

Critérios Metodológicos para a Avaliação de Programas de Intervenção Em Psicologia

Helen Durgante¹, Débora Dalbosco Dell’Aglío
Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, Brasil

RESUMO

A área de avaliação de programas de intervenção em Psicologia ainda é incipiente no Brasil, embora o processo de monitoramento e avaliação de programas seja fundamental para o aperfeiçoamento de práticas e alocação de recursos para fins de desenvolvimento social. Este artigo teórico visa discutir conceitos/definições na área de avaliação de programas, além de apresentar critérios científicos adotados para avaliação de eficácia e efetividade de programas de intervenção. São apresentados critérios metodológicos, de acordo com agências internacionais de referência, para intervenções de promoção de saúde e prevenção de doenças, assim como diferentes indicadores de resultados de programas de intervenção. O artigo destaca também critérios para avaliar estudos de viabilidade/piloto, como prática ainda pouco comum, tanto em contexto nacional quanto internacional. Por fim, é ressaltada a importância dos critérios científicos na elaboração, desenvolvimento e avaliação de programas de intervenção na área de saúde.

Palavras-chave: programa, avaliação, eficácia, efetividade, viabilidade.

ABSTRACT – Methodological criteria for the evaluation of intervention programs in Psychology

The evaluation area of intervention programs in Psychology is still incipient in Brazil, although the process of monitoring and evaluating programs is fundamental for the improvement of practices and allocation of resources for social development purposes. This theoretical article aims to discuss concepts/definitions in the area of program evaluation, besides presenting scientific criteria adopted to evaluate the efficacy and effectiveness of intervention programs. Methodological criteria are presented, according to international reference agencies, for interventions to promote health and disease prevention, as well as different indicators of the results of intervention programs. The article also outlines criteria to evaluate feasibility/pilot studies, as a practice still uncommon, both in the national and international context. Finally, the importance of scientific criteria in the elaboration, development and evaluation of intervention programs in the health area is emphasized.

Keywords: Program; Evaluation; Efficiency; Effectiveness; Viability.

RESUMEN – Criterios metodológicos para evaluación de programas de intervención en Psicología

En Brasil, el área de evaluación de programas de intervención en Psicología todavía es incipiente, aunque el proceso de monitoreo y evaluación de programas es fundamental para el perfeccionamiento de prácticas y asignación de recursos para fines de desarrollo social. Este artículo teórico busca discutir conceptos/definiciones en el área de evaluación de programas, además de presentar criterios científicos adoptados para evaluación de eficacia y efectividad de programas de intervención. Se presentan criterios metodológicos de acuerdo con agencias internacionales de referencia, para intervenciones de promoción en salud y prevención de enfermedades, así como diferentes indicadores de resultados de programas de intervención. El artículo destaca también criterios para evaluar estudios de viabilidad/piloto, como práctica todavía poco común, tanto en el contexto nacional como internacional. Por último, se resalta la importancia de los criterios científicos en la elaboración, desarrollo y evaluación de programas de intervención en el área de salud.

Palabras clave: Programa; Evaluación; Eficacia; Efectividad; Viabilidad.

A área de avaliação de programas de intervenção em Psicologia ainda é incipiente no Brasil. Por “programa”, entende-se uma proposta estruturada de intervenção profissional, com base em dados da literatura científica e modelo teórico fundamentado, de modo a atingir objetivos específicos, tanto para a promoção de saúde, quanto para a prevenção de doenças, em todos os níveis

(Fernandez-Ballesteros, 2001). O processo de monitoramento e avaliação de programas é fundamental para o aperfeiçoamento de práticas, otimização da alocação de recursos humanos e materiais, dinamização da gestão do programa para fins de desenvolvimento social, reestruturação e planejamento de novos programas (UNDP, 2009). Isso quer dizer que a avaliação é tão importante

¹ Endereço para correspondência: Instituto de Psicologia, Universidade Federal do Rio Grande do Sul – UFRGS. Rua Ramiro Barcellos, 2600, sala 115, Santa Cecília, 90035-003, Porto Alegre, RS. Tel.: (55) 98163-1201. E-mail: helen.durga@gmail.com
Apoio Financeiro: Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq e Fundação de Amparo à Pesquisa do Estado do RGS (FAPERGS).

quanto a implementação do programa propriamente dito e deveria compor parte integrante do processo de desenvolvimento do programa per se. Nesse caso, durante a fase de desenvolvimento ou planejamento do programa, a avaliação de resultados deveria ser incluída a fim de fornecer subsídios baseados em evidência empírica e possibilitar a inferência sobre melhorias e adequações necessárias nos processos de desenvolvimento e implementação dele (UNDP, 2009).

Este artigo teórico visa esclarecer e aprofundar conceitos fundamentais acerca da avaliação de programas em Psicologia, mais especificamente, com relação às definições e aos critérios científicos adotados para avaliação de eficácia e efetividade de programas de intervenção. São apresentados critérios metodológicos, até então estabelecidos como “boas práticas” para a avaliação de programas, de acordo com agências internacionais de referência na área, para intervenções de promoção de saúde e prevenção de doenças: *American Psychological Association* (APA, 2002) e *Society for Prevention Research* (Gottfredson et al., 2015; SPR, 2004).

