

# A Process for Systematically Reviewing the Literature: Providing the Research Evidence for Public Health Nursing Interventions

B.H. Thomas, RN, MSc, D. Ciliska, RN, PhD, M. Dobbins, RN, PhD, S. Micucci, BA, MSc

## ABSTRACT

**Background:** Several groups have outlined methodologies for systematic literature reviews of the effectiveness of interventions. The Effective Public Health Practice Project (EPHPP) began in 1998. Its mandate is to provide research evidence to guide and support the Ontario Ministry of Health in outlining minimum requirements for public health services in the province. Also, the project is expected to disseminate the results provincially, nationally, and internationally. Most of the reviews are relevant to public health nursing practice.

**Aims:** This article describes four issues related to the systematic literature reviews of the effectiveness of public health nursing interventions: (1) the process of systematically reviewing the literature, (2) the development of a quality assessment instrument, (3) the results of the EPHPP to date, and (4) some results of the dissemination strategies used.

**Methods:** The eight steps of the systematic review process including question formulation, searching and retrieving the literature, establishing relevance criteria, assessing studies for relevance, assessing relevant studies for methodological quality, data extraction and synthesis, writing the report, and dissemination are outlined. Also, the development and assessment of content and construct validity and intrarater reliability of the quality assessment questionnaire used in the process are described.

**Results:** More than 20 systematic reviews have been completed. Content validity was ascertained by the use of a number of experts to review the questionnaire during its development. Construct validity was demonstrated through comparisons with another highly rated instrument. Intrarater reliability was established using Cohen's Kappa. Dissemination strategies used appear to be effective in that professionals report being aware of the reviews and using them in program planning/policymaking decisions.

**Conclusions:** The EPHPP has demonstrated the ability to adapt the most current methods of systematic literature reviews of effectiveness to questions related to public health nursing. Other positive outcomes from the process include the development of a critical mass of public health researchers and practitioners who can actively participate in the process, and the work on dissemination has been successful in attracting external funds. A program of research in this area is being developed.

*Worldviews on Evidence-Based Nursing 2004; 1(3):176–184. Copyright ©2004 Sigma Theta Tau International*

**KEYWORDS** systematic literature reviews, effectiveness, public health, nursing, quality assessment, dissemination

**B.H. Thomas**, Associate Professor, McMaster University School of Nursing, and clinical consultant, Hamilton Public Health and Community Services, PHRED Program. **D. Ciliska**, Professor, McMaster University School of Nursing, and clinical consultant, Hamilton Public Health and Community Services, PHRED Program. **M. Dobbins**, Associate Professor, McMaster University School of Nursing. **S. Micucci**, Project Coordinator, Effective Public Health Practice Project.

Address correspondence to Helen Thomas, School of Nursing, Room 3N28C, McMaster University, 1200 Main St. W., Hamilton, Ontario L8N 3Z5, Canada; thomash@mcmaster.ca

## INTRODUCTION

Numerous systematic literature reviews have appeared in the peer-reviewed literature over the last 10 years.

This article is adapted from a paper presented at the Sigma Theta Tau 14th International Nursing Research Congress, July 2003, St. Thomas, U.S. Virgin Islands.

Submitted 25 November 2003; Accepted 21 January 2004  
Copyright ©2004 Sigma Theta Tau International  
1545-102X/04

The Cochrane Collaboration has established a standardized method for conducting reviews of health care issues (Clarke & Oxman 2003). To date more than 2,500 reviews have been completed within the Cochrane Collaboration. The Campbell Collaboration has also established a process for conducting systematic reviews related to education, social justice, and crime issues (Davies & Boruch 2001). The Guide to Community Preventive Services (GCPS) has established a process for reviewing topics related to public health and has conducted many reviews (Briss et al. 2000; Truman et al. 2000; Hopkins et al. 2001). In the UK, the Centre for Reviews and Dissemination has updated its guidelines for producing systematic reviews of literature to assist in policies and decisions about the organization and delivery of health care (Khan, ter Riet, Glanville, Sowden & Kleijnen 2001). In addition, the Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre provides a number of reviews of the effectiveness of health promotion practices (Oliver & Peersman 2001).

Rimer, Glanz, and Rasband (2001) present a cogent discussion of the challenges of systematic reviewing for health education and health behavior interventions, many of which are relevant to public health nursing. They explain the role of both quantitative and qualitative evidence in determining whether or not an intervention works and why, respectively. They explain that scientific evidence is available for many interventions and that reviews should be undertaken and the results disseminated. When this evidence is not available, they recommend that sound evaluation be undertaken. They outline five challenges to the use of evidence within health promotion/health education: obstacles to finding the evidence, limitations in the state of the science, negative professional values, barriers to dissemination, and communication and training. Also, they provide potential solutions to the barriers.

