

CAPÍTULO 4 – POR QUE DEVERÍAMOS NOS PREOCUPAR COM PRÁTICAS DE PROMOÇÃO DA SAÚDE BASEADA EM EVIDÊNCIAS?

LOUISE POTVIN¹

Muito tem sido feito no campo da promoção da saúde ao longo da última década, com esforços para se construir um corpo de práticas baseadas em evidências para uso no planejamento e implementação de iniciativas e programas voltados à promoção da saúde. Naturalmente, isso acontece na esteira do formidável ímpeto no sentido de se desenvolver uma prática baseada em evidências na área da medicina clínica, particularmente em áreas altamente especializadas. Infelizmente, muitos grupos que atuam na promoção da saúde aderiram ao movimento sem questionar suas raízes e pressupostos. Acreditava-se que, com a síntese dos estudos de avaliação das intervenções, estaríamos bem posicionados para oferecer orientações práticas, extraídas de forma racional, aos profissionais atuantes na área. Penso que nós, como comunidade atuante na avaliação da promoção da saúde, nos associamos a essa corrente de forma muito apressada e que devemos recuar um pouco a fim de examinar criticamente os pressupostos subjacentes às sínteses de pesquisa e também se esses pressupostos podem ser colocados quando se trata de informar as práticas de promoção da saúde.

Prática baseada em evidências: uma ferramenta simplificadora

O pressuposto mais importante na prática baseada em evidências é o de que as decisões práticas, exigidas em situações únicas para um dado problema, são mais bem informadas pela sistematização dos resultados de iniciativas ou programas avaliados e que foram conduzidos para lidar com aquele mesmo problema em outras situações. Muito embora esse pressuposto possa parecer razoável à primeira vista, ele também pressupõe uma série de outros corolários. **O primeiro deles é o de que a situação particular de interesse para a qual se busca uma decisão faça parte do mesmo universo de situações daquelas das quais nossas evidências foram retiradas.** Esse princípio foi amplamente discutido por Lee J. Cronbach *et al.*¹ em sua teoria da generalizabilidade. De fato, o que fica claro no trabalho de Cronbach é que não há uma realidade ontológica que defina qualquer universo de objetos. Os critérios para julgar se um objeto qualquer faz parte de um universo particular são todos derivados empiricamente, em sua maioria, por meio da aplicação do princípio da similaridade. Objetos com propriedades similares, acredita-se, pertencem ao mesmo universo. Essa asserção aparentemente simples apresenta três problemas para o propósito prático de decidir se um objeto pertence ou não a um universo específico. O primeiro deles diz respeito à seleção das propriedades com base nas quais as semelhanças devem ser avaliadas. Em seguida, emergem questões relacionadas à mensuração das diferenças naquelas propriedades entre os objetos que possivelmente serão classificados como pertencentes ao mesmo universo. Finalmente, é preciso que uma decisão seja tomada com relação a um ponto de corte na escala, além do qual as diferenças sejam clas-

1. Centre de Recherche Léa-Roback sur les Inégalités Sociales de Santé. Université de Montréal. Montréal, QC

sificadas como muito grandes para que aqueles objetos pertençam ao mesmo universo. Assim, definir o universo de situações problemáticas que sejam relevantes para se levantar evidências a fim de informar decisões em um caso específico não é uma questão puramente técnica. Ela envolve uma série de julgamentos normativos que raramente são documentados de forma adequada e levados em conta na interpretação da evidência.

Na essência, o que está em jogo na definição dos universos relevantes para se derivar evidências é a questão das variações entre os objetos e as operações relacionadas às decisões sobre as composições daqueles universos, o que acaba por se resumir a um problema de classificação. Uma das funções cognitivas mais importantes do cérebro humano é a de ser capaz de criar categorias; classes de objetos com as quais um indivíduo pode se relacionar da mesma maneira. Essa é uma função essencial porque ela nos permite selecionar comportamentos apropriados sem ter que, constantemente, produzir análises completas para cada objeto que compõe nosso ambiente. Encontros com objetos que não podem ser classificados dentro de uma categoria já existente colocam problemas que somente podem ser resolvidos por meio do processo de aprendizagem, que consiste em criar uma classe totalmente nova de objetos ou em modificar as regras de classificação que se mostraram inadequadas para lidar com esse novo objeto.

