

Why we should be worried about evidence-based practice in health promotion

Louise Potvin ¹

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

Porque deveríamos estar preocupados com práticas de promoção da saúde baseadas em evidências

Abstract

The author analyzes the health promotion considering the use of evidence-based practice. Sometimes problems could arise from this approach. Indeed decisions concerning a specific case, are not only a technical issue, but involve normative judgments rarely well documented.

Key words *Evidence-based medicine, Health promotion*

Resumo

O autor discute a promoção da saúde considerando o uso da prática baseada em evidências. Algumas vezes podem surgir problemas com essa abordagem. De fato, decisões sobre casos específicos não se baseiam exclusivamente em aspectos técnicos, mas envolvem julgamentos normativos nem sempre adequadamente documentados.

Palavras-chave *Medicina baseada em evidências, Promoção da saúde,*

For about a decade, there has been a great deal of activity in the field of health promotion in various attempts to build bodies of evidence-based practices for planning and implementing initiatives and programs aimed at promoting population health. This follows of course the overwhelming impetus to develop evidence-based practice in the field of clinical medicine especially in highly specialised areas. Unfortunately, many groups involved in health promotion have entered the ban wagon of the evidence-based movement without questioning its roots and assumptions. It was believed that by synthesising studies evaluating interventions we would be in a position to provide rationally derived practice guidance for practitioners in the field. I think that we, as a community of practice in health promotion evaluation, have entered this evidence-based movement too hastily and that we should backtrack a little in order to critically examine the assumptions underlying the research syntheses and whether or not these assumptions can be met when it comes to provide guidance for health promotion practices.

Evidence-based practice: a simplifying device

The single most important assumption in evidence-based practice is that practical decisions required in singular situations for acting on a given problem are best informed by synthesising the results of evaluated initiatives or programs undertaken to address that problem in other situations. While this assumption may seem reasonable at first sight, it also requires a series of corollary assumptions. The first of those assumptions is that the singular situation of interest for which a decision is sought, is part of the same universe of situations than those from which we drew evidence. This principle was thoroughly discussed by Lee J. Cronbach *et al.*¹ in his generalizability theory. Indeed, what is clear from Cronbach's work is that there is no ontological reality that defines any universe of objects. The criteria for judging whether or not any single object belongs to a particular universe are all empirically derived mostly through the application of a principle of similarities. Objects with similar properties, it is believed, belong to the same universe. This apparently very simple assertion poses three problems for the practical purpose of deciding whether or not an object belongs to a particular universe. The first pertains to selecting the properties on which similarities are to be assessed. Subsequently come issues

related to measuring differences on those properties among objects that can potentially be classified as belonging to the same universe. Finally, a decision has to be made regarding a cut point in the scale beyond which differences are deemed too large for those objects to belong to the same universe. So, defining the universe of problematic situations relevant for deriving evidence to inform decisions for a specific case is not purely a technical issue. It involves a series of normative judgements that are rarely adequately documented and taken into account in the interpretation of evidence.

Essentially what is at stake in the definition of the relevant universes for deriving evidence is the question of variations between objects and the operations involved in decisions about the compositions of those universes, which boils down to classification problems. One of the most important cognitive function in the human brain is to be able to create categories; classes of objects with which one can relate in the same manner. This is an essential function because it provides guidance to select appropriate behaviours without having to constantly produce thorough analyses for each object that composes our environment. Encounters with objects that cannot be classified into an existing category pose problems that can only be solved through a learning process, which consists in either creating a whole new class of objects or modifying the classification rules that are proved inadequate in order to deal with this new object.

A recent experience enlightening about this fact was when I installed a computer for the first time in my aging parent's home. Most of us have been interacting with this class of objects for so long that we tend to forget the awkwardness of the specific behaviours required from humans who want to interact with them. Whenever we see something that looks like a computer mouse with two or three levers that can be pressed and that are associated with the a mechanism that moves a cursor on a screen, we know that one click on the right lever will trigger a roll down menu to appear on the screen and that a double click on the left lever will produce an operation as defined by the location of the cursor on the screen. We have developed those automatic responses through thousands and thousands of working sessions with computers and we can generalise those behaviours whatever the shape of a particular mouse. What struck me when I was instructing my 70 year-old parents about their new computer was that they were not even able to figure out the functions that the mouse would be performing for them and how to operate those func-

tions. Even more disturbing was the fact that even a computer mouse is such a complex object that describing what it does for someone who has never seen it is close to impossible. So I decided that they should learn about the mouse by using it. This is when I realised that there is no gene for the double click. This simple behaviour that we now take for granted and that even two year old kids perform without thinking has to be learned through practice and repetition. Even more so, nowadays with the multiplicity of shapes and forms that computer mice take, recognising that a specific object does indeed perform computer mouse functions sometimes takes a great deal of imagination and experimentation with the actual object.