One of the major issues in conducting systematic reviews is the method for assessing the quality of included primary studies. Two reviews of the instruments to assess the methodological rigor of the primary studies have recently been reported. Deeks et al. (2003) reviewed 194 scales and checklists in current use for assessing the methodological rigor of nonrandomized intervention studies included in systematic literature reviews. They concluded that although a few seemed to have strong theoretical basis and content validity, the description of tool development was generally poor. Most lacked information about reliability and validity. However, six instruments were suggested for future use. In 2002, the Agency for Healthcare Research and Quality (AHRQ) produced a report that overviewed 109 methods for assessing systems to rate the strength of scientific evidence (West et al. 2002). They concluded that there were eight accept-

**TABLE 1**Domains<sup>a</sup> included in evaluation of scales and checklists

<b>OBSERVATIONAL STUDIES</b>	<b>RCTs</b>
Study question	Study question
Study population	Study population
Comparability of subjects	Randomization
Exposure/intervention	Intervention
Outcome measure blinding	Outcome measure blinding
Statistical analysis	Statistical analysis
Results	Results
Discussion	Discussion
Funding	Funding

<sup>a</sup>West et al. (2002), pp. 37, 39.

able approaches to reviewing randomized controlled trials (RCTs), and six to evaluating observational studies. The domains used to assess the approaches for observational studies and for RCTs are outlined in Table 1. Each domain has one to six criteria based on empirical evidence and/or best practice.

In 1998, the Effective Public Health Practice Project was funded by the Ontario Ministry of Health and Long-Term Care (MOHLTC), Public Health Branch, Public Health Research, Education and Development (PHRED) Program. Its mandate is to provide research evidence to guide/support the provincial Mandatory Health Programs and Services Guidelines (MHPSG; Ontario Ministry of Health, Public Health Branch 1997), which outline minimum requirements for public health in the province. These guidelines encompass a wide range of content areas (e.g., chronic disease prevention, early detection of cancer, injury prevention, substance abuse prevention, sexual health, reproductive health, child and family health, and control of infectious diseases). A variety of intervention strategies such as one-to-one, small groups, and community formats to provide health education, behavior change education, screening, immunization, and community development are also outlined in the MHPSG.

The project began because of the many challenges facing public health and health promotion practitioners and policymakers in finding and interpreting the relevant literature. For many topics there is a huge amount of literature available in a variety of databases and journals. These are not easily accessible to all policymakers and program planners. Also, the process of locating, retrieving, and reading the literature is very time-consuming. The situation is further complicated by the fact that for those who do not understand research methodology, the contradictions in conclusions from one study to the next can be confusing.

In policy/program decision making, a number of factors including the clinical skills of the professionals within

the organization, individual/group/community preferences for services, current financial resources available, and opportunity costs of program implementation must be considered. However, the availability of summaries of the current research literature is also crucial (Dobbins, Cockerill, Barnsley & Ciliska 2001b).

The purposes of this article are to describe the following four issues related to systematic literature reviews of the effectiveness of public health nursing interventions produced by the EPHPP: (1) the process of systematically reviewing the literature, (2) the development of a quality assessment instrument, (3) the results of the EPHPP to date, and (4) some results of the dissemination strategies used.

## METHODS

### Steps in an EPHPP Systematic Review

There are seven steps in an EPHPP review (see Figure 1). These include question formulation, literature retrieval, developing relevance criteria, assessing studies for relevance and then for methodological quality, data extraction and narrative synthesis, peer review of the report, and dissemination. Each review takes between 7 to 9 months to complete.

Once a question has been formulated, a review group is constituted, consisting of four to six people including a review group leader. Review group membership includes at least one methodological expert in systematic reviews and two or more topic content experts. Members are recruited from the PHRED programs in the province. The PHRED program consists of five university/public health unit affiliations, which have as their mandate relevant research, education of future public health practitioners, and development and evaluation of innovative public health programs. PHRED staff members are jointly appointed between the respective university and public health unit. The EPHPP staff provides support for the reviews.