Uma experiência iluminadora relacionada a esse fato foi quando eu instalei um computador pela primeira vez na casa dos meus pais idosos. A maioria de nós interage com essa classe de objetos há tanto tempo que tendemos a esquecer os estranhos comportamentos exigidos das pessoas que desejam interagir com esses equipamentos. Sempre que vemos algo que se parece um *mouse* de computador, com dois ou três botões que podem ser pressionados e que estão associados a um mecanismo que mexe um cursor em uma tela, nós sabemos que um clique com o botão direito irá acionar um *menu* com opções de escolha que aparecerá na tela e que um clique duplo com o botão esquerdo produzirá uma operação definida pela posição do cursor na tela. Nós desenvolvemos essas respostas automáticas após milhares e milhares de sessões de trabalho com computador e podemos generalizar esses comportamentos seja qual for o formato do *mouse*. O que me impressionou quando eu estava ensinando meus pais de 70 anos a usar o computador novo deles foi que eles nem mesmo conseguiam compreender as funções que o *mouse* desempenharia para eles e como poderiam realizar essas funções. Ainda mais perturbador foi o fato de que até mesmo um *mouse* de computador é um objeto tão complexo que descrever o que ele faz para alguém que nunca viu um é algo quase impossível. Então decidi que eles deveriam aprender o que era um *mouse* usando-o. Foi quando percebi que não há um gene para o clique duplo. Esse comportamento simples que hoje em dia nós temos como natural e que mesmo crianças de dois anos de idade faz sem pensar precisa ser aprendido por meio da prática e da repetição. Especialmente nos dias de hoje, em que um *mouse* pode ter várias formas e formatos, reconhecer que um objeto específico pode, de fato, realizar as funções de um *mouse* de computador requer uma grande imaginação e experimentação com o objeto real.

Assim, uma classe de objetos funciona como uma grande caixa dentro da qual nós armazenamos os objetos conforme nos deparamos com eles. Cada caixa é associada a um conjunto particular de ações e comportamentos. Colocar um objeto específico em uma caixa específica oferece algum nível de segurança de que os comportamentos associados àquela caixa, provavelmente, são

os mais apropriados para qualquer objeto que saia dela. Essa é uma ferramenta simplificadora, que oculta as variações entre os vários objetos que são reagrupados em uma classe. Conseqüentemente, quanto mais homogêneos forem os objetos que compõe uma classe, mais provável será a adequação e a relevância das ações e comportamentos associados àquela classe para qualquer um dos objetos que compõe a classe. Essa é exatamente a ideia do intervalo de confiança. Quanto maior as variações entre os objetos que formam uma classe, maior o intervalo de confiança em torno de qualquer estimativa de propriedade ou ação associada com qualquer objeto individualmente daquela classe. Dessa forma, para que essa ferramenta simplificadora da prática baseada em evidências seja relevante para oferecer orientação para ações específicas em uma dada situação problemática, duas condições precisam ser satisfeitas: é preciso primeiro ser capaz de avaliar a que universo de situações problemáticas a situação em questão pertence; e, em segundo lugar, deve existir um claro curso de ação associada àquele universo de situações problemáticas. Duas ferramentas metodológicas desenvolvidas no fim dos anos 1970 são úteis para oferecer *insights* acerca dessas duas condições.