Salienta-se que este trabalho não pretende esgotar todas as possibilidades sobre critérios existentes para a avaliação de programas, uma vez que outras agências internacionais também acrescentam/sugerem critérios nessa área. Pretende-se, portanto, facilitar a compreensão acerca de parâmetros comuns, consensuais, utilizados para avaliação de programas de promoção e prevenção em saúde. Ressalta-se também que novos critérios podem e devem ser introduzidos, tanto quanto forem demonstrados úteis para essas práticas e de acordo com o rigor científico-metodológico exigido. Por esse motivo, os(as) pesquisadores(as) são encorajados a divulgarem seus achados empíricos, uma vez cumpridos os critérios já existentes para estudos de avaliação de programas. Isso pode dinamizar práticas futuras e servir como dados adicionais para a elaboração de novos critérios científicos a serem utilizados na avaliação de programas de intervenção, também em contexto nacional. Por fim, o artigo aponta critérios existentes para avaliar estudos de viabilidade como prática ainda pouco comum, tanto em contexto nacional quanto internacional.

A Avaliação de Programas em Psicologia

O estabelecimento de mudanças recentes na ciência psicológica demarcou um novo paradigma científico para investigar aspectos relacionados à saúde em sua integralidade (Johnson, Weinman, & Charte, 2011). Esse paradigma ressaltou a importância do papel do psicólogo, enquanto membro fundamental nas equipes de saúde, para transpor questões puramente relativas a transtornos mentais, ampliar seu modo de atuação e abranger também a promoção e manutenção da saúde. Por promoção de saúde entende-se um amplo campo de atuação e medidas gerais, educativas, que visam à capacitação, empoderamento e autonomia do indivíduo e comunidades, a fim

de modificar determinantes em saúde e promover melhor qualidade de vida e bem-estar (Sparks, 2013; WHO, 2010). Isso inclui o desenvolvimento de competências pessoais e sociais e a participação ativa do indivíduo para a criação de ambientes mais saudáveis. De acordo com a Organização Mundial de Saúde (WHO, 2010), a saúde é compreendida não somente como a inexistência de doenças, mas também com relação aos aspectos de vida dos indivíduos, como fatores multicausais e fundamentais para a constituição da saúde.

Como resultado dessas mudanças, as abordagens em Psicologia começaram a priorizar também intervenções primárias, com base na promoção, prevenção e manutenção da saúde, por meio de programas de tratamentos sistemáticos, diagnóstico precoce, diretrizes para ampliar a qualidade das intervenções em saúde e subseqüente aperfeiçoamento de pessoal e de políticas públicas na área (Fernandez-Ballesteros, 2001; Straub, 2012). Nesse sentido, intervenção inclui serviços disponibilizados por profissionais, programas de saúde, avaliação psicológica, diagnósticos, tratamentos preventivos, psicoterapias e consultas, tanto para o(a) paciente quanto para as demais partes envolvidas na intervenção (Baker & McFall, 2014).

Com relação aos aspectos metodológicos, os programas de intervenção na área da Psicologia passaram a considerar critérios específicos para a avaliação de resultados das abordagens utilizadas pelos profissionais para a promoção da saúde e redução da doença – programas preventivos (APA, 2002). Esses critérios podem ser melhor compreendidos como pré-requisitos mínimos (essenciais), ou desejáveis (suplementares), que devem ser satisfeitos/atingidos, para que se possa considerar que o programa de intervenção foi submetido a testes, cumprindo com rigor científico, e apresenta “verificada eficácia” ou “verificada efetividade” para aquilo que se propõe a atingir (Gottfredson et al., 2015). Assim, a eficácia dos programas pode ser avaliada no que se refere aos efeitos produzidos pela intervenção em um contexto controlado, para “tal” população e contexto, em um tempo “T”, resultando em maiores ganhos do que perdas e com base nos objetivos propostos pelo programa *a priori*, quando comparados a outras intervenções ou grupo controle (APA, 2002; Flay et al., 2005; Gottfredson et al., 2015). Já a efetividade, diz respeito aos efeitos da intervenção para uma amostra e/ou população específica, em condições não controladas, “do mundo real”, ou fora do *setting* experimental (Flay et al., 2005).

De acordo com a *American Psychological Association*, o conceito de “Prática Baseada em Evidências” representa a integração de dados empíricos e especialização clínica profissional, levando em consideração as características do paciente, sua cultura e preferências (APA, 2006). Esse método de pesquisa aplicada, baseado na utilização prático-metodológica de evidências empíricas, tem contribuído para avanços na constatação de resultados

de diferentes programas para a promoção de saúde para amostras específicas (Flay et al., 2005). Isso tende a favorecer não somente o aperfeiçoamento de práticas clínicas no atendimento aos pacientes como também fornece sustentação para a tomada de decisões assertivas quanto ao investimento, ou não, de recursos humanos e materiais em determinados programas oferecidos à população (Gottfredson et al., 2015).