### Question Formulation

Topics for review are selected through a number of processes. First, through a biennial survey of the 37 public health departments in the province, relevant and timely topics are identified. Second, specific topics may be submitted by the Public Health Branch, MOHLTC. Third, practitioners around the province may submit relevant topics. The EPHPP Steering Committee makes the final decision based on relevance to the Mandatory Health Programs and Services Guidelines and whether the topic can be formulated into a researchable question. The question has to clearly identify a population of interest (e.g., school-aged

children and youth), an intervention(s) (e.g., school-based awareness campaigns and/or strategies to increase physical activity), and an outcome (e.g., increased physical activity). In this example, the question is "Are school-based interventions effective in increasing physical activity among children and youth?"

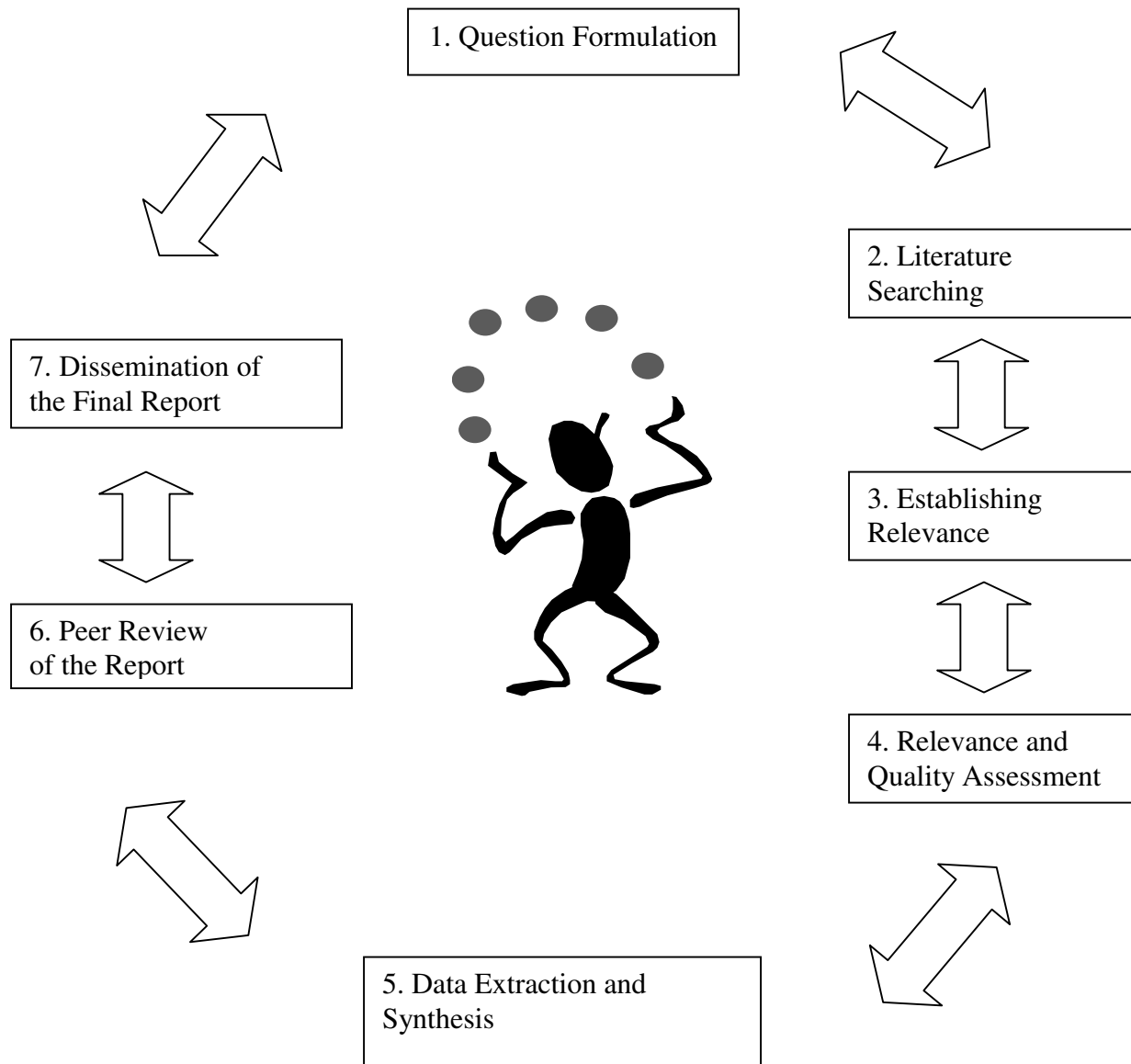
A major issue in defining a question is whether to keep it broad ("lumping") or make it very specific ("splitting"). While others (e.g., EPPI-Centre) have opted to answer very broad questions in their reviews, we have chosen to keep our questions quite specific. We have done this for two reasons. First, practitioners and policymakers want answers to specific questions, and so our reviews and their summary statements provide this. Second, keeping questions specific limits any one literature search and retrieval. Given that the "hit" rate for relevant articles in an electronic search regarding public health topics is about 10%, any review requires a lot of reviewer time to select the relevant articles from those identified. When topics are broad, the "hit" rate can be even lower, requiring more resources.

