

Using Systematic Reviews for Evidence-based Health Promotion: Basic Methodology Issues

JEFFERSON A. BUENDÍA-RODRÍGUEZ y
JUANA P. SÁNCHEZ-VILLAMIL

Received 30th May 2006/Sent for Modification 10th Jun 2006/Accepted 20th September 2006

ABSTRACT

Systematic reviews and evidence-based recommendations are becoming increasingly important for decision-making in health and medicine. Systematic reviews of population-health interventions are challenging and methods will continue evolving. This paper provides an overview of how evidence-based approaches in public health and health promotion are being reviewed to provide a basis for Colombian Guide to Health Promotion, analysing limitations and recommendations for future reviews.

Key Words: Health promotion, evidence-based medicine (*source: MeSH, NLM*).

RESUMEN

Uso de las Revisiones Sistemáticas en Promoción de la Salud Basada en la Evidencia: Aspectos Metodológicos Básicos

La importancia de las revisiones sistemáticas y de las recomendaciones basadas en la evidencia está aumentando cada vez más, para la toma de decisiones en salud y medicina. Las revisiones sistemáticas de intervenciones poblacionales están aun en crecimiento y sus métodos en continuo desarrollo. Este artículo provee una mirada de cómo una aproximación de salud pública y promoción de la salud basada en la evidencia es útil para la formulación de guías nacionales en promoción de la salud; analizando sus limitaciones y recomendaciones para futuras revisiones.

Palabras Clave: Promoción de la salud, medicina basada en la evidencia (*fuentes: DeCS, BIREME*)

Health promotion interventions tend to be complex and context-dependent. Evaluating evidence must distinguish between an evaluation process's fidelity in detecting the success or failure of an intervention and the relative success or failure of the intervention itself (1). Furthermore, proper interpretation of evidence depends upon the availability of suitable descriptive information regarding the intervention and its context so that the transferability of evidence can be determined. The best available knowledge has to be used in making decisions for populations or groups of patients, as a systematic review (although it may not always be available) provides the best possible knowledge along with clinical decision-making. Moreover, it may be difficult to obtain policy support without evidence of effective health promotion (2). This paper provides an overview of how evidence is reviewed within the context of systematic reviews and how such evidence is translated into recommendations provided in the *Guidelines*.

Traditional literature review of systematic review

The authors of traditional reviews, who may be experts in their field, use informal, unsystematic and subjective methods for collecting and interpreting information, which is often subjectively summarised and is narrative (3). Processes such as searching, quality appraisal and data synthesis are not usually described and, as such, they are very prone to bias. An advantage of these reviews is that they are often conducted by experts who may have a thorough knowledge of the research field; however, a disadvantage lies in the authors possibly having preconceived notions or biases leading them to overestimate the value of some studies.

Many systematic research synthesis tools were developed by American social scientists during the 1960s (4). However, today's systematic evidence-based reviews are very much driven by the evidence-based medicine movement, particularly from the methods developed by the Cochrane Collaboration. A systematic review is defined as being, "a review of the evidence on a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyze data from the studies that are included in the review." (5). A meta-analysis is the statistical combination of studies to produce a single estimate of the effect of the healthcare intervention being considered. A meta-analysis is simply the statistical combination of results from studies; the final estimate of effect may not always be the result of a systematic review of the literature, meaning that It should therefore not be considered as a type of review.

Complexities and challenges of public health reviews

Issues and difficulties which arise and need to be taken into account when synthesising the results of multiple studies are (6) (a) focusing on populations and communities rather than individuals; (b) difficulties characterising and simplifying complex multi-component interventions rather than single interventions; (c) analysing process as well as measuring outcomes; (d) the effect of community members or potential participants' involvement in programmed design and evaluation; (e) the effect of using health promotion theories and beliefs; (f) analysing the use of different types of qualitative and quantitative research; (g) the need for multiple primary papers which may cover the complexity and long-term nature of public health intervention outcomes, and (h) the intervention's integrity highlighting which factors may have influenced the effectiveness of the intervention, such as participation (including suitability), exposure of programmed or intervention, resources, quality of delivery (including training and enthusiasm) and safeguards against intervention contamination (7-8). Conducting systematic reviews of all the available evidence can thus be a complex task, requiring reviewers to have (or have access to) sound content and methodological knowledge and expertise.