Duas ferramentas metodológicas necessárias à prática baseada em evidências

O movimento da **prática baseada em evidências no campo da clínica médica** tornou-se possível por conta dos grandes avanços **metodológicos que ocorreram no fim da década de 1970**, relacionados às sínteses de resultados de pesquisas. Até então, não havia um procedimento consensual para resumir os resultados de uma série de estudos. O número de estudos empíricos, então, começou a crescer exponencialmente, em particular sobre questões aplicadas que não eram necessariamente ligadas a uma escola de pensamento teórico forte. De fato, quando um estudo se relaciona com teorias sofisticadas e bem desenvolvidas, essa teoria em si oferece um arcabouço para a síntese dos dados de pesquisa. No entanto, quando essas teorias não existem, outras ferramentas precisam ser criadas. No caso das perguntas aplicadas, nas quais as observações empíricas exercem um papel, no mínimo, igual às explicações teóricas ao derivar conhecimento, havia uma necessidade de oferecer uma ferramenta metodológica que cumprisse a função de formatar a integração de vários e dispersos resultados de pesquisa em um número de estimativas mais administráveis para um tamanho-efeito. As ferramentas metodológicas da meta-avaliação e da meta-análise foram desenvolvidas para fazer justamente isso. Apesar de semelhantes, esses dois campos lidam com a questão da síntese de resultados de pesquisa a partir de ângulos um pouco diferentes.

A meta-análise é normalmente compreendida como o verdadeiro procedimento técnico de resumo de resultados a partir de uma variedade de estudos diferentes que testa a mesma hipótese para uma estimativa única.² Idealmente, estudos adequados para uma meta-avaliação usariam as mesmas medidas, tanto para variáveis dependentes quanto independentes. Eles deveriam controlar as variáveis de confusão. Finalmente, esses estudos deveriam usar critérios de inclusão e de exclusão similares para definição das amostras do estudo. No caso muito raro em que todas essas condições são satisfeitas, e quando os dados originais dos estudos estão disponíveis para uma análise secundária, seria possível agrupar esses dados em uma amostra

estratificada maior e computar uma estimativa tamanho-efeito com duas propriedades principais: a) a estimativa sintética seria uma média ponderada dos tamanhos efeitos calculados na meta-análise; e b) o intervalo de confiança da estimativa sintética seria menor do que o do tamanho efeito em cada estudo. Devido ao fato de existir grandes variações entre as pesquisas realizadas para estudar um mesmo fenômeno, o principal avanço da meta-análise foi criar procedimentos de estimativa estatística que acomodassem desvios do caso ideal. Embora os avanços na meta-análise tendam, geralmente, a ter o foco voltado para os procedimentos mais técnicos de estimativa estatística, muitas questões normativas que dizem respeito às variações nas técnicas de mensuração também foram discutidas e os seus efeitos avaliados em estudos simulados ou reais de meta-análise.

Diferente da meta-análise, o campo da meta-avaliação é caracterizado, logo de início, como um esforço normativo que, basicamente, busca definir padrões e normas com relação as quais a qualidade de qualquer avaliação individual possa ser comparada e testada em termos de qualidade.³ Outra diferença entre meta-análise e meta-avaliação é o fato de que na última a variável dependente é a exposição a um programa ou a uma intervenção concebida para lidar com uma situação problemática identificável. Portanto, no caso da meta-avaliação, três dimensões ou tipos diferentes de propriedades devem ser avaliados subsequentemente a fim de se poder encontrar o universo adequado para qualquer estudo de avaliação em questão. A primeira é a situação problemática em si. Para deixar as coisas ainda mais complicadas, pode-se argumentar que há muitas formas diferentes de identificar uma situação problemática e que o meio pelo qual a situação foi identificada como problemática deve ser, pelo menos, documentado e, talvez, levado em consideração. A segunda dimensão é a intervenção e/ou programa concebido para lidar com a situação problemática. De fato, idealmente, para uma situação problemática dada, as séries de estudos de meta-avaliação conduzidos em diferentes intervenções deveriam constituir o requisito mínimo da prática baseada em evidências. O terceiro tipo de propriedade é o resultado de interesse. É bem sabido que as intervenções e programas possuem uma gama de efeitos esperados que podem ser avaliados, e que isso nem leva em consideração aqueles efeitos não esperados que raramente são examinados.