### Searching and Retrieving the Literature

At least seven electronic databases (e.g., CINAHL, MEDLINE, EMBASE, Eric, Cochrane Library, Sociological abstracts, PsychINFO) are searched. The dates for searching are dependent upon the topic. Usually databases are searched from their inception. The electronic search is organized around a number of subheadings that are generated related to the topic of interest. Then all text-word synonyms are included within each heading. The review group leader and one of the staff independently review the initial captured titles and abstracts and mark all those that appear relevant for retrieval. Full articles of titles selected by either reviewer are retrieved. Because electronic searching sometimes misses relevant articles, five or six relevant journals are hand searched for the previous five years. Key informants and known experts in the field are also contacted to supply other references. The gray literature is also searched. Bibliography lists of all retrieved articles are searched and relevant studies retrieved.

### Establishing Relevance Criteria

Each review group establishes its unique relevance criteria. These include a description of the population of interest, the intervention(s), the outcomes, and the study design. In public health, RCTs are available in some areas of research, but sometimes these are either not feasible or unethical, so observational studies with comparison groups are included as well. The other relevance criterion is usually that the intervention is within the scope of public health nursing practice.



**Figure 1. Steps in an EPHP systematic review.**

The inclusion of observational studies in systematic reviews has been the subject of much discussion (Kunz & Oxman 1998; Moher et al. 1998; McKee et al. 1999; Guyatt, DiCenso, Farewell, Willan & Griffith 2000; MacLehose et al. 2000). Some suggest it increases the effect size in a meta-analysis (Guyatt et al. 2000). Others refute this (Kunz & Oxman 1998; McKee et al. 1999; MacLehose et al. 2000). It appears that the effects of including observational studies in meta-analysis depend on at least three factors: (1) the topic of the review, (2) the quality of RCTs available, and (3) the quality of observational studies available.

Using the relevance criteria developed for the specific review, all articles are independently assessed for relevance by two reviewers. Discrepancies are resolved through discussion and an informal consensus process. If the two reviewers are unable to agree, a third person assesses the article and determines its relevance. Studies that meet all the relevance criteria are eligible for quality assessment.

**Quality Assessment of Relevant Studies**

Many of the evaluations of effectiveness studies in public health/health promotion interventions include not only

**TABLE 2**  
Quality assessment components and ratings for EPHPP instrument

COMPONENTS	STRONG	MODERATE	WEAK
Selection bias	Very likely to be representative of the target population and greater than 80% participation rate	Somewhat likely to be representative of the target population and 60–79% participation rate	All other responses or not stated
Design	RCT and CCT	Cohort analytic, case-control, cohort, or an interrupted time series	All other designs or design not stated
Confounders	Controlled for at least 80% of confounders	Controlled for 60–79% of confounders	Confounders not controlled for, or not stated
Blinding	Blinding of outcome assessor and study participants to intervention status and/or research question	Blinding of either outcome assessor or study participants	Outcome assessor and study participants are aware of intervention status and/or research question
Data collection methods	Tools are valid and reliable	Tools are valid but reliability not described	No evidence of validity or reliability
Withdrawals and dropouts	Follow-up rate of >80% of participants	Follow-up rate of 60–79% of participants	Follow-up rate of <60% of participants or withdrawals and dropouts not described

RCTs, but also nonrandomized studies. Therefore, it was necessary to develop an appropriate quality assessment tool to encompass a variety of research designs.

The quality assessment instrument for primary studies was developed using the following process. The research team reviewed available instruments in the peer-reviewed literature and devised a list of components to be included (Mulrow & Oxman 1994; Jadad et al. 1996). The components were sample selection, study design, identification and treatment of confounders, blinding of outcome assessors and of participants, reliability and validity of data collection methods, and withdrawals and dropouts. The components are rated strong, moderate, or weak according to a standardized guide and dictionary (see Table 2 for details). The overall rating for the study is determined by assessing the six component ratings. Those with no weak ratings and at least four strong ratings are considered strong. Those with less than four strong ratings and one weak rating are considered moderate. Finally, those with two or more weak ratings are considered weak. (Copies of the instrument and dictionary are available from the first author on request.) Strong and moderate studies are included in a review.