Quanto à investigação científica dos benefícios de programas oferecidos aos pacientes/clientes, diferentes tipos de intervenções – *Empirically Supported Treatments* – foram constatadas como apresentando validade empírica adequada (testagem randomizada e manualização) para as práticas profissionais (APA, 2006). No que diz respeito

aos critérios utilizados para a análise da qualidade e resultados dos programas, o uso sistemático de métodos empíricos tende a auxiliar julgamentos subjetivos dos profissionais quanto à validade ou fracasso de suas práticas (Flay et al., 2005).

Neste estudo, foi realizado um processo de busca e seleção de critérios, considerados essenciais, e subcritérios, considerados desejáveis, para que possam ser aferidas a eficácia e/ou efetividade de programas, de acordo com dados da American Psychological Association (APA, 2002) e *Society for Prevention Research* (Gottfredson et al., 2015; SPR, 2004). Esse processo é ilustrado no diagrama de fluxo apresentado na Figura 1, adaptado de acordo com o modelo PRISMA da Cochrane (Higgins & Green, 2011).

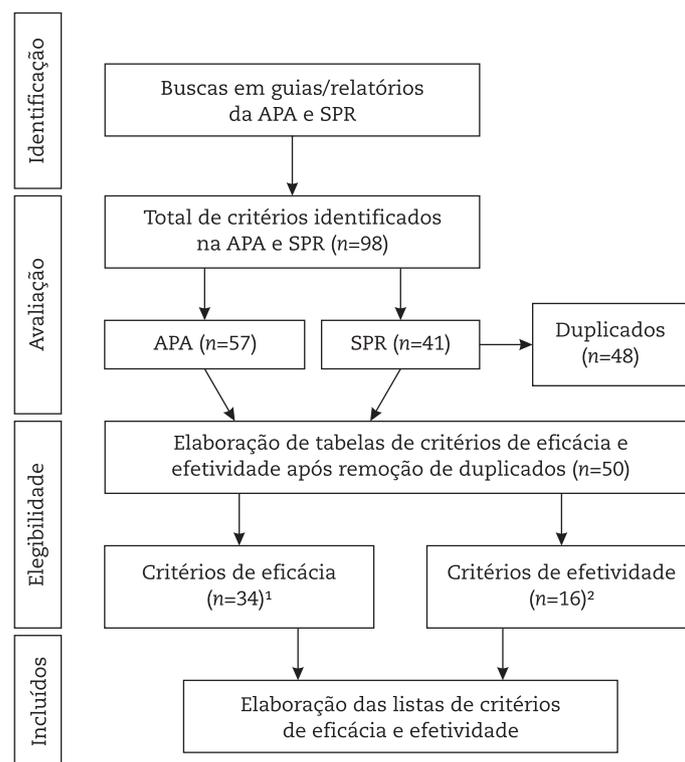


Figura 1. Estrutura e cargas fatoriais padronizadas da GPSRBS

Nota. 1=17 critérios essenciais e 17 critérios desejáveis; 2=16 critérios essenciais

Assim, a partir da busca realizada, aos moldes de convenções da APA (2002) e SPR (Gottfredson et al., 2015; SPR, 2004) são apresentados os critérios fundamentais para a avaliação de eficácia dos programas:

1. As diretrizes do programa devem ser baseadas em ampla e análise da literatura empírica relevante;
2. As diretrizes do programa devem considerar opinião clínica, observação clínica sistematizada e consenso entre os especialistas reconhecidos, representando diferentes pontos de vista na área;
3. As diretrizes devem especificar quais resultados a

intervenção visa produzir, e evidências devem ser fornecidas para cada resultado;

4. É importante considerar o método e lógica de seleção dos participantes e como a amostra representa a população (equivalência entre grupos) e fenômenos de interesse;

5. Deve ser utilizado o método experimental (randomização da amostra) ou quase experimental;

6. Medidas utilizadas para avaliar os resultados devem ter validade psicométrica;

7. Os resultados devem ser melhores que: (a) não tratamento; (b) resultados de outras intervenções;

(c) benefícios percebidos por simplesmente estar recebendo um tratamento;

8. Uma declaração de eficácia deve ser formulada de que “o programa ou política X é eficaz para produzir resultados Y para a população Z”;

9. Diferenças individuais pré-teste (antes da intervenção) devem ser avaliadas;

10. Para resultados a longo prazo, deve haver pelo menos um acompanhamento de seguimento (seis meses após a intervenção). Esse intervalo pode variar para diferentes tipos de intervenções;

11. Todos os resultados devem ser apresentados (positivos, não significativos ou negativos);

12. Devem ser utilizadas estatísticas que assegurem estimativas não enviesadas nos resultados;

13. Os resultados devem considerar a significância clínica (benefícios clínicos) da intervenção. Por exemplo: (a) funcionamento dentro dos limites normais; (b)

melhorou, mas não está funcionando dentro dos limites normais; (c) melhorou; (d) nenhuma mudança; e (e) deterioração;

14. Resultados consistentes são necessários por pelo menos dois outros estudos, de alta qualidade, que atendam a todos os critérios acima referidos;

15. O(s) componente(s) principal(is) da intervenção, tempo para obter efeitos, variáveis mediadoras/moderadoras devem ser descritos nos resultados;

16. A intervenção deve ser detalhada de modo que replicações possam ser conduzidas por outros pesquisadores;

17. Deve-se valiar e relatar o nível de perdas e desistências de participantes, que possam comprometer a generalização dos resultados para outras amostras/populações. Já os subcritérios (desejáveis/completos) para aferir eficácia dos programas foram compilados conforme a Tabela 1.