Devido a maior complexidade das situações aplicadas com que os estudos incluídos nas meta-avaliações lidam, comparadas àquelas incluídas nas meta-análises, as dimensões normativas das meta-avaliações são muito mais desenvolvidas. De fato, muitas vezes, o trabalho do meta-avaliador é ponderar as múltiplas opções de intervenção para uma situação problemática específica. Ademais, devido à natureza aplicada das situações, existe uma grande variação nos formatos de estudo em estudos de avaliação, e essas variações, muitas vezes, são restringidas e relacionadas às várias opções de intervenção. Uma das ferramentas normativas mais usadas desenvolvidas pela meta-avaliação é a classificação dos formatos de avaliação em termos da confiabilidade dos resultados no que tange a relações causais que ligam programas e resultados. As normas e padrões definidos pelos órgãos de acreditação de avaliação são, de fato, bastante úteis na preparação do projeto de avaliação e do relatório de avaliação, no sentido de que eles oferecem orientações aos avaliadores

para ajudá-los a produzir avaliações aceitáveis pela comunidade de avaliadores. Assim, a meta-avaliação nos ensinou que fazer síntese de estudos de avaliação, envolve um grande quantidade de decisões arbitrárias acerca do que constitui resultados comparáveis a serem sintetizados e também sobre como ponderar aquelas decisões em termos da validade ou confiabilidade das conclusões alcançadas nas meta-avaliações.

Reunindo tudo: por que deveríamos nos preocupar com promoção da saúde baseada em evidências?

As discussões acerca da meta-avaliação são úteis para se examinar as dificuldades enfrentadas no campo da promoção da saúde para satisfazer à primeira condição da prática baseada em evidências, qual seja, a capacidade de encontrar um universo de situações adequado para categorizar uma situação em análise. Para as práticas de promoção da saúde baseadas em evidências, o universo de referência é composto de estudos de avaliação que relacionam uma situação problemática a uma intervenção documentada e, assim, a resultados específicos. Essa é a cadeia de eventos que queremos ser capaz de antecipar antes de escolher um curso de ação em uma dada situação. Eu já me referi aqui ao fato de que ao reduzir-se a variância entre os objetos que compõe o universo de referência, naquele caso os estudos de avaliação específicos, aumenta-se o grau de adequação da prática baseada em evidências. Há pelo menos duas maneiras de reduzir essa variância. A primeira opção é dividir o grupo original de objetos em categorias mais homogêneas, baseadas em dimensões relevantes. A fim de que isso seja feito, no entanto, é preciso, de início, acessar um conjunto muito grande de objetos. A segunda opção é reduzir a complexidade dos objetos que perfazem o universo de interesse, porque a complexidade é um dos maiores geradores de variação. No caso da avaliação da promoção da saúde, nenhuma dessas opções existe.

Em primeiro lugar, o conjunto total de estudos de avaliação é muito pequeno comparado aos mais de cem anos de pesquisa que abastece a prática baseada em evidências extraída da medicina experimental. Como uma prática nova na saúde pública, a promoção da saúde ainda não teve tempo para construir uma base de pesquisa que permitisse dividir os estudos de avaliação em categorias, dentro das quais o número de estudos seja suficiente para oferecer estimativas precisas quando submetidos a uma meta-avaliação. Nós temos a tendência de agregar estudos que são muito diversos e de extrair estimativas de efeito a partir de um número muito pequeno de estudos. Em segundo lugar, e para piorar ainda mais as coisas, a complexidade das intervenções no campo da promoção da saúde é bem maior do que a da maioria dos outros tipos de intervenções na área da saúde. Normalmente, as intervenções de promoção da saúde envolvem uma grande diversidade de atores, que abordam a ação com base nas suas próprias histórias e interesses. Além disso, essas intervenções acontecem em sistemas abertos, o que significa que elas são transformadas por meio das interações com as condições e o contexto da implantação. Finalmente, somando-se a essa complexidade, essa evolução normalmente acontece ao longo de um extenso período de tempo durante o qual o contexto muda e novos atores se envolvem. A diminuição da complexidade das avaliações

das intervenções no campo da promoção da saúde, a fim de criar categorias mais homogêneas, não pode ser feita sem se alterar a natureza da intervenção.