Two other areas are included in the assessment, but they are not included in the overall score. First, the integrity of the intervention refers to what proportion of participants actually received the duration and intensity of the intervention as it was designed. In public health, particularly in group interventions, this is very important because outcomes can differ depending upon the amount of intervention received. Also, consistency of the intervention is important because if few participants received the intervention as designed, the outcome of no differ-

ence between groups relates to program delivery rather than its effectiveness (Rychetnik, Frommer, Hawe & Shiell 2002). Since many public health interventions relate to population-wide health concerns (e.g., smoking cessation), it is important to know if contamination or cointervention occurred in order to accurately interpret results of a specific study.

Second, use of appropriate analysis is also an important consideration in the strength of a study. In the pilot testing of the instrument, this item for quality assessment was included; however, many reviewers had difficulty responding to the questions. After training, there was still very low interrater reliability. Data are collected about the unit of allocation, the unit of analysis, appropriateness of the statistical methods, and whether or not the analyses are performed by intervention allocation status (i.e., intention to treat) or by actual intervention received. Information about the integrity of the intervention and the use of appropriate analysis is used in the summary of the article, if it is included in the review. These data are often helpful in understanding the results of the studies.

To assess content validity of the instrument, the draft was circulated to six experts in appraising study quality for their comments and feedback. Once their comments were incorporated, the final draft of the instrument was independently pilot tested on ten primary studies by four experts in critical appraisal and community health. Further clarification of the instrument was completed. A detailed dictionary to describe the items and to clarify the criteria for rating each item was developed. Finally, guidelines for assessing the overall strength of the articles were described. (The instrument and dictionary are available from the first author.)

Construct validity was assessed by comparing the component ratings to those of the Guide to Community Preventive Services (GCPS) instrument. Both of these instruments were recommended for future use (Deeks et al. 2003). The two instruments scored some of the components slightly differently. For example, all components of the GCPS instrument were scored either yes/no or not applicable, whereas the Effective Public Health Practice Project (EPHPP) instrument had ordinal scoring for four of the components (i.e., selection bias, controlling for confounders, exposure to the intervention, and number of withdrawals/dropouts). In order to compare these components, the EPHPP items were collapsed into yes/no responses. To determine the levels of agreement of the 11 similar components between the two quality assessment instruments, 70 relevant primary studies from two of the EPHPP reviews were assessed independently by two reviewers using the GCPS instrument (Thomas et al. 1999; Black, Yamada, Mann, Cava & Micucci 2000). Prior to this assessment, the two assessors were trained in the use of the GCPS instrument. The EPHPP data from the previously completed reviews were used for comparison. Percentage agreement between ratings of each component was calculated using McNemar's chi-square (Norman & Streiner 2003). This test was used because it takes into account the matching of samples.

Two reviewers assessed a random selection of primary studies a second time to determine test-retest reliability. The Kappa statistic was used to establish the level of test-retest reliability (Norman & Streiner 2003).

For each review, the review group leader and one of the support staff independently assess the quality of the relevant articles. Differences are resolved through discussion and an informal consensus process.

### **Data Extraction and Synthesis**

Data are extracted from each strong and moderate study using a standardized format that includes the following variables: the funding source; the number of participants in each group and the number of dropouts; a description of the target population, the intervention(s), and the outcome(s); and the length of follow-up.

The results are usually synthesized narratively. Meta-analysis is not used for three reasons. Clinical heterogeneity of the populations in different studies often makes it not sensible to combine the results (Egger, Davey Smith & Schneider 2001). For example, interventions may be directed at community samples as well as those at high risk for the problem of interest. For some topics (i.e., heart health), there are multiple interventions. Also, the multiple outcomes found in many public health interventions make determining the most important one, and therefore the one upon which to base the meta-analysis, difficult. For

example, in a review of parenting groups with professional involvement in improving child and parent outcomes, 42 reliable and valid outcomes related to the parent, child, or parent-child interactions were identified (Thomas et al. 1999). Although these fell into the three groups described, the outcomes were different. The other challenge in this situation was that each outcome was available for a very small number of studies. Consequently, results are summarized narratively.

### **The Written Report**

The review group leader synthesizes all of the information into a written report. Members of the review group critique the report and changes are made. At this stage, it is peer-reviewed by at least five experts in the field. Final revisions are made and it is disseminated.