Tabela 1

Subcritérios (Desejáveis) Científicos para a Avaliação de Eficácia de Programas de Intervenção

1. É desejável utilizar várias medidas e/ou fontes (i.e., auto relatos de estudantes, relatos de pais e professores, observadores independentes) para coleta e análise dos resultados.
2. Deve haver pelo menos uma fonte de dados (medida) coletados por pessoa(s) diferente(s) de quem aplica ou entrega a intervenção.
3. Resultados devem ser analisados utilizando diferentes métodos e fontes convergentes de informação.
4. Os resultados devem ser compatíveis com as metas/objetivos do programa.
5. É desejável considerar medidas de resultados proximais (i.e., variáveis moderadoras).
6. O programa deve ter pelo menos uma condição de grupo controle que não receberá a intervenção testada, para comparar resultados.
7. É desejável medir a implementação/processo do programa.
8. Os resultados da intervenção para comportamentos, ou saúde pública devem ser medidos.
9. Realizar análises de subgrupos (gênero, etnia/raça, nível sócioeconômico, religião, etc.) e apontar para qual grupo o tratamento é melhor indicado.
10. Diferenças pré-teste devem ser analisadas e corrigidas, quando possível.
11. Não deve haver graves efeitos negativos (iatrogênico) em variáveis importantes.
12. É desejável medir os efeitos colaterais, ou iatrogênicos e satisfação dos participantes.
13. Pelo menos uma replicação do programa deve ser feita por investigadores independentes.
14. Aceitação, adesão e envolvimento do público-alvo devem ser avaliados e relatados.
15. É desejável relatar possíveis impactos na saúde pública.
16. É desejável ter informações sobre e custo-benefício do programa.
17. Desejável apresentar as implicações/contribuições do programa para a saúde pública.

Similarmente, para cumprir com critérios de efetividade, os resultados do programa devem atender a todos os critérios de eficácia, mais os seguintes:

1. Manuais, treinamento e suporte técnico devem estar prontamente disponíveis para que outros profissionais (não pesquisadores) possam aplicar o programa “no campo”;

2. É necessária replicação em diferentes populações. Sem replicação, os efeitos encontrados não podem ser generalizados;

3. A replicação deve envolver diferentes agentes de implementação da intervenção;

4. É necessário declarar para quem e em quais condições a intervenção está prevista para ser efetiva;

5. A população-alvo e o método de amostragem devem ser explicados, a fim de explicitar o quanto a amostra representa a população geral;

6. Amostra deve ter tamanho suficiente para possíveis generalizações;

7. A dose deve ser medida, a partir de dois fatores: (a) Integridade-fidelidade da implementação; e (b) Engajamento (aceitação, conformidade, adesão ou envolvimento) do público-alvo da intervenção;

8. A vontade e capacidade dos pacientes de participar das intervenções-tarefas recomendadas devem ser consideradas;

9. Adaptações e modificações de componentes da intervenção devem ser avaliadas e relatadas;

10. É necessário levar em conta a complexidade e a idiosincrasia do estado clínico dos pacientes, incluindo a gravidade, comorbidade e estressores externos (variáveis moderadoras/mediadoras);

11. Recomenda-se levar em conta dados sobre como as diferenças (habilidade profissional, experiência, sexo, idioma, etc.) entre os profissionais de saúde podem afetar a efetividade do tratamento;

12. Randomização da amostra é o ideal, mas outras alternativas podem ser sugeridas (descontinuidade da regressão, séries temporais ou delineamentos pareados controlados);

13. Se a amostra é heterogênea com relação a variáveis importantes (idade, sexo, etnia/raça, níveis de risco), é desejável relatar análises de subgrupos. Ensaios de efetividade devem incluir número suficiente de indivíduos para cada subgrupo (poder estatístico);

14. As avaliações devem relatar alguma evidência da importância prática, para além dos resultados obtidos (como é o caso de estudos de eficácia), e também análise de impacto na saúde pública;

15. Quando diretrizes considerarem os custos, devem relatar custos diretos, indiretos, a curto e longo prazo para o paciente, para o profissional e ao sistema de saúde, bem como os associados à retenção do tratamento.

