Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Grace Oguntebi Date

The Association between Methodological Reporting Quality of Systematic Reviews Produced by the *Guide to Community Preventive Services* and the Recommendations and Findings Issued by the Task Force on Community Preventive Services, 2000-2010

By

Grace Oguntebi
Master of Public Health

Global Epidemiology

_____

Anne Spaulding, MD, MPH
Committee Chair

_____

Theresa Sipe, PhD, MPH, RN
Committee Member

_____

Stephen Thacker, MD, MSc
Committee Member

The Association between Methodological Reporting Quality of Systematic Reviews Produced by the *Guide to Community Preventive Services* and the Recommendations and Findings Issued by the Task Force on Community Preventive Services, 2000-2010

By

Grace Oguntebi

B.S., Georgia Institute of Technology, 2008

Thesis Committee Chair: Anne Spaulding, MD, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2012

ABSTRACT

The Association between Methodological Reporting Quality of Systematic Reviews Produced by the *Guide to Community Preventive Services* and the Recommendations and Findings Issued by the Task Force on Community Preventive Services, 2000-2010

By Grace Oguntebi

Background: An increasing demand for evidence-based healthcare is placing new emphasis on the methodological quality of systematic reviews. The *Guide to Community Preventive Services* (the Community Guide) was developed to conduct systematic reviews evaluating the effectiveness of public health interventions. Each intervention is issued a finding from the Task Force on Community Preventive Services (Task Force), based on the strength of the evidence obtained in the review. These findings and recommendations are increasingly being implemented in public health policies and programs.

Objective: To determine the association between methodological reporting quality in Community Guide systematic reviews and the corresponding Task Force finding. We also sought to evaluate the effect that three covariates (publication year, topic area, or the type of intervention considered in the review) have on the methodological reporting quality.

Methods: Community Guide systematic reviews were selected from five topic areas, and from each review that satisfied the inclusion and exclusion criteria, data were extracted on descriptive information, quality of the included studies, methodological reporting quality (based on five selected methodological characteristics), and the Task Force finding. Associations between the methodological reporting quality and the Task Force finding, and the effects of the three covariates on the methodological reporting quality, were evaluated using logistic regression modeling techniques.

Results: In the 72 systematic reviews included, the average number of methodological characteristics addressed was 1.72 out of a possible 5. A Task Force finding of strong evidence vs. insufficient evidence was significantly associated with increased methodological reporting score (OR=2.31) and increased study quality score (OR=2.24). Significant associations of similar magnitude were also observed for the comparison of sufficient vs. insufficient evidence. Reporting of the individual methodological characteristics was significantly different at varying values of the three covariates evaluated.

Conclusions: All findings are of importance in public health, whether a recommendation with strong evidence supports an intervention's effectiveness, or a finding of insufficient evidence suggests areas for future research. For any finding, thorough and transparent methodological reporting contributes to the validity of the systematic review, increasing its usefulness to public health policy-makers and professionals.

The Association between Methodological Reporting Quality of Systematic Reviews Produced by the *Guide to Community Preventive Services* and the Recommendations and Findings Issued by the Task Force on Community Preventive Services, 2000-2010

By

Grace Oguntebi

B.S., Georgia Institute of Technology, 2008

Thesis Committee Chair: Anne Spaulding, MD, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2012

ACKNOWLEDGEMENTS

I would like to thank my faculty advisor, Dr. Anne Spaulding, for her guidance and encouragement throughout all the ups and downs of my thesis work. I would also like to acknowledge Dr. Stephen Thacker and Dr. Theresa Sipe, who made their time and their expertise available to me through the development of this project.

To my colleagues at the Rollins School of Public Health, thank you for inspiring me to broaden my horizons. And to my friends, thank you for your constant encouragement – whether from the next room over or from many miles away.

Finally, I want to express the utmost gratitude to my family for their unending support – emotionally, mentally, and spiritually – and for believing in me every step of the way.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Tables

Figures

CHAPTER I

LITERATURE REVIEW

Overview

Systematic reviews have in recent years taken on an increasingly important role in social, behavioral, and biomedical research. The Cochrane Collaboration, an international organization focused on the production and promotion of systematic reviews addressing the effects of interventions in health care, defines the *systematic review* as "a review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review" (1).

This chapter will begin with a discussion of the general methodology for conducting a review, the considerations needed when including observational studies as opposed to randomized controlled trials only, and the implications of current practice in the reporting of reviews. Later, the chapter will introduce the *Guide to Community Preventive Services* and its role in producing evidence-based recommendations for public health interventions through the conduct and reporting of high-quality systematic reviews. The chapter will conclude with a brief description of selected reporting guidelines for various types of research, and the role and impact that reporting guidelines have on the quality of reviews.