So, a class of objects acts as a big box into which we store objects as we encounter them. Each box is associated with a particular set of actions and behaviours. Putting a specific object in a specific box allows some reassurance that the behaviours associated with that box is the most likely appropriate behaviour for whatever object comes out of that box. It is a simplifying device that obliterates variations between the various objects that are regrouped within a class. Consequently, the more homogeneous the objects composing a class, the more likely are the appropriateness and relevance of the actions and behaviours associated with that class for anyone of the object composing the class. This is the whole idea of confidence interval. The greater the variations across objects forming a class, the wider the confidence interval around any estimation of property or action associated with any one object from that class. So, in order for the simplifying device of evidence based practice to be relevant for providing guidance for specific actions in a given problematic situation two conditions need to be met: one should first be able to assess to which universe of problematic situations the situation at hands belongs to and second, there should exist a clear course of actions associated with that universe of problematic situations. Two methodological tools developed at the end of the 1970's are helpful to provide insights concerning those two conditions.

Two methodological tools necessary for evidence-based practice

The evidence-based practice movement in clinical fields was made possible by the tremendous methodological developments that occurred at the end of the 1970's regarding the synthesis of research findings. Up to then there was no generally agreed upon

procedure for summarising results from a variety of studies. The number of empirical studies then started to grow exponentially, in particular about applied issues that were not necessarily linked to a strong theoretical school of thought. Indeed, when studies are linked to sophisticated and well developed theories, the theory itself provides the frame for synthesising research data but when such theories are lacking other devices have to be created. In the case of applied questions for which empirical observations play a role that is at least equal as theoretical explanations in deriving knowledge, there was a need to provide a methodological tool that would fulfil the function of framing the integration of various and disperse research results into a more manageable number of estimates for an effect size. The methodological tools of meta-evaluation and meta-analyses were developed for doing just that. Although similar, those two fields address the issue of synthesizing research results in slightly different perspectives.

Meta-analysis is usually understood as the very technical procedure of summarising results from a number of different studies testing the same hypothesis into a single estimate.² Ideally, studies suitable for meta-evaluation would have used the same measures for both dependent and independent variables. They should have controlled for the same confounding variables. Finally, they should have used similar inclusion and exclusion criteria for defining study samples. In the very rare case when all these conditions are met, and when the original studies' individual data are available for secondary analysis, one could pool these data into a stratified larger sample and compute a synthetic effect-size estimate with two main properties: a) the synthetic estimate would be a weighted average of the effect sizes calculated in each study included in the meta analysis; and b) the confidence interval of the synthetic estimate would be smaller than that of the effect size in each study. Because there exist wide variations between research undertaken to study the same phenomenon, the main developmental task in meta analysis was to create statistical estimation procedures that would accommodate deviations from this ideal case. Although developments in meta-analysis tend to be generally focussing on the more technical statistical estimation procedures, a number of normative issues mainly pertaining to variations in measurement techniques have also been discussed and their effect assessed in simulated or real meta-analysis studies.

Contrary to that of meta-analysis, the field of meta-evaluation is characterised upfront as a norma-

tive enterprise that ultimately seeks to define standards and norms against which the quality of any single evaluation study can be compared and assessed for quality.³ Another distinction between meta-analysis and meta-evaluation is the fact that in the latter, the dependent variable is the exposure to a program or an intervention designed to address an identifiable problematic situation. Therefore, in the case of meta evaluation, three different dimensions or types of properties should be sequentially assessed in order to be able to find the proper universe for any single evaluation study. The first is the problematic situation itself. It should be the same for all evaluation studies in a meta evaluation. To make things even more complicated, one could also argue that there are many different ways of identifying a problematic situation and that the mean by which the situation was identified as problematic should be at least documented and maybe taken into account. The second dimension is the intervention and/or program designed to address the problematic situation. Indeed, ideally for a given problematic situation, series of meta-evaluation studies performed on various interventions should constitute the minimum requirement for evidence-based practice. The third type of properties is the outcome of interest. It is well known that interventions and programs have numerous intended effects that can be evaluated and that does not even account for those unintended effects that are only rarely examined.

