência.

A avaliação altamente favorável da métrica “adaptabilidade” salienta a importância da flexibilidade e resiliência dos médicos em diferentes situações clínicas e contextos, competências essenciais evidenciadas pela pandemia da Covid-19. Da mesma forma, as avaliações favoráveis de “métricas de desempenho clínico” e “competência procedural” refletem a priorização de indicadores de competência técnica e clínica, com ênfase na capacidade dos residentes de realizar procedimentos médicos de forma segura e eficaz. Por sua vez, “empatia” e “satisfação do paciente” enfrentaram avaliações mistas e aquém do esperado pelos autores, possivelmente devido à sua natureza subjetiva e aos desafios de medição e interpretação dessas métricas. Influenciados por uma série de fatores externos, esses indicadores podem não refletir de forma isolada a qualidade dos programas. Diante disso, os painelistas enfatizaram que a avaliação das *soft skills* na prática médica, como habilidades interpessoais, é mais desafiadora do que a avaliação das *hard skills*, como competências técnicas específicas. A preocupação levantada durante as rodadas do painel sobre a ênfase excessiva na “satisfação do paciente” (ou “espetacularização da medicina”) reflete um fenômeno paralelo observado nas mídias sociais, em que profissionais de saúde e instituições frequentemente compartilham conteúdo visando maximizar a popularidade, a satisfação e o engajamento

do público, em detrimento da precisão e da profundidade científica. Concluindo as métricas, “serviço comunitário” e “ativismo em saúde” foram reconhecidos por ressaltarem o papel social e político da medicina, embora tenham suscitado debates sobre até que ponto o engajamento dos residentes na defesa da saúde pública pode impactar sua formação técnico-científica. As diferentes perspectivas sobre essas métricas ilustram as complexas interações entre a prática clínica, os interesses acadêmicos e as responsabilidades sociais dos médicos, evidenciando a necessidade de buscar um equilíbrio nos processos avaliativos.

A inclusão de diversos *stakeholders* no processo de avaliação da qualidade da residência médica não apenas enriquece o sistema, mas também o valida, assegurando sua relevância e aceitação. Dada a diversidade e complexidade dos atores envolvidos nos programas, pode ser desafiador atender integralmente a todas as suas necessidades e expectativas, que podem ser variadas e, ocasionalmente, conflitantes. Portanto, torna-se essencial que cada programa identifique, em um dado momento, quais são as partes interessadas mais relevantes para suas metas e objetivos específicos, priorizando as métricas que mais ressoam com esses grupos-chave. Essencialmente, o programa deve consultar esses “clientes” sobre os indicadores críticos e as informações que moldam suas percepções acerca da eficácia do programa. A organização desses dados em um *framework* de análise, como as matrizes apresentadas neste estudo, facilita a identificação de parâmetros centrais para múltiplos *stakeholders* e aqueles cruciais para os decisores de alto impacto. Paralelamente, é vital assegurar que as métricas selecionadas estejam alinhadas com a visão, a missão e os objetivos do programa, garantindo que a avaliação da qualidade esteja em conformidade com o seu planejamento estratégico. Por fim, a revisão periódica dessas métricas é importante para manter seu alinhamento às evoluções nas perspectivas dos *stakeholders* e na prática médica. Essa abordagem dinâmica fomenta uma avaliação da qualidade dos programas justa/equilibrada, objetiva e confiável.

Algumas limitações desta revisão merecem atenção. Primeiramente, as revisões de escopo não contemplam avaliação do risco de viés ou análises críticas dos estudos incluídos, concentrando-se mais na abrangência do que na qualidade metodológica. Isso limita nossa capacidade de emitir recomendações para políticas ou práticas na avaliação da qualidade dos programas de residência médica. Ademais, existe o risco de não termos capturado todos os estudos relevantes, apesar dos esforços para completude. Essa limitação pode ser atribuída à seleção restrita de bases de dados, à possível omissão de termos de busca relevantes, à não inclusão de “literatura cinzenta” ou à exclusão de estudos em idiomas não

previstos inicialmente. Por fim, a estratégia de busca pode ter privilegiado publicações nas áreas da saúde e educação, negligenciando contribuições de outras disciplinas.