### **Dissemination**

Evidence suggests that passive dissemination of information does not result in much knowledge transfer (Grimshaw et al. 2001). Consequently a variety of strategies are used to enhance information uptake and utilization. Hard copies of the full reports and two-page, user-friendly summary statements are circulated to appropriate personnel at the Public Health Branch of the Ontario Ministry of Health and Long-Term Care and to all of the provincial medical officers of health in the 37 health units. Also, they are placed in all the health unit libraries. All of the reports and summary statements are also available on the Web site (<http://www.hamilton.ca/phcs/ephpp>). Presentations are made to academic and clinical practice conferences locally, nationally, and internationally. Several of the reviews have also been published in peer-reviewed journals. Because of their expertise, different members of the project sit as its representative on several national and international committees. The director is also involved with a Cochrane methods group to assure that the methodology of the review process is current. Reviewers have been invited to be members of relevant provincial policymaking committees. Finally, some of the reviews have been used in externally funded knowledge transfer research projects.

## **RESULTS**

Systematic reviews conducted to date include those within the following areas of public health nursing: chronic disease prevention ( $n = 6$ ), early detection of cancer ( $n = 2$ ), injury and substance abuse prevention ( $n = 2$ ), sexual health ( $n = 3$ ), reproductive health ( $n = 4$ ), child and youth health ( $n = 5$ ), and infectious diseases ( $n = 3$ ).

The EPHPP instrument has been used in all of the reviews. Based on feedback from reviewers, it is relatively

**TABLE 3**  
EPHPP and GCPS instruments: Quality assessment components and specific items used in measuring agreement

COMPONENTS	EPHPP	GCPS
Study design	Included	Not included
Sample description	Not included	Included
Selection bias	A1	2C
Confounding	C1 & 2	5B
Blinding of assessors	D1	3C
Blinding of participants	Included	Not included
Validity of data collection tools	E1	3C
Reliability of data collection tools	E2	3C
Exposure to intervention	G1	3A
Integrity of intervention	G2	1B
Contamination/co-intervention	G3	5C
Participation rates	A2	2D
Withdrawals/dropouts	F2	5A
Data analysis	H3	4A or B

easy to use. The dictionary that accompanies it clarifies any questions related to the components. The use of content experts in its development and other recent work indicate that it has adequate content validity. The Effective Public Health Practice Project (EPHPP) instrument was among the six recommended for future use by Deeks et al. (2003). Table 3 outlines the components of both the EPHPP and the Guide to Community Preventive Services (GCPS) instruments. The numbers and letters in each column indicate where the components are found in each instrument. As outlined, the EPHPP instrument shares 11 of the 14 components with the GCPS questionnaire, which was rated highly by both Deeks et al. (2003) and West et al. (2002). The EPHPP quality assessment tool does not collect data about the sample description. This is included in the standardized data extraction form. The GCPS instrument does not include the study design component, although this is factored into the decision about study quality and inclusion in a review. Although both tools include assessing blinding of outcome assessors, the GCPS does not address blinding of participants to intervention.

A comparison of the results of rating the component items of the GCPS and the EPHPP questionnaires for primary studies from two reviews provides some evidence of construct validity for the EPHPP instrument. Table 4 outlines the overall frequency of agreement by component and review topic for the 70 primary studies. Agreement ranges from 63% to 92% for the parenting review and between 53% and 88% for the cervical cancer screening review. More than 79% agreement was found for eight of the 11 components in the parenting review. In the cervical cancer screening review, the level of agreement was more than 75% for

**TABLE 4**  
Levels of agreement for components by review topic

COMPONENTS	PARENTING REVIEW <sup>a</sup>	CERVICAL CANCER SCREENING REVIEW <sup>b</sup>
	<i>N</i> = 30 % AGREEMENT ( <i>p</i> VALUE)	<i>N</i> = 40 % AGREEMENT ( <i>p</i> VALUE)
Selection bias	87.5 (1.000)	52.5 (0.004)
Confounding	70.8 (1.000)	75.0 (0.754)
Blinding	95.8 (1.000)	77.5 (1.000)
Validity	91.7 (1.000)	65.0 (0.791)
Reliability	91.7 (1.000)	52.5 (0.359)
Intervention exposure	70.8 (0.016)	62.5 (0.035)
Intervention integrity	62.5 (0.004)	42.5 (0.001)
Contamination	91.7 (1.000)	25.0 (0.001)
Participation rate	79.2 (0.063)	72.5 (0.012)
Withdrawals and dropouts	79.2 (0.063)	77.5 (0.180)
Data analysis	66.7 (0.070)	87.5 (1.000)

<sup>a</sup>Thomas et al. (1999).