General Methodology in the Conduction of Systematic Reviews

The primary objective in the conduct of reviews is to perform a systematic summarization of all of the available evidence pertaining to a particular topic and to obtain a conclusion regarding the effect of the intervention or treatment under

consideration. The "systematic" feature of this process involves precise methodology in order to minimize bias and maintain the reliability of the results produced by the review (2). Many groups and authors have thoroughly explained the process of conducting a review (3, 4).

To begin, the researchers must define an explicit research question which they hope to address in the review. An important aspect of the review is the completion of a thorough search, including both published and unpublished sources, for the entirety of the evidence that is relevant to the selected research question. Such a search would yield references that are outside the scope of the desired evidence, based on inclusion and exclusion criteria that have been defined prior to the beginning of the study. Complete documentation of the search results and the decisions made to include or exclude certain references along each phase of the screening process, along with the reasoning behind those decisions, contributes to the replicability of the review (4).

Once the set of references relevant to the review topic has been obtained, the researchers abstract the specific data needed in order to calculate the summary measure from each study. In addition, the quality of each study is assessed in order to determine the potential for biases that might affect internal validity. The review process does not change the individual study quality scores; instead, the researchers might comment on the role that the biases present in the individual studies might play in the results (5).

A second group of biases might affect the external validity of the review because these biases arise from the methodology used in the conduction of the review. The publication and dissemination of the results of primary research are strongly influenced by the actual results of the study, and these reporting biases are known to lead to over-representation of studies with statistically significant, or positive, results (6). Studies

with these positive results are more likely to be present in the review because they are more likely to be published (publication bias), published quickly, published in English, published multiple times, and more likely to be references in other research (3). These types of biases can be minimized by employing as broad a search strategy as possible, in order to capture a higher proportion of the existing evidence that is pertinent to the research question, and not only those studies with positive results. In addition, various statistical methods can be used to assess the presence of publication bias, the results of which are factored into the discussion and explanation of the results obtained.

As noted earlier, the overall objective of the review is to obtain a summarizing conclusion that answers the research question. The researchers must decide whether it is appropriate to combine the results from each individual study into one conclusion for the review (3). An important issue to consider when making this decision is the extent of between-study heterogeneity that exists due to differences in the population, setting, interventions, and outcomes considered by each study (4, 7). For example, it may not be appropriate to combine the study-specific results if there is much heterogeneity between studies. Whether or not the study-specific results are pooled together, interpretations can be drawn from the review and generalized to the appropriate populations and/or circumstances.

Observational Studies vs. Randomized Controlled Trials

Because randomized controlled trials (RCTs) are considered to be the strongest study designs, systematic reviews of RCTs are also considered to provide the strongest summary of research (8). However, conducting RCTs may be difficult or impossible when addressing certain research questions. For instance, it is unethical to randomize a

study participant to an exposure that is likely to be harmful or to a treatment group where the side effects of the treatment are unknown. In these situations, observational studies fill the research gap by contributing to the overall foundation of knowledge regarding an exposure-outcome association.

Some experts argue that reviews of observational studies should be abandoned altogether (9). Indeed, there are several arguments in favor of including only RCTs in a review. Because randomization would ideally control for any known or unknown confounding factors, these studies are less prone to bias due to confounding, as compared to observational studies. Also, blinding of the study subjects and the investigators and data analysts would minimize biases due to the placebo effect; that is, with both the study participants and the researchers unaware of the members of each treatment group, any outcomes can be confidently attributed to the intervention itself (10).

Thirdly, observational studies can take on one of several different study designs, including cohort, cross-sectional, and case-control, each with varying strengths of evidence. These different study designs make comparability between studies, a necessary characteristic of a high-quality review, difficult. Also, straightforward procedures for pooling effect estimates become more complicated when working with studies of varying designs (11). Lastly, the use of observational studies will increase the likelihood of having to address heterogeneity between the studies (12).

It is generally accepted that observational study designs are weaker than RCTs; however, recent improvements in study designs, such as enhanced methods for controlling for potential confounders, have increased the quality of some observational studies. Multiple groups of researchers have obtained similar effect estimates when

conducting reviews using high-quality, prospective cohort studies as compared to reviews using RCTs (10). Furthermore, lapses in quality can also occur with RCTs, especially in the process of random allocation and in methods of allocation concealment (13). Therefore, a distinction of RCT does not automatically translate into a better quality study.