16. A avaliação da utilidade clínica dos resultados deve considerar análise de generalização dos efeitos do tratamento.

Vale lembrar que o rigor exigido, em termos de controle metodológico sobre a implementação do programa para ensaios de eficácia (validade interna), sobressai-se àquele exigido para ensaios de efetividade (Shaughnessy, Zechmeister, & Zechmeister, 2012). Por outro lado, aferir efetividade de um programa, quer dizer que, além de ter cumprido com todas as exigências metodológicas necessárias para um ensaio de eficácia, o programa também satisfaz as exigências dos critérios estabelecidos para ensaios de efetividade (Flay et al., 2005; Gottfredson et al., 2015). Ou seja, ensaios de eficácia poderiam ser considerados, em outras palavras, um passo inicial enquanto tentativa/avaliação de resultados de programas. Já ensaios de efetividade pressupõem menor ênfase na validade

interna, porém, incluem uma série de outros tantos critérios, que fornecem base para aferir validade externa da intervenção. Por isso, pode-se conjecturar que, investigar a eficácia de resultados do programa, quer dizer avaliar se o programa atinge seus objetivos em um contexto controlado; enquanto que efetividade de resultados seria, supostamente, um passo subsequente a ser atingido em um modelo lógico, em estudos de replicações do programa (os resultados se sustentam/permanecem em um contexto não controlado, do mundo real? Logo, podem ser generalizados?) (Gottfredson et al., 2015).

Similarmente, desde elaborações iniciais de critérios para avaliação de programas de intervenção, há mais de uma década, as agências têm constantemente atualizado critérios por meio de observações sistemáticas, ensaios randomizados e avaliações de resultados compilados em diferentes estudos de promoção e prevenção em saúde, de modo que os critérios possam ser compreendidos de maneira menos linear. Ou seja, compreender as evidências obtidas em cada estudo (resultados de avaliações de diferentes critérios), como sendo cumulativas, onde cada novo ensaio com a intervenção possa acrescentar dados sobre sua eficácia e/ou efetividade. Nesse caso, a partir de diferentes estudos, pode-se cumprir com novos critérios, até então não satisfeitos em estudos passados, e somar dados para um corpo de conhecimento acerca dos resultados daquela intervenção, seja para aferir sua eficácia ou efetividade (Gottfredson et al., 2015). Desse modo, é possível que, em um mesmo estudo, o(a) pesquisador(a) ou profissional de saúde, compile dados (cumpram com critérios) de eficácia, bem como de efetividade de sua intervenção em saúde, de acordo com delineamento cuidadosamente traçado de sua intervenção.

Por isso, relatórios recentes enfatizam a necessidade de dados de pesquisa serem divulgados, incluindo informações sobre o método utilizado nas intervenções e resultados obtidos, para que o conhecimento avance na área e novos critérios possam ser introduzidos para avaliar tanto eficácia quanto efetividade de programas (Gottfredson et al., 2015). Da mesma forma, sem a elaboração e utilização de guias ou diretrizes para avaliar intervenções, o resultado dos programas é inferido com base apenas no julgamento do profissional que desenvolve o programa, ficando as análises sujeitas às incertezas e subjetividade do profissional. Guias ou diretrizes para serviços de saúde podem ser definidas como declarações, ou sugestões, que norteiam comportamentos e práticas profissionais, para uma otimização dos serviços oferecidos (APA, 2002). É esperado que diretrizes possam servir como padrão a ser seguido na avaliação de programas, com o profissional mantendo sempre seu senso crítico para avaliar necessidades e diferenças individuais, culturais e históricas dos indivíduos.

As diretrizes apresentadas não pretendem servir como sistemas rígidos para a avaliação de programas em Psicologia, mas sim para melhor indicar padrões de

mudanças ou de ajustes no tratamento oferecido aos usuários dos programas (APA, 2002). Assim, faz-se necessário que os profissionais avaliem também, o quanto cada critério sugerido é aplicável ou útil para avaliar sua prática e, de acordo com o delineamento traçado para cada programa, evitando a falta de escrutínio, debates e cientificidade na utilização desses critérios em um dado contexto.

Crítérios para Avaliação de Estudos de Viabilidade

Estudos de viabilidade, por definição, são pedaços/recortes de uma pesquisa maior, conduzidos previamente ao estudo principal, e visam responder, primordialmente, à questão: Este estudo deve realmente ser conduzido. Um dos objetivos de estudos de viabilidade é averiguar o potencial de sucesso, ou o quanto a intervenção principal tende a ser bem-sucedida, após medidas para retificações no delineamento serem tomadas, como resultado de estudos de viabilidade (NIHR, 2017). Assim, busca-se reduzir ameaças à validade dos resultados do estudo principal ou da intervenção propriamente dita (Tickle-Degnen, 2013).

Dessa forma, estudos de viabilidade são usados como parâmetros para a elaboração do delineamento de estudos maiores e são apropriados quando há evidências para, como hipótese, seguir com um estudo maior, acerca de um determinado tema ou problema a ser resolvido (HRC, 2017; NIHR, 2017). Nesse caso, um estudo de viabilidade prévio ao estudo maior – estudo piloto ou intervenção – pode gerar dados, não apenas sobre condições adequadas/necessárias para implementação da intervenção, testes, adaptações/refinamentos de procedimentos e possíveis correções de erros metodológicos, mas também sobre a viabilidade de um estudo maior no futuro (estudo piloto ou intervenção propriamente dita). Como resultado, estudos de viabilidade servem para constituir as bases de intervenções ‘padrão ouro’, sendo estas, atualmente, ensaios clínicos randomizados para a investigação de eficácia-efetividade de intervenções em saúde (Tickle-Degnen, 2013).