Because of the greater complexity of the applied situations that studies included in meta-evaluations are dealing with compared to those included in meta-analyses, the normative dimensions of meta-evaluations are much more developed. Indeed, often the task for meta evaluators is to weight various intervention options for a single problematic situation. In addition, because of the applied nature of the situations, there exists a huge variation in the study designs implemented in evaluation studies and these variations are often constrained by, and related to, the various intervention options. One of the most widely used normative tool developed by meta evaluation is the ranking of evaluation designs in terms of the trustfulness of results about causal relationships that link programs and outcomes. Norms and standards agreed upon by evaluation accreditation bodies are indeed quite useful in the preparation of evaluation design and evaluation report in the sense that they provide guidelines for evaluators to help them produced evaluations acceptable by the community of evaluators. So, meta evaluation has taught us that synthesising evaluation studies

involves a great number of arbitrary decisions regarding what constitutes comparable results to be synthesised and how to weight those decisions in terms of the validity or trustworthiness of the conclusions reached in meta evaluations.

Putting it together: why we should worry about evidence-based health promotion

Discussions regarding meta evaluation are helpful for examining the difficulties for health promotion to meet the first condition for evidence-based practice that is the capacity to find an appropriate universe of situations in which to categorise the situation at hand. For evidence-based health promotion practices, the universe of reference is made up of evaluation studies that link a problematic situation to a documented intervention and then to specific outcomes. This is the chain of events that we want to be able to anticipate before choosing a course of action in a given situation. I have already made the point that reducing the variance between the objects that make up the universe of reference, in that case specific evaluation studies, increases the appropriateness of evidence-based practice. There are essentially two ways of reducing that variance. The first option is to divide up the original pool of objects into more homogeneous categories based on relevant dimensions. In order to do this however one needs to have access to a very large pool of objects to start with. The second option is to reduce the complexity of the objects that made up the universe of interest because complexity is one of the main generators of variation. In the case of health promotion evaluation, none of these options is available.

First, the total pool of evaluation studies in evaluation is very small compared to the hundred more years of research that feed into evidence-based practice derived from experimental medicine. As a new practice in public health, health promotion has not yet had the time to build the research base that would allow dividing up evaluation studies into categories in which the number of studies is sufficient to lend to precise estimates when subjected to meta evaluation. We have the tendency to aggregate studies that are very diverse and to derive estimates of effects from a very small number of studies. Second, and to make matters even worse, the complexity of health promotion interventions is much greater than that in most of the other types of health interventions. Typically, health promotion interventions involve a great diversity of actors who approach the intervention with their own history and interest. In addition those interventions take place in open systems,

meaning that they are transformed through interactions with the implantation conditions and context. Finally, adding to the complexity, this evolution usually occurs over an extended period of time during which the context changes and new actors get involved. Reducing the complexity of health promotion intervention evaluations in order to create more homogeneous categories cannot be done without changing the nature of the intervention.

This is not to say that evaluation research is useless in health promotion. Quite on the contrary, in order to be able at one point to have the critical mass of evaluation studies that will allow to derive valid estimates, there is a need to intensify our efforts to evaluate health promotion practices. One should beware however that using the pool of existing studies to derive estimates of effects and to then extrapolate this estimate to decisions in specific situations is hazardous at best and can often be misleading. Indeed, one of Cronbach's major contribution to the field of applied research is the warning that ignoring the various higher order interactions that are at play in the production of any scientific fact derived from observation or even from controlled experimentation often results in misleading conclusions.⁴ So, how is it possible to inform decisions and actions in a given situation by tapping into knowledge derived from other similar experience?

It is my impression that in our haste to provide practitioners with guidance for intervention in health promotion, we are making too little use of the other device to frame and make sense of research data: theory. We have a tendency to underrate and invali-

date knowledge derived from a deductive process applied to theoretical knowledge and to overrate knowledge that comes from the accumulation of empirical observations even if the empirical basis is not sufficient. By doing so we have the misleading impression of being more "scientific" and rational since the room for interpretation appears to be smaller in extrapolating from empirically derived evidence than in deducting courses of action for a specific situation from a more general theory about the functioning of the mechanisms at play. This is where in my opinion we should be concentrating our efforts in evaluation.

In any single situation, especially in the open systems of community interventions, there is a variety of social, biological and psychological mechanisms, triggered or not by the intervention, interacting in a specific manner to produce the observed outcomes. It is only in the closed systems of laboratories that one can effectively control for other potential mechanisms in order to isolate one mechanism of interest. In an open system one can only estimate those interaction effects. One can do that inductively by aggregating empirical observations and calculating synthetic estimates whenever the number of observations is sufficient, or one can use theoretical knowledge about the mechanisms at play and deductively disentangle their interactions⁵ according to the characteristics of the situation in which the knowledge needs to be applied. To do this, however, requires a confidence in the practicality of theoretical knowledge that has been lost in public health and in health promotion.

Referências

1. Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. The dependability of behavioural measurements: theory of generalizability of score and profiles. New York: 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. London: Sage; 1997.

Recebido em 2 setembro de 2005

Versão final apresentada em 10 de outubro de 2005

Aprovado em 11 de novembro de 2005