Levels of evidence in evidence-based health promotion

The assessment of causality for evidence-based health promotion has mostly depended upon the level of evidence, which has been traditionally defined by the study design used in evaluative research. Study designs are graded by their potential for eliminating bias. A hierarchy of study designs was first suggested by Campbell and Stanley in 1963 (9); levels of evidence based on study design were proposed by Fletcher and Sackett for the Canadian Task-force on Periodic Health Examination in 1979 (10). Systematic reviews of randomised controlled trials (RCT) have become widely accepted as providing the best evidence (level 1) regarding the effects of preventative, therapeutic, rehabilitative, educational or administrative interventions in medicine (11). The concept of levels of evidence has been widely adopted for determining the degree of recommendations for clinical practice, e.g. in US Preventive Services Task Force recommendations and the Canadian Task Force on Periodic Health Examination (12). Levels of evidence have also been applied to other areas of evidence-based decision-making in health, including prognosis, diagnosis and economic analysis.

Steps by structured, systematic review process

1. Formulate the question
2. Comprehensive search
3. Unbiased selection and abstraction
4. Critical appraisal of data
5. Synthesis of data (may include meta-analysis)
6. Interpreting results

Components of an answerable question (PICO)

PICO (population, intervention, comparison, outcome) provides a formula for creating an answerable question. It is also worthwhile at this stage to determine the types of study-design to be included in the review: PICOT.

Population. Health promotion and public health may include populations, communities or individuals. It must be considered whether there is value in limiting the population (e.g. street youth, problem drinkers). These groups are often under-studied and may be different in all sorts of important respects from the study populations usually included in health promotion and public health reviews. Reviews may also be limited to the effects of an intervention on disadvantaged populations to investigate the effect of an intervention on reducing inequalities. Further information on reviews addressing inequality is provided below.

Intervention. Reviewers may choose to lump similar interventions together in a review, or split the review by addressing a specific intervention. Reviewers may also consider approaches to health promotion rather than topic-driven interventions, for example, peer-led strategies for changing behaviour. Reviewers may also want to limit a review by focusing on the effectiveness of a particular type of theory-based intervention.

Comparison. It is important to specify comparison intervention for a review. Comparison interventions may consist of no intervention, another intervention or standard care/practice. The choice of comparison or control has large implications for interpreting results. A question addressing one intervention *cf* no intervention is a different question than one comparing one intervention to standard care/practice.

Outcome. Outcomes chosen for a review must be meaningful for a review's users. The discrepancy between outcomes and interventions which reviewers choose to include in a review and the outcomes and interventions which laypeople prefer to be included has been well-described. Reviewers will need to include process indicators as well as outcome measurement

when investigating both the implementation of an intervention and its effects. Unanticipated (side-effects) and anticipated effects should be investigated in addition to cost-effectiveness.

Examples of review questions

Poorly designed questions: 1. ¿Which interventions reduce health inequalities among people with TBC?

Answerable questions: 1. ¿Does peer-based interventions reduce health inequality in women with TBC?

Evidence of effectiveness

The methods used for providing evidence of effectiveness (by conducting a systematic review) must be sufficiently comprehensive to encompass the complexity of public-health interventions (13). There are many different types of design in public health research, such as randomised controlled trials studies (RCTs), quasi-randomised trials and non-randomised controlled studies. It is useful to distinguish differences regarding such designs before evaluating the quality of the evidence.

Uncontrolled studies are generally not included in reviews, because they have trouble in distinguishing the effects of an intervention from the placebo effect or from what would naturally have occurred. However, RCTs may be uncommon in many areas of public health as they tend to be suited to more simple and straightforward interventions. Randomised controlled trials and quasi-randomised controlled trials refer to trials where participants or populations are randomly allocated (using random number tables) to an intervention or control/comparison group and are followed-up to assess differences in outcome rates (14). A quasi-randomised trial uses a method of allocation which differs from genuine randomisation for methodological (allocation by date of birth, alternate allocation) or pragmatic and policy reasons (allocation by housing-sector). Non-randomised controlled studies (before and after studies) refer to a study design where participants or populations are non-randomly allocated by the investigator to an intervention or control group. The outcome of interest is measured both at baseline and after the intervention period, comparing final values if the groups are comparable at baseline or changes in outcome if not. The lack of randomisation in these types of study may result in groups being different at baseline, as randomisation is the only way of controlling confounders which are not known or not measured (15). Interrupted time series designs are “multiple observations over time that are ‘interrupted’ usually by an intervention or treatment” (16).