Para fins de esclarecimentos, é fundamental diferenciar, conceitualmente, estudo de viabilidade de estudo piloto. O primeiro, refere-se a pequenos recortes de pesquisa de uma intervenção que está por vir, servindo para traçar ajustes na elaboração da intervenção principal. O último, diz respeito à tentativa de operacionalização do estudo principal, ou seja, um estudo piloto é uma versão menor, implementado tal qual a intervenção principal, para verificar se os diferentes componentes da intervenção, em conjunto, funcionam, inclusive a avaliação de medidas de resultado (NIHR, 2017).

Tendo em vista que essas definições são relativamente novas no campo de avaliação de programas, é recomendado que ambos os estudos (de viabilidade e piloto) não incluam em seu delineamento o teste de hipóteses do estudo principal, resumindo-se em avaliar a viabilidade e validade de seguir, ou não, com a intervenção principal (Tickle-Degnen, 2013). Isso se justifica, inclusive, pois, em ambos

os ensaios, não são necessárias amostras grandes, como é o caso de intervenções em saúde (ensaios clínicos randomizados). Sendo assim, em casos de estudos com amostras menores, como é o caso de estudos de viabilidade/piloto, pode haver prejuízo no poder estatístico para a constatação de efeitos em variáveis de resultado importantes para a intervenção. Nesse caso, relatar tão somente resultados estatísticos (provavelmente não significativos para a maioria das variáveis de resultado) muda o foco do objetivo principal de estudos de viabilidade e piloto, que é chegar a conclusão se o estudo deve, ou não, ser recomendado para futuro teste de eficácia-efetividade (Bowen et al., 2009). Por isso, é recomendado que resultados de estudos de viabilidade e piloto, sejam aferidos a partir de estatísticas descritivas, plano de análise qualitativa de dados e segundo alguns critérios já sugeridos (ainda pouco explorados) para avaliação de resultados, com base em: (a) processo; (b) recursos; (c) gerenciamento/gestão; (d) embasamento científico no planejamento da intervenção (Thabane et al., 2010).

De acordo com Thabane et al. (2010), a avaliação de processo em estudos de viabilidade e piloto inclui decisões que precisam ser tomadas para a implementação da intervenção, tais como número de participantes selecionados, desistentes, número de recusas (aqueles que se recusaram a participar da intervenção ou randomização), retenção/estudo de seguimento, adesão/engajamento aos procedimentos, critérios de inclusão (muitos flexíveis ou rígidos), coleta de dados (pouco ou muitos dados coletados), compreensão dos participantes (com relação aos protocolos de avaliação e procedimentos da intervenção), tempo para coleta e análise de dados. Recursos referem-se aos problemas técnicos que podem suscitar durante o estudo, como sobrecarga nos serviços (sala de espera, linhas telefônicas e e-mails congestionados), recursos audiovisuais, financeiros para a implementação, coleta e análise de dados (in)disponíveis. Já gerenciamento/gestão visa avaliar problemas humanos na implementação do estudo, como disponibilidade de profissionais treinados e com tempo hábil para o serviço, desafios impostos para o local de serviço onde será implementada a intervenção, capacidades técnicas para a implementação, coleta, gerenciamento e análise de dados. Por fim, de acordo com os autores Thabane et al. (2010), embasamento científico no planejamento da intervenção inclui avaliação sobre aspectos relativos à dose (a quantidade de tratamento é suficiente para se obter os resultados objetivados?), resposta dos participantes ao tratamento (positiva, negativa, neutra, mista), estimativa de efeitos do tratamento e efeitos a longo prazo.

Cabe lembrar que os critérios mencionados servem como guia para o delineamento de estudos de viabilidade e piloto. Porém, dada a (ainda) falta de consenso sobre critérios metodológicos para avaliação desses tipos de estudo (viabilidade e piloto), diferentes autores tendem a sugerir critérios conforme avanços são atingidos em pesquisas como tentativa de ampliar a

compreensão atual sobre estas formas de delineamento de intervenções (Bowen et al., 2009; Tickle-Degnen, 2013). Essa lógica pode parecer contraproducente, uma vez que, para seguir para avaliação sistemática de resultados de eficácia e efetividade, uma intervenção deve, presumidamente, ter passado por verificações prévias em estudos de viabilidade, seguido de estudo piloto (não necessariamente em estudos distintos, ou seja, o mesmo estudo por incluir avaliações de viabilidade, já estando em formato de estudo piloto).