<sup>b</sup>Black et al. (2000).

five of the items. Where differences occurred, the EPHPP item consistently rated more negatively than the GCPS item. Test-retest (intrarater) reliability as measured by the Kappa statistic was 0.74 and 0.61 for the two reviewers, respectively.

Data extracted from the primary studies have been narratively summarized for all of the reviews. Statistically significant positive changes from pretest to posttest are considered positive outcomes, and therefore the intervention is effective. Negative changes are considered to be negative outcomes, and the intervention is not effective. No statistically significant outcome means that the effectiveness of the intervention is unknown. In this situation, the power of the sample is examined to detect a difference (this is a frequent flaw in public health nursing evaluation studies). The integrity of the intervention is also examined to try to explain the causes of the result.

The dissemination strategies seem to be effective in that surveys indicate that the products of the project are well known; professionals find them very helpful and report using them for program planning decisions (Dobbins, Cockerill & Barnsley 2001a; Dobbins et al. 2001b). During a recent update of the provincial Mandatory Health Programs and Services Guidelines, the reviews and their authors were actively involved in developing new policies. Ninety-two percent of those participating in the development of these new policies reported that the EPHPP reviews were used in recommending changes in the guidelines for public health practice in Ontario (Dobbins et al. 2001b).

The Web site and the project office receive numerous requests for the reviews. Recently, all of the reviews were entered into the Database of Abstracts of Reviews of Effects (DARE) of the Cochrane Library, which allows relatively easy access to them for an international audience.

## CONCLUSIONS

The EPHPP has demonstrated the ability to adapt the most current methods for systematic literature reviews of effectiveness to questions related to public health. There are at least two limitations to the work and a number of positive outcomes from the process to date. First, a standardized way to use qualitative work in the reviews has not yet been determined. One method would be to use it, when possible, to explain the outcomes. Second, producing systematic reviews is very labor intensive and requires staff with sophisticated skills in data management, design methodology, critical appraisal of literature, communication, and organization. This means that a well-developed infrastructure is necessary to complete systematic literature reviews.

The process for completing systematic literature reviews of public health nursing interventions has had a number of positive outcomes. First, over 4 years, 25 reviews have been completed. Second, the process used has engaged public health researchers and practitioners around the province and developed a critical mass of professionals who can actively participate in the process, understand its importance, and keep their respective constituents updated about new reviews as they become available.

Third, the instrument developed to assess the methodological quality of the primary studies with a variety of study designs has demonstrated content and construct validity. Although the levels of agreement between the EPHPP and the GCPS instruments vary for each item and between the two reviews, most are within acceptable ranges. Further work may be needed to determine if this variation in agreement is topic dependent. Fourth, the test-retest reliability appears to be adequate. Because the current empirical evidence demonstrates the potential dangers in summing scores of the quality assessment components, in the future we will determine a priori the most important components and include studies in a review that include these components.

Fifth, the project has been involved in a number of externally funded dissemination activities that appear to be effective. This is a very complex field that is in its infancy. We plan to continue to seek external funding to develop a program of research in this area.

In summary, this project has demonstrated that systematic reviews of the effectiveness of public health nursing

interventions can be efficiently completed and can impact positively upon policymakers and program planners.

### Implications of systematic literature reviews of effectiveness of interventions related to public health nursing

- Current methods for systematically reviewing the quantitative literature can be successfully adapted for public health nursing questions.
- Quality assessment of primary studies for inclusion in a review requires the use of instruments with known psychometric properties.
- The results of systematic literature reviews are useful for policy development, program planning, and developing future research questions.

## References

- Black M., Yamada J., Mann V., Cava M. & Micucci S. (2000). *Effectiveness of strategies to increase cervical cancer screening: A systematic review of the literature 1989–1999*. Ontario, Canada: Public Health Research, Education and Development Program (PHRED).
- Briss P.A., Zaza S., Pappaioanou M., Fielding J., Wright-De Agüero L., Truman B.I., et al. (2000). Developing an evidence-based Guide to Community Preventive Services—Methods. The Task Force on Community Preventive Services. *American Journal of Preventive Medicine*, 18(1 Suppl), 35–43.
- Clarke M. & Oxman A.D. (Eds.). (2003). *Cochrane reviewers' handbook. The Cochrane Library* (Issue 2). Oxford, UK: Update Software.
- Davies P. & Boruch R. (2001). The Campbell Collaboration. Does for public policy what Cochrane does for health. *British Medical Journal*, 323(7308), 294–295.
- Deeks J.J., Dinnes J., D'Amico R., Sowden A.J., Sakarovich C., Song F., Petticrew, M. & Altman, D.G. (2003). Evaluating non-randomised intervention studies. *Health Technology Assessment*, 7(27), 1–173.
- Dobbins M., Cockerill R. & Barnsley J. (2001a). Factors affecting the utilization of systematic reviews. A study of public health decision makers. *International Journal of Technology Assessment in Health Care*, 17(2), 203–214.
- Dobbins M., Cockerill R., Barnsley J. & Ciliska D. (2001b). Factors of the innovation, organization, environment, and individual that predict the influence five systematic reviews had on public health decisions. *International Journal of Technology Assessment in Health Care*, 17(4), 467–478.
- Egger M., Davey Smith G. & Schneider M. (2001). Systematic reviews of observational studies. In M. Egger, G. Davey Smith & D.G. Altman (Eds.), *Systematic reviews*