These designs may include a control group. Process evaluations (often published separately from outcome evaluations) may also be included in the review, alongside quantitative studies, to assess the adequacy of the delivery of the intervention, and the context in which the intervention was evaluated. Process data have conventionally been drawn from observational quantitative research but increasingly use qualitative and quantitative research methodologies, as appropriate (17).

Cluster RCTs and cluster non-randomised studies

Allocating the intervention by group or cluster is being increasingly adopted within the field of public health because of administrative efficiency, lessened risk of experimental contamination and the likely enhancement of subject compliance (18). Some studies (e.g. class-based nutritional intervention) dictate its application at cluster level.

Interventions allocated at cluster level (e.g. school, class, worksite, community, geographical area) involve particular problems with selection bias where groups are not formed by random allocation but rather through some physical, social, geographical or other connection amongst their members (19). Cluster trials also require a larger sample size than would be required in similar, individually-allocated trials, because the correlation between cluster members reduces the study's overall power (20). Other methodological problems with cluster-based studies include the level of intervention differing from the level of evaluation (analysis) and, frequently, the small number of clusters in a study. Issues surrounding cluster trials have been well described in a Health Technology Assessment report, which should be read for further information if cluster designs are to be included in a systematic review.

The role of qualitative research within effectiveness reviews

“To provide an in-depth understanding of people's experiences, perspectives and histories within the context of their personal circumstances and settings” (18). Qualitative studies can contribute towards reviews of effectiveness in a number of ways, including (5):

- Helping to frame the review question (selecting interventions and outcomes of interest to participants);
- Identifying factors which enable/impede implementing the intervention (human factors, contextual factors);
- Describing the experience of the participants receiving the intervention
- Providing participants' subjective evaluations of outcomes;

- Helping to understand the diversity of effects in studies, settings and groups; and
- Providing a means of exploring the 'fit' between subjective needs and evaluated intervention for developing new interventions or refining existing ones.

Methods commonly used in qualitative studies may include one or a number of the following; interviews (structured around respondents' priorities/interests), focus groups, participant and/or non-participant observation, conversation (discourse and narrative analysis) and documentary and video analysis. The unit of analysis in qualitative studies need not necessarily be an individual or single case; communities, populations or organisations may also be investigated. Anthropological research, which may involve some or all of these methods within the context of wide-ranging fieldwork, can also be a valuable source of evidence, although it may be difficult to subject it to many aspects of a systematic review.

Importance in evaluating health inequality

Health inequality is defined as being, "the gap in health status and access to health services between different social classes and ethnic groups and between populations in different geographical areas." (21) Systematic reviews should consider health inequality when assessing intervention effectiveness. This is because it is thought that many interventions may not be equally effective for all population subgroups. The effectiveness for the disadvantaged may be substantially lower. Evans and Brown (22) have suggested that there are a number of factors (PROGRESS) which may be used in classifying health inequality:

- Place of residence
- Race/ethnicity
- Occupation
- Gender
- Religion
- Education
- Socio-economic status
- Social capital

It may thus be useful for a review of public health interventions to measure the effect in different subgroups. The following data is required for reviews addressing inequality:

- A valid measurement of health status (or change in health status)
- A measurement of disadvantage (e.g. defining socio-economic position)
- A statistical measurement for summarising differential effectiveness.

Retrieving information regarding public health literature

Retrieving information re clinical medicine is facilitated by the clinical medical literature being comparatively well-organised, comparatively easily-accessible through large sophisticated bibliographical databases, the domination of the peer-reviewed journal format and comparatively well-controlled and stable technical terminology (23). Retrieval in public health is much more complicated due to more diverse literature (reflecting its multi-disciplinary nature), a wider range of bibliographical tools of varying coverage and quality and terminological difficulties (24). Identifying public health studies is also problematic because of database indexing, as many studies may not be well-indexed, or indexed differently amongst the databases. Moreover, a great deal of public health research is widely dispersed and may not always be available in the public domain (25).