Tendo em vista que não há critérios consensualmente estabelecidos para ensaios de viabilidade e piloto, como é o caso de ensaios de eficácia e efetividade, o(a) pesquisador(a) precisa ter em vista a maior gama de variáveis intervenientes possíveis que possam impactar, tanto positiva quanto negativamente, em seu delineamento de intervenção (Orsmond & Cohn, 2015). Isso é particularmente importante quando: (a) existem poucos estudos sobre o tema de interesse; (b) é necessário firmar parcerias com locais de serviço ou na própria rede de atendimento, com a oferta de um novo tratamento de saúde; (c) estudos prévios conduzidos com população/contexto/técnicas semelhantes forneceram insuficientes evidências acerca dos efeitos atingidos pela intervenção, enquanto critério para se seguir com teste de eficácia/efetividade (Bowen et al., 2009). Para maior clareza na compreensão de critérios, até então definidos, Bowen et al., (2009) salientam áreas fundamentais a serem verificadas em estudos de viabilidade, a exemplo:

1. Aceitabilidade/Satisfação: Identificar como o público-alvo e equipe técnica reagem antes, durante e após intervenção. Verificar a aceitabilidade relativa dos meios de divulgação e seleção da amostra e intenção dos participantes em continuar com outras propostas em saúde, após a intervenção;

2. Demanda: Avaliar o quanto a intervenção foi procurada por parte do público-alvo. Há a necessidade desse tipo de serviço? O quanto é provável que esse serviço seja usado?;

3. Implementação: Avaliar a extensão, modo e probabilidade real da implementação da intervenção em contexto não controlado. O que precisa ser ajustado? Avaliar sucessos e fracassos no processo de implementação do programa, quantidade de recursos necessários (humanos, materiais e físicos) e fatores dificultadores para a implementação do programa;

4. Adaptação: Observar modificações necessárias nos conteúdos, estrutura, processos de implementação e avaliação do programa, compatível ao público-alvo;

5. Integração: Considerar as mudanças necessárias para integrar o programa em um local de trabalho/organização. É viável introduzir esse tipo de serviço na cultura organizacional? Verificar adaptações, custos e efeitos à cultura organizacional em manter o programa;

6. Praticidade: Investigar possibilidades de recursos (humanos e materiais, tempo, instalações, etc.) escassos

e/ou acesso limitado (ao público-alvo, local de serviço, etc.), rapidez, qualidade de implementação, custos, habilidades dos participantes de conduzirem as atividades;

7. Expansão: Verificar se é possível expandir a proposta do programa para outras populações. O quanto o programa pode ser expandido para gerar novos serviços?;

8. Testes limitados de eficácia: Verificar aspectos preliminares de critérios de eficácia, como efeito em variáveis importantes, tamanho do efeito, manutenção da mudança (estudos de seguimento *follow-up*). Testes conduzidos em amostra por conveniência.

Considerações Finais

A avaliação de programas configura-se como área indispensável para que avanços sejam alcançados em saúde, tanto em termos de prevenção de doenças quanto em abordagens mais amplas de promoção de saúde. No entanto, essa avaliação é ainda desafiadora devido à vasta gama de critérios já definidos em estudos internacionais para avaliação de resultados de eficácia e efetividade dos programas. É necessário divulgar e aprofundar esses conceitos, bem como conhecimentos sobre critérios já introduzidos para avaliação de resultados de programas de intervenção em Psicologia.

Porém, o nível de rigor de sistematização de critérios para o delineamento e avaliação de resultados, não parece seguir a mesma lógica quando se trata de estudos de viabilidade e estudos piloto. Isso, pois, há ainda uma grande inconsistência entre autores/pesquisadores e respectivos critérios sugeridos, quanto à avaliação de estudos de viabilidade e piloto. Nessa mesma perspectiva, pode-se pensar que, ao traçar métodos criteriosos para atingir dado resultado, no caso de ensaios de eficácia e/ou efetividade, uma intervenção deveria, supostamente, ter cumprido com rigor prévio, em um modelo cautelosamente planejado, em estudo de viabilidade, seguindo para, mas não necessariamente separado, de um estudo piloto. Após atingir os mínimos desejados, em termos de resultados satisfatórios para seguir com ensaios de eficácia, o pesquisador, calçado em suas evidências empíricas obtidas em estudos de viabilidade e/ou piloto, teria maior precisão de resultados futuros, evitando gastos desnecessários, esgotamento de participantes submetidos às intervenções e favorecendo, assim, o alcance de metas em saúde.

No entanto, essa realidade ainda parece distante, quando se percebe que a maioria dos programas ofertados em saúde em contexto nacional ainda carece de rigor empírico, seja na avaliação de eficácia ou efetividade, quanto mais no que diz respeito à verificação prévia, em estudos de viabilidade/piloto. Por isso, maior conhecimento por parte dos pesquisadores/gestores em saúde, com relação aos critérios já existentes para avaliar suas práticas profissionais, é fundamental para disseminar melhores estratégias e evitar desgastes em todos os níveis nos serviços (humanos, materiais, de tempo, etc.).