O método Delphi também possui suas limitações. Primeiro, sua eficácia é fortemente dependente da experiência e qualificação dos participantes, estando suscetível a vieses dos *stakeholders*. Além disso, o *feedback* controlado limita a discussão aberta e aprofundada sobre os temas, particularmente em cenários de opiniões divergentes. Finalmente, as conclusões derivadas das opiniões dos painelistas podem não ser universalmente aplicáveis, restringindo a generalização dos resultados. É importante ressaltar que a estratégia de validação por especialistas, embora valiosa, não elimina a necessidade de investigações empíricas adicionais sobre o assunto. Portanto, os resultados devem ser interpretados à luz dessas limitações, reconhecendo a importância de pesquisas futuras para validar e ampliar os nossos achados.

CONCLUSÃO

Ao mapear os *stakeholders* da residência médica, bem como identificar, categorizar e analisar as métricas de avaliação da qualidade mais prevalentes, este estudo ampliou o debate sobre a complexidade das perspectivas em torno da formação médica. A diversidade de atores envolvidos justifica valorizações distintas das várias dimensões da qualidade, reforçando a conclusão de que métricas isoladas não capturam integralmente a qualidade dos programas. Na prática, especialmente em um país de dimensões continentais como o Brasil, com realidades sociais e regionais muito distintas, os resultados sublinham a importância da implementação de sistemas de avaliação da qualidade que sejam equilibrados e alinhados com as expectativas e necessidades dos principais *stakeholders*.

CONTRIBUIÇÃO DOS AUTORES

Marcos Lopes de Miranda participou da concepção do projeto, do delineamento metodológico, da coleta e análise estatística dos dados, da interpretação dos resultados, da redação do manuscrito e da redação e aprovação da versão final do artigo. Sérgio da Cunha participou da concepção do projeto, do delineamento metodológico, da interpretação dos resultados e da redação e aprovação da versão final do artigo. Leila Barbosa participou da concepção do projeto, do delineamento metodológico, da redação do manuscrito e da redação e aprovação da versão final do artigo. Michelle de Moura Balarini e Suzane Almeida dos Santos Pinto participaram da coleta e análise estatística dos dados, da interpretação dos resultados, da redação do manuscrito e da redação e aprovação da versão final do artigo. Denise Herdy Afonso participou da interpretação dos resultados, da

redação do manuscrito e da redação e aprovação da versão final do artigo.

CONFLITO DE INTERESSES

Declaramos não haver conflito de interesses.

FINANCIAMENTO

Declaramos não haver financiamento.