The key components of the search strategy consist of subject headings and text words describing each PICO(T) question element (population, intervention, comparison, outcome and type of study). However, it is usually recommended not to include the O (outcome) in the PICO question because outcomes are described in many different ways and may not be described in an article's abstract. Search terms for describing outcomes should only be used if the number of citations is too large to apply the inclusion and exclusion criteria. The search strategy should be piloted first (a scoping search on a database most likely to yield studies should be completed by using a sample of keywords to locate a few relevant studies). The subject headings used to index the studies and relevant text words in the citation's abstract should be checked. A combination of subject headings and text words for each PICO element should always be used. The London EPPI-Centre provides support and resources for assisting review researchers conduct sensitive searches (<http://eppi.ioe.ac.uk>) The HP&PH Field has nearly completed a project which will provide recommendations for search terms and hand searching strategies for published and "grey" literature (26).

Critical appraisal of public health interventions

Critical appraisal of intervention research entails assessing the validity of the evidence, the completeness of implementation (intervention integrity) and applicability of such evidence. There are many quality tools or checklists available to help reviewer assess the extent to which a study's methodology sought to minimise bias.

Appraising public health studies is particularly challenging because of (a) difficulty in blinding participants to certain interventions (particularly educational initiatives), (b) the potential for the control/comparison group to become "contaminated" (within schools where participants in intervention and control groups are highly likely to come into contact with each other) and (c) the potential threats to the validity and reliability of data collection methods, particularly where outcomes are subjective (reported behaviour). Incomplete reporting of vital study information hinders complete assessment of study quality (27).

Experienced public health research reviewers thus advocate that public health reviews should assess each included study to determine: complete reporting of the number of participants in control and intervention groups; complete reporting of pre-test and post-test data for all participants in both groups; and the provision of complete data for all outcomes. Such minimum information is needed before studies can be further assessed for random allocation and blinding participants.

Synthesis of results

Public health interventions are often difficult to synthesise because of the complexity of the characteristics of an intervention, the study population, the outcomes measured and other methodological issues (including study design) relating to conducting primary studies (28). Complexity becomes introduced because the effectiveness of an intervention may become modified by the context within which it operates. Because of characteristics' potential variability between studies, reviewers may choose to use a narrative synthesis of results, as calculating a statistical overall estimate would not be meaningful as one would be comparing "apples with oranges." Meta-analysis remains a useful tool for synthesising the outcomes of multiple studies where interventions are homogenous and outcomes comparable; however, meta-analysis may be generally inappropriate because of the degree of heterogeneity between studies.

Applicability

Determining how the results of a review relate to another specific situation, context, or intervention is called applicability, transferability, or generalisability. These terms are essentially synonymous with external validity. This information provided in reviews is particularly relevant to users and their decisions to enable them to assess the applicability of the results to their individual settings. Systematic reviews of public health interventions encompass a number of issues which may complicate determining applicability. Systematic reviews which include a number of studies (with consistent results) which have been conducted in a range of settings, would suggest wide applicability (29). Context refers to the social, organisational and political setting in which an intervention is implemented. Examples of contextual factors which may affect intervention effectiveness include literacy, income, cultural values and access to media and health services.

Conclusions

Evidence-based reviews identify the most effective and efficacious interventions and provide information to help ensure efficient use of resources. The findings of these reviews are targeted to those needing to make decisions about the type of strategies that should be developed and implemented. The advice provided by such reviews should be seen as complementing rather than replacing the practical experience and critical judgments of planners and practitioners. The recommendations need to be carefully considered in the light of the particular context for implementation to ensure a balanced and realistic application. Significant logistical and methodological challenges are associated with reviewing the evidence base for health promotion. The amount of available evidence is often very limited and the quality highly varied. For this reason, these reviews should be seen as a first step only, requiring ongoing enhancement and critical application.

REFERENCES

1. Hawe P. How much trial and error should we tolerate in community trials? *BMJ* 2000; 320:119.
2. Tang KC, Ehsani JP, McQueen DV. Evidence-based health promotion: recollections, reflections and reconsiderations. *J Epidemiol Community Health* 2003; 57: 841–843.
3. Klassen TP, Jadad AR, Moher D. Guidelines for reading and interpreting systematic reviews. 1. Getting Started. *Arch Pediatr Adolesc Med* 1998; 152: 700–704.