One group suggested that observational studies be included along with RCTs in reviews, and also mentioned that it may be appropriate to forgo the meta-analysis process, which is the statistical combination of effect estimates (10). However, if a meta-analysis were conducted, the multiple different study designs could be investigated as a potential explanation for any between-study heterogeneity (9). In any case, the additional studies in the review would increase the overall sample size for the review and would add to the pool of evidence that contributes to answering the research question.

Implications of Current Practice in Reporting of Systematic Reviews

Experts across many fields in health research have noted that, despite the efforts of journal editors and peer reviewers, the overall quality of reporting is inadequate (14). These observations have stemmed from both primary studies and the reviews that synthesize their results. Although what the investigators report is not always the same as what is actually done, the only insight that the consumers of the review have into the methods of that review is what is documented in the report. Clear documentation of methods allows for a higher degree of transparency when examining the strengths and weaknesses of the review. Therefore, it is necessary that investigators are thorough in the reporting of their review methodologies.

A well-conducted review has several potential practical applications for several different consumer groups, including healthcare organizations, clinicians, funding groups, other experts in that research area, and the public. First, the results and conclusions of reviews are often used to inform the development of health care policies, resulting in the practice of evidence-based medicine. Fund-granting organizations may also rely on reviews to determine areas in which financial resources are needed to conduct additional research (15). To facilitate the use of reviews in this way, the results and conclusions regarding the overall effect of the intervention being considered should be clearly reported (4). Furthermore, detailed reporting of the methodology facilitates the process of replication of the review by other researchers, which increases the strength of the findings (16).

When important aspects of the review methodology are not reported, the result can be a decreased level of confidence in the results and conclusions of the review (17). Specifically, certain steps in the review process, when omitted or inadequately addressed, can affect the validity of the results. Biases can result from the following issues: search strategies may not be broad enough to identify all relevant trials, researchers may fail to address clinical or statistical heterogeneity, and statistical methods for pooling may not be appropriate for the studies involved (2). Another issue that can limit the usefulness of a review is time; reviews that are not updated in a timely manner may become outdated and no longer applicable (17).

The Community Guide

The *Guide to Community Preventive Services: Systematic Reviews and Evidence-Based Recommendations* (the Community Guide) was developed by the Task Force on

Community Preventive Services (Task Force) (18). The primary purpose of the Community Guide is to offer recommendations on public health interventions that focus on preventing disease and promoting health in populations. Supported by staff from the Centers for Disease Control and Prevention (CDC) and other organizational partners, the Community Guide makes these recommendations via systematic reviews that are conducted according to a very specific, standardized methodology. The original Community Guide was expected to be published in 2001, consisting of a set of reviews on interventions that were selected by each review team. Over time, updated chapters and new chapters considering additional topic areas have been completed (18).

Several aspects of the Community Guide development process make it a particularly useful resource for those involved in planning, funding, and implementing population-based services and policies. For instance, scientific knowledge forms the foundation for the decisions made, and this knowledge has been considered, analyzed, and interpreted by a panel of experts (19). However, because reviews vary in many ways, including the topic areas, the types of study designs included, and the staff conducting the review, there are many opportunities for inconsistencies between reviews. The use of a standardized procedure for data collection contributes to minimizing these variations (20). In addition, the conduction of every Community Guide review includes eight steps that transform the evidence obtained into recommendations (18):

1. Form multidisciplinary chapter development teams
2. Develop a conceptual approach to organizing, grouping, and selecting the interventions evaluated in each chapter
3. Select interventions to be evaluated

4. Search for and retrieve evidence
5. Assess the quality of and summarize the body of evidence of effectiveness
6. Translate evidence of effectiveness into recommendations
7. Consider evidence other than effectiveness
8. Identify and summarize research gaps

Based on the strength of the evidence obtained, the Task Force issues a finding for each intervention that is evaluated. The finding can fall into one of three categories: recommended, in which the review provides strong or sufficient evidence for the effectiveness of the intervention; recommended against, in which the review provides evidence that the intervention is harmful or not effective; and insufficient evidence, in which the review does not provide enough evidence to determine any effectiveness of the intervention. In the recommended category, the distinction between strong evidence and sufficient evidence is based on the number of studies considered, the study designs of those studies, and the consistency of the effect across studies (21).