REFERÊNCIAS

- Scheffer M, Guilloux AGA, Miotto BA, Almeida CJ. Demografia médica no Brasil 2023. São Paulo: FMUSP, AMB; 2023.
- Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467-73.
- Peters MDJ, Godfrey C, Mclnerney P, Munn Z, Tricco AC, Khalil, H. Scoping reviews (2020 version). In: Aromataris E, Munn Z, editors. *JBI Manual for Evidence Synthesis*. Adelaide: JBI; 2020.
- Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan – a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210-9.
- van de Schoot R, Bruin J de, Schram R, Zahedi P, Boer J de, Weijdemans F, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nat Mach Intell*. 2021;3:125-33.
- Campbell SM, Shield T, Rogers A, Gask L. How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings? *Qual Saf Health Care*. 2004;13(6):428-34.
- Niederberger M, Spranger J. Delphi technique in health sciences: a map. *Front Public Health*. 2020;8:457-66.
- Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007;30(4):459-67.
- Elliott RL, Juthani NV, Rubin EH, Greenfeld D, Skelton WD, Yudkowsky R. Quality in residency training: toward a broader, multidimensional definition. *Acad Med*. 1996;71(3):243-7.
- Simpson D, Riddle JM, Hamel Jr DL, Balmer DF. blueprinting program evaluation evidence through the lens of key stakeholders. *J Grad Med Educ*. 2020;12(5):629-30.
- Abreu-Reis P, Oldoni C, de-Souza GAL, Bettega AL, Góes MN, Sarquis LM, et al. Aspectos psicológicos e qualidade de vida na Residência Médica. *Rev Col Bras Cir*. 2019;46(1):e2050.
- Agarwal V, Bump GM, Heller MT, Chen LW, Branstetter BF 4th, Amesur NB, et al. Resident case volume correlates with clinical performance: finding the sweet spot. *Acad Radiol*. 2019;26(1):136-40.
- Andrews J, Jones C, Tetrault J, Coontz K. Advocacy training for residents: insights from Tulane's internal medicine residency program. *Acad Med*. 2019;94(2):204-7.
- Rosas JBM, Lopes Junior A, Moreira JV, Afonso MPD, Sarno MM, Borret RH do ES, et al. Recomendações para a qualidade dos Programas de Residência de Medicina de Família e Comunidade no Brasil. *Rev Bras Med Fam Comunidade*. 2020;15(42):2509-19.
- Blanchard J. Board scores and resident performance: is there a link? *Ann Emerg Med*. 2000;36(1):64-7.
- Busireddy KR, Miller JA, Ellison K, Ren V, Qayyum R, Panda M. Efficacy of interventions to reduce resident physician burnout: a systematic review. *J Grad Med Educ*. 2017;9(3):294-301.
- Chole RA, Ogden MA. Predictors of future success in otolaryngology residency applicants. *Arch Otolaryngol Head Neck Surg*. 2012;138(8):707-12.
- Colthart I, Bagnall G, Evans A, Allbutt H, Haig A, Illing J, et al. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Med Teach*. 2008;30(2):124-45.
- Cutrer WB, Miller B, Pusic MV, Mejicano G, Mangrulkar RS, Gruppen LD, et al. Fostering the development of master adaptive learners: a conceptual model to guide skill acquisition in medical education. *Acad Med*. 2017;92(1):70-5.
- Oliveira GS de, Chang R, Fitzgerald PC, Almeida MD, Castro-Alves LS, Ahmad S, et al. The prevalence of burnout and depression and their association with adherence to safety and practice standards: a survey of United States anesthesiology trainees. *Anesth Analg*. 2013;117(1):182-93.
- Farley H, Enguidanos ER, Coletti CM, Honigman L, Mazzeo A, Pinson TB, et al. Patient satisfaction surveys and quality of care: an information paper. *Ann Emerg Med*. 2014;64(4):351-7.
- Giroto JA, Adams NS, Janis JE, Brandt KE, Slezak SS. Performance on the plastic surgery in-service examination can predict success on the American Board of Plastic Surgery written examination. *Plast Reconstr Surg*. 2019;143(5):1099e-1105e.
- Goldstein AO, Calleson D, Curtis P, Hemphill B, Gamble G, Steiner B, et al. Community service by North Carolina family physicians. *J Am Board Fam Pract*. 2005;18(1):48-56.
- Grewal SG, Yeung LS, Brandes SB. Predictors of success in a urology residency program. *J Surg Educ*. 2013;70(1):138-43.
- Hart D, Franzen D, Beeson M, Bhat R, Kulkarni M, Thibodeau L, et al. Integration of entrustable professional activities with the milestones for emergency medicine residents. *West J Emerg Med*. 2019;20(1):35-42.
- Herbers JE, Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. How accurate are faculty evaluations of clinical competence? *J Gen Intern Med*. 1989;4(3):202-8.
- Jewell CM, Kraut AS, Miller DT, Ray KA, Werley EB, Schnapp BH. Metrics of resident achievement for defining program aims. *West J Emerg Med*. 2022;23(1):1-8.
- Kay C, Jackson JL, Frank M. The relationship between internal medicine residency graduate performance on the ABIM certifying examination, yearly in-service training examinations, and the USMLE Step 1 examination. *Acad Med*. 2015;90(1):100-4.
- Ko CY, Whang EE, Longmire WP, McFadden DW. Improving the surgeon's participation in research: is it a problem of training or priority? *J Surg Res*. 2000;91(1):5-8.
- Kruger J, Dunning D. Unskilled and unaware of it: how difficulties recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol*. 1999;77(6):1121-34.
- Lang SC, Weygandt PL, Darling T, Gravenor S, Evans JJ, Schmidt MJ, et al. Measuring the correlation between emergency medicine resident and attending physician patient satisfaction scores using Press Ganey. *AEM Educ Train*. 2017;1(3):179-84.
- Levy D, Dvorkin R, Schwartz A, Zimmerman S, Li F. Correlation of the emergency medicine resident in-service examination with the American Osteopathic Board of Emergency Medicine Part I. *West J Emerg Med*. 2014;15(1):45-50.
- Loeb AE, Rao SS, Ficke JR, Morris CD, Riley LH 3rd, Levin AS. Departmental experience and lessons learned with accelerated introduction of telemedicine during the Covid-19 crisis. *J Am Acad Orthop Surg*. 2020;28(11):e469-76.
- Meier AH, Gruessner A, Cooney RN. Using the ACGME Milestones for Resident Self-Evaluation and Faculty Engagement. *J Surg Educ*. 2016;73(6):e150-7.
- Neumann M, Edelhäuser F, Tauschel D, Fischer MR, Wirtz M, Woopen C, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. *Acad Med*. 2011;86(8):996-1009.
- O'Sullivan PS, Cogbill KK, McClain T, Reckase MD, Clardy JA. Portfolios as a novel approach for residency evaluation. *Acad Psychiatry*. 2002;26(3):173-9.