Enquanto profissionais não conduzirem, com prudência suficiente, a investigação criteriosa de suas práticas, a avaliação e relato transparente de seus resultados, com base na cientificidade do melhor método disponível para aquele momento e contexto, resultados obtidos, ou alegações de supostos resultados de intervenções, continuam à mercê de serem tidos como mera especulação. Isso

impede que avanços sejam alcançados dentro do modelo científico vigente. Portanto, pesquisadores/gestores em saúde são encorajados a seguirem suas práticas, com base em evidências e em prol de gerar novas evidências (ferramentas, técnicas, tecnologias, programas, etc.) que possam somar ganhos em termos de respostas a problemas sociais e subtrair custos para a saúde e sociedade como um todo.

Referências

- American Psychological Association, APA (2002). Criteria for evaluating treatment guidelines. *American Psychologist*, 57(12), 1052-1059. doi: 10.1037/0003-066X.57.12.1052
- American Psychological Association, APA (2006). Evidence-based practice in psychology: APA presidential task force on evidence-based practice. *American Psychologist*, 61(4), 271-285. doi:10.1037/0003-066X.61.4.271
- Baker, T. B., & McFall, R. M. (2014). The promise of science-based training and application in psychological clinical science. *Psychotherapy*, 51(4), 482-486. doi:10.1037/a0036563
- Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., ... Fernandez, M. (2009). How we design feasibility studies. *American Journal of Preventive Medicine*, 36(5), 452-457. doi: 10.1016/j.amepre.2009.02.002
- Fernandez-Ballesteros, R. (2001). *Introducción a la evaluación psicológica*. Madrid: Piramides.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., ... Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151-175. doi: 10.1007/s11121-005-5553-y
- Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: next generation. *Prevention Science*, 16(7), 893-926. doi: 10.1007/s11121-015-0555-x
- Health Research Council of New Zealand, HRC (2017). *Feasibility studies in population health or clinical research provide support to test feasibility issues affecting a planned, larger study*. Recuperado de <http://www.hrc.govt.nz/funding-opportunities/researcher-initiated-proposals/feasibility-studies>
- Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]*. The Cochrane Collaboration, 2011. Recuperado de www.handbook.cochrane.org
- Hollon, S. D., Stewart, M. O., & Strunk, D. (2006). Enduring effects for cognitive behavior therapy in the treatment of depression and anxiety. *Annual Review of Psychology*, 57(1), 285-315. doi: 10.1146/annurev.psych.57.102904.190044
- Johnson, M., Weinman, J., & Chater, A. (2011). A healthy contribution. *The Psychologist*, 24(12), 890-892. Recuperado de <https://thepsychologist.bps.org.uk/getfile/1061>
- National Institute for Health Research, NIHR (2017). *Feasibility and pilot studies: Which programme should I apply to?* Recuperado de <http://www.nets.nihr.ac.uk/glossary>
- Norte, C. E., Souza, G. G. L., Pedrozo, A. L., Mendonça-de-Souza, A. C. F., Figueira, I., Volchan, E., & Ventura, P. R. (2011). Impacto da terapia cognitivo-comportamental nos fatores neurobiológicos relacionados à resiliência. *Revista de Psiquiatria Clínica*, 38(1), 43-45. doi: 10.1590/S0101-60832011000100009
- Orsmond, G. I., & Cohn, E. S. (2015). The distinctive features of a feasibility study. *OTJR: Occupation, Participation and Health*, 35(3), 169-177. doi: 10.1177/1539449215578649
- Sparks, M. (2013). The changing contexts of health promotion. *Health Promotion International*, 28(2), 153-156. doi: 10.1093/heapro/dat034
- Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2012). *Research methods in psychology*. 9th Ed. New York: McGraw Hill.
- Society for Prevention Research, SPR (2004). *Standards of evidence: criteria for efficacy, effectiveness and dissemination*. Recuperado de <http://www.preventionresearch.org/StandardsofEvidencebook.pdf>
- Straub, R. (2012). *Health psychology. A biopsychological approach*. 3rd Ed. New York: Worth Publishers.
- Thabane, L., Ma, J., Chu, R., Cheng, J., Ismail, A., Rios, L. P., ... Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. *BMC Medical Research Methodology*, 10(1), 1-10. doi: 10.1186/1471-2288-10-1
- Tickle-Degnen, L. (2013). Nuts and bolts of conducting feasibility studies. *American Journal of Occupational Therapy*, 67(2), 171-176. doi: 10.5014/ajot.2013.006270
- United Nations Development Programme – UNDP (2009). *Handbook on planning, monitoring and evaluating for development results*. New York: A. K. Office Supplies.
- World Health Organization –WHO (2010). *Global status report on noncommunicable diseases*. Geneva: Library Cataloguing-in-Publication Data.

recebido em junho de 2017
aprovado em dezembro de 2017

Sobre as autoras

Helen Bedinoto Durgante é Psicóloga e Mestre em Desenvolvimento e Direitos pela Goldsmiths University of London, Doutoranda em Psicologia pelo PPG Psicologia UFRGS, bolsista CNPq.

Débora Dalbosco Dell'Aglio é Psicóloga, Mestre e Doutora em Psicologia do Desenvolvimento (UFRGS), Professora Pesquisadora PPG em Educação UniLasalle/Canoas e Professora Colaboradora PPG Psicologia UFRGS, RS. Pesquisadora 1A CNPq.