4. Hedin A, Kallestål C. Knowledge based public health work. Part 2: Handbook for compilation of reviews on interventions in the field of public health. National Institute of Public Health [Internet]. Sweden 2004 Disponible en: http://www.fhi.se/shop/material_pdf/r200410Knowledgebased2.pdf Consultado: Abril 16 de 2006.
5. NHS Centre for Reviews and Dissemination, University of York. Undertaking systematic reviews of research on effectiveness. CRD's Guidance for those carrying out or commissioning reviews [Internet]. 2nd Ed. 2001. Disponible en: <http://www.york.ac.uk/inst/crd/report4.htm>. Consultado: Febrero de 2006.
6. Jackson SF, Edwards RK, Kahan B, Goodstadt M. An assessment of the methods and concepts used to synthesize the evidence of effectiveness in health promotion: A review of 17 initiatives [Internet]. Sep. 2001. Disponible en <http://www.utoronto.ca/chp/CCHPR/synthesisfinalreport.pdf>. Consultado: Octubre 10 de 2006.
7. Waters E, Doyle J, Jackson N, Howes F, Brunton G, Oakley A. Evaluating the effectiveness of public health interventions: the role and activities of the Cochrane Collaboration. *J Epidemiol Community Health* 2006; 60: 285 - 289.
8. Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clin Psychol Rev* 1998; 18: 23 - 45.
9. Campbell DR, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand McNally College Publishing, 1963.
10. Canadian Task Force on Periodic Health Examination. Periodic Health Examination. *Can Med Assoc J* 1979;121: 1193 - 254.
11. Sackett DL. The Cochrane Collaboration. *ACP Journal Club* 1994 120; (suppl 3): A -11.
12. US Preventive Services Task Force. Guide to clinical preventive services. 2nd ed. Baltimore: Williams and Wilkins; 1996.
13. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health*. 2002. 56; 119 - 27.
14. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess*. 2003; 7(27): III-X, 1-173.
15. Higgins JPT, Green S, editors. Assessment of study quality. *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.5. Actualizado Mayo 2005; Sección 6. [Internet]. Disponible en <http://www.cochrane.dk/cochrane/handbook/hbook.htm>. Consultado: Junio 29 de 2005
16. University of Ottawa. Effective practice and organisation of care group. EPOC methods paper: including interrupted time series (ITS) designs in a EPOC Review [Internet]. Disponible en: <http://www.epoc.uottawa.ca/inttime.pdf>. Consultado: Enero 31 de 2006
17. Steckler A, Linnan LE. Eds. Process evaluation for public health interventions and research. San Francisco: Jossey-Bass, 2002.

18. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organization-based interventions in health and health care: a systematic review. *Health Technol Assess.* 1999; 3(5): III-92.
19. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ.* 2001; 322(7282): 3557.
20. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group randomized trials: a review of recent methodological developments. *Am J Public Health.* 2004 Mar; 94(3): 423-32.
21. Public Health Electronic Library. [Internet].. Disponible en: <http://www.phel.gov.uk/toolsandresources/toolsandresources.asp>. Consultado: Octubre 22 de 2005.
22. Evans T, Brown H. Road traffic crashes: operationalizing equity in the context of health sector reform. *Injury Control and Safety Promotion* 2003; 10(2): 11-12.
23. Grayson L, Gomersall A. A difficult business: finding the evidence for social science reviews, working paper 19. Town: ESRC UK Centre for Evidence Based Policy and Practice; 2003.
24. Beahler CC, Sundheim JJ, Trapp NI. Information retrieval in systematic reviews: challenges in the public health arena. *Am J Prev Med* 2000; 18(suppl 4): 6 - 10.
25. Peersman G, Oakley A. Learning from research. In: Oliver S, Peersman G, eds. *Using research for effective health promotion.* Buckingham: Oxford University Press; 2001. pp. 33 - 43.
26. Armstrong R, Jackson N, Doyle J, Waters E, Howes F. It's in your hands: the value of handsearching in conducting systematic reviews of public health interventions. *J Public Health (Oxf)* 2005; 27: 388 - 91.
27. Peersman G, Oliver S, Oakley A. Systematic reviews of effectiveness. In: Oliver S, Peersman G, eds. *Using research for effective health promotion.* Buckingham, UK: Open University Press; 2001. pp. 100 - 1.
28. Grimshaw JM, Freemantle N, Langhorne P, Song F. Complexity and systematic reviews: report to the US Congress Office of Technology Assessment. Washington, DC: Office of Technology Assessment, 1995.
29. Irwig L, Zwarenstein M, Zwi A, Chalmers I. A flow diagram to facilitate selection of interventions and research for health care. *Bull World Health Organ.* 1998; 76: 17 - 24.