- in health care: Meta-analysis in context*, 2nd ed. (pp. 211–227). London: BMJ Books.
- Grimshaw J.M., Shirran L., Thomas R., Mowatt G., Fraser C., Bero L., et al. (2001). Changing provider behavior: An overview of systematic reviews of interventions. *Medical Care*, 39(8 Suppl 2), I12–45.
- Guyatt G.H., DiCenso A., Fawcett V., Willan A. & Griffith L. (2000). Randomized trials versus observational studies in adolescent pregnancy prevention. *Journal of Clinical Epidemiology*, 53(2), 167–174.
- Hopkins D.P., Briss P.A., Ricard C.J., Husten C.G., Carande-Kulis V.G., Fielding J.E., et al. (2001). Reviews of evidence regarding interventions to reduce tobacco use and exposure to environmental tobacco smoke. *American Journal of Preventive Medicine*, 20(2 Suppl), 16–66.
- Jadad A.R., Moore R.A., Carroll D., Jenkinson C., Reynolds D.J., Gavaghan D.J., et al. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, 17(1), 1–12.
- Khan K.S., ter Riet G., Glanville J., Sowden A.J. & Kleijnen J. (2001). *Undertaking systematic reviews of research on effectiveness. CRD's guidance for those carrying out or commissioning reviews*. York, UK: NHS Centre for Reviews and Dissemination, University of York.
- Kunz R. & Oxman A.D. (1998). The unpredictability paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. *British Medical Journal*, 317(7167), 1185–1190.
- MacLehose R.R., Reeves B.C., Harvey I.M., Sheldon T.A., Russell I.T. & Black A.M.S. (2000). A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technology Assessment*, 4(34), 1–154.
- McKee M., Britton A., Black N., McPherson K., Sanderson C. & Bain C. (1999). Methods in health services research. Interpreting the evidence: Choosing between randomised and non-randomised studies. *British Medical Journal*, 319(7205), 312–315.
- Moher D., Pham B., Jones A., Cook D.J., Jadad A.R., Moher M., et al. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*, 352(9128), 609–613.
- Mulrow C.D. & Oxman A.D. (1994). *The Cochrane collaboration handbook*. Oxford, UK.
- Norman G. & Streiner D. (Eds.). (2003). *PDQ statistics*, 3rd ed. Hamilton, Ontario, Canada: B.C. Decker Inc.
- Oliver S. & Peersman G. (Eds.). (2001). *Using research for effective health promotion*. Buckingham, UK: Open University Press.
- Ontario Ministry of Health, Public Health Branch. (1997). *Mandatory health services and program guidelines*. Toronto: Queen's Printer.
- Rimer B.K., Glanz D.K. & Rasband G. (2001). Searching for evidence about health education and health behavior interventions. *Health Education Behavior*, 28(2), 231–248.
- Rychetnik L., Frommer M., Hawe P. & Shiell A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health*, 56(2), 119–127.
- Thomas H., Camiletti Y., Cava M., Feldman L., Underwood J. & Wade K. (1999). *Effectiveness of parenting groups with professional involvement in improving parent and child outcomes*. Ontario, Canada: Public Health Research, Education and Development Program (PHRED).
- Truman B.I., Smith-Akin C.K., Hinman A.R., Gebbie K.M., Brownson R., Novick L.F., et al. (2000). Developing the Guide to Community Preventive Services—Overview and rationale. The Task Force on Community Preventive Services. *American Journal of Preventive Medicine*, 18(1 Suppl), 18–26.
- West S., King V., Carey T.S., Lohr K.N., McKoy N., Sutton S.F., et al. (2002). *Systems to rate the strength of scientific evidence*. Evidence Report/Technology Assessment No. 47. Rockville, MD: Agency for Healthcare Research and Quality.