37. Olivero W, Wang H, Vinson D, Pierce C, Trumbull S. Correlation between Press Ganey scores and quality outcomes from the national neurosurgery quality and outcomes database (lumbar spine) for a hospital employed neurosurgical practice. *Neurosurgery*. 2018;65(CN_suppl_1):34-6.
38. Parlier AB, Galvin SL, Thach S, Kruidenier D, Fagan EB. The road to rural primary care: a narrative review of factors that help develop, recruit, and retain rural primary care physicians. *Acad Med*. 2018;93(1):130-40.
39. Potts JR. Surgical program accreditation and case logs: what is the meaning of the minima? *J Am Coll Surg*. 2019;229(4):431-5.
40. Riall TS, Teiman J, Chang M, Cole D, Leighn T, McClafferty H, et al. Maintaining the fire but avoiding burnout: implementation and evaluation of a resident well-being program. *J Am Coll Surg*. 2018;226(4):369-79.
41. Romão GS, Fernandes CE, Silva Filho AL, Ribeiro Rocha S, Sá MFS de. Como avaliar programas de residência a partir do Teste de Progresso? *Rev Bras Educ Med*. 2022;46:e149.
42. Romão GS, Fernandes CE, Silva Filho AL, Sá MFS de. O Teste de Progresso na residência médica em ginecologia e obstetrícia: a experiência nacional. *Rev Bras Educ Med*. 2022;46:e153.
43. Schumacher DJ, Martini A, Sobolewski B, Carraccio C, Holmboe E, Busari J, et al. Use of resident-sensitive quality measure data in entrustment decision making: a qualitative study of clinical competency committee members at one pediatric residency. *Acad Med*. 2020;95(11):1726-35.
44. Shah D, Haisch CE, Noland SL. Case reporting, competence, and confidence: a discrepancy in the numbers. *J Surg Educ*. 2018;75(2):304-12.
45. Shappell E, Schnapp B. The F word: how "fit" threatens the validity of resident recruitment. *J Grad Med Educ*. 2019;11(6):635-6.
46. Sharp B, Johnson J, Hamedani AG, Hakes EB, Patterson BW. What are we measuring? Evaluating physician-specific satisfaction scores between emergency departments. *West J Emerg Med*. 2019;20(3):454-9.
47. Shellito JL, Osland JS, Helmer SD, Chang FC. American Board of Surgery examinations: can we identify surgery residency applicants and residents who will pass the examinations on the first attempt? *Am J Surg*. 2010;199(2):216-22.
48. Smith KE, Norman GJ, Decety J. The complexity of empathy during medical school training: evidence for positive changes. *Med Educ*. 2017;51(11):1146-59.
49. Spitzer AB, Gage MJ, Looze CA, Walsh M, Zuckerman JD, Egol KA. Factors associated with successful performance in an orthopaedic surgery residency. *J Bone Joint Surg Am*. 2009;91(11):2750-5.
50. Thompson RH, Lohse CM, Husmann DA, Leibovich BC, Gettman MT. Predictors of a successful urology resident using medical student application materials. *Urology*. 2017;108:22-8.
51. Torbeck L, Canal DF, Choi J. Is our residency program successful? Structuring an outcomes assessment system as a component of program evaluation. *J Surg Educ*. 2014;71(1):73-8.
52. Van Rosendaal GM, Jennett PA. Comparing peer and faculty evaluations in an internal medicine residency. *Acad Med*. 1994;69(4):299-303.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.