Isso não quer dizer que pesquisas de avaliação são inúteis para promoção da saúde. Bem ao contrário, para que, em um momento, seja possível ter uma massa crítica de estudos de avaliação que permita extrair estimativas válidas, é preciso intensificar nossos esforços no sentido de avaliar as práticas de promoção da saúde. É preciso estar alerta, no entanto, para o fato de que usar o conjunto de estudos existentes para extrair estimativas de efeitos e, em seguida, extrapolar essa estimativa para decisões em situações específicas é algo, no mínimo, perigoso e pode, muitas vezes, levar a erros. De fato, uma das maiores contribuições de Cronbach ao campo da pesquisa aplicada é o alerta feito por ele no sentido de que ignorar as várias interações de ordem mais alta que acontecem na produção de qualquer fato científico extraído a partir da observação (ou mesmo da experimentação controlada), muitas vezes, resulta em conclusões enganadoras.⁴ Dessa forma, como é possível informar as decisões e ações em uma dada situação por meio do uso do conhecimento derivado de outra experiência similar?

A minha impressão é que, em nossa pressa para oferecer orientações aos profissionais que atuam na intervenção no campo da promoção de saúde, estamos fazendo pouquíssimo uso de outra ferramenta usada para formatar e extrair sentido dos dados de pesquisa: a teoria. Nós temos a tendência a subestimar e invalidar o conhecimento extraído de um processo dedutivo aplicado ao conhecimento teórico e a superestimar o conhecimento que advem da acumulação das observações empíricas, mesmo que a base empírica não seja suficiente. Ao fazermos isso, temos a impressão enganosa de sermos mais “científicos” e racionais, dado que o espaço para interpretação parece ser menor quando se extrapolar a partir de evidências derivadas empiricamente do que quando se deduzem cursos de ação para uma situação específica a partir de uma teoria mais geral sobre o funcionamento dos mecanismos em ação. É nesse ponto, em minha opinião, que deveríamos concentrar nossos esforços na avaliação.

Em qualquer situação específica, particularmente nos sistemas abertos das intervenções comunitárias, há uma multitudine de mecanismos sociais, biológicos e psicológicos, ativados ou não pela intervenção, que interagem de maneira específica para produzir os resultados observados. Somente nos sistemas fechados dos laboratórios é que se pode, efetivamente, controlar outros mecanismos potenciais a fim de isolar um mecanismo de interesse. Em um sistema aberto, pode-se tão-somente estimar esses efeitos da interação. Isso pode ser feito de forma indutiva, agregando-se observações empíricas e calculando estimativas sintéticas sempre que o número de observações for suficiente, ou fazendo uso do conhecimento teórico sobre os mecanismos em ação e, dedutivamente, desbaratando suas interações⁵ de acordo com as características da situação à qual o conhecimento precisa ser aplicado. Para fazer isso, no entanto, é preciso ter confiança na praticidade do conhecimento teórico, algo que foi perdido na saúde pública e na promoção da saúde.

Referências Bibliográficas

1. Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. The dependability of behavioural measurements: theory of generalizability of score and profiles. Nova Iorque: Wiley; 1972.
2. Glass GV, McGraw B, Smith ML. Meta-analysis in social research. Beverly Hills, CA: Sage; 1981.
3. Stufflebeam D. Meta evaluation: an overview. *Eval Health Prof* 1997;1:17-43.
4. Cronbach LJ. Beyond the two disciplines of scientific psychology. *Am Psychol* 1975;30:671-84.
5. Pawson R, Tilley N. Realistic evaluation. Londres: Sage; 1997.