Impact of Standards for Reporting Reviews

One recent development that has aided reviewers towards the goal of complete, transparent reporting has been the creation and dissemination of various reporting guidelines. Multiple groups and initiatives have purposed to describe sets of reporting guidelines for the conduct of reviews under varying circumstances. One such group is CONSORT (Consolidated Standards of Reporting Trials), which in 1996 first published its proposal for the structured reporting of RCTs (22). These standards included a checklist of information items that should be included in the report and a flow diagram

template outlining the process of subjects' flow from eligibility and registration into the study to randomization to withdrawal or completion of the study. The completed flow diagram should contain the numbers of participants and the rationale for any losses at each phase of the study. The CONSORT reporting standards have since been updated twice to reflect feedback from users of the guidelines and to include new developments in the methodology for conducting RCTs (23, 24).

One limitation of these reporting standards is that most of the developing groups do not have formalized plans for the dissemination, implementation, and evaluation of the standards (14). However, various studies evaluating the use of CONSORT have shown that reports of RCTs have improved in quality, especially within journals that have formally adopted the guidelines (13). Journals can endorse the use of CONSORT by adding a statement to their instructions to authors that they follow the CONSORT guidelines, or by requiring that authors submit the checklist noting where in their manuscript each item is referenced (25). One study compared the methodology of RCTs published in four medical journals before and after the release of the CONSORT statement. The researchers found that there was a significant increase over time in the number of items on the CONSORT checklist that were adequately reported (23).

Standards for reporting have also been developed for systematic reviews and meta-analyses. One of the first such proposals was published by the QUOROM group in 1999, with the aim of improving the Quality of Reporting of Meta-analyses, focusing on meta-analyses of RCTs (26). As with CONSORT, the QUOROM statement consists of a checklist of items that are important to include in the report, in addition to a flow diagram. The flow diagram for a review follows the progress of potentially relevant studies obtained by the initial search through each step of the screening process,

including reasons for exclusion at each stage (Figure 1). The QUOROM statement was recently updated, and the new statement, now known as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) has expanded its scope to explicitly include systematic reviews and to incorporate updates in review methodology (27).

As discussed earlier, reviews of observational studies require certain considerations and methodologies that are not applicable to reviews of RCTs. The reporting guidelines proposal for meta-analyses of observational studies in epidemiology (MOOSE) is similar to those mentioned previously in that it includes a checklist of methodological recommendations that provide guidance and instruction for the conduction and reporting of meta-analyses of observational studies (28). The checklist suggests those aspects of review methodology that should be included in each section of the report: background, search strategy, methods, results, discussion, and conclusion (Figure 2). The MOOSE workgroup hoped that these guidelines might address the variability that had been observed in the reporting of these meta-analyses.

CHAPTER II

MANUSCRIPT

The Association between Methodological Reporting Quality of Systematic Reviews Produced by the *Guide to Community Preventive Services* and the Recommendations and Findings Issued by the Task Force on Community Preventive Services, 2000-2010

Grace Oguntebi

ABSTRACT

Background: An increasing demand for evidence-based healthcare is placing new emphasis on the methodological quality of systematic reviews. The *Guide to Community Preventive Services* (the Community Guide) was developed to conduct systematic reviews evaluating the effectiveness of public health interventions. Each intervention is issued a finding from the Task Force on Community Preventive Services (Task Force), based on the strength of the evidence obtained in the review. These findings and recommendations are increasingly being implemented in public health policies and programs.

Objective: To determine the association between methodological reporting quality in Community Guide systematic reviews and the corresponding Task Force finding. We also sought to evaluate the effect that three covariates (publication year, topic area, or the type of intervention considered in the review) have on the methodological reporting quality.

Methods: Community Guide systematic reviews were selected from five topic areas, and from each review that satisfied the inclusion and exclusion criteria, data were extracted on descriptive information, quality of the included studies, methodological reporting quality (based on five selected methodological characteristics), and the Task Force finding. Associations between the methodological reporting quality and the Task Force finding, and the effects of the three covariates on the methodological reporting quality, were evaluated using logistic regression modeling techniques.

Results: In the 72 systematic reviews included, the average number of methodological characteristics addressed was 1.72 out of a possible 5. A Task Force finding of strong evidence vs. insufficient evidence was significantly associated with increased methodological reporting score (OR=2.31) and increased study quality score (OR=2.24). Significant associations of similar magnitude were also observed for the comparison of sufficient vs. insufficient evidence. Reporting of the individual methodological characteristics was significantly different at varying values of the three covariates evaluated.

Conclusions: All findings are of importance in public health, whether a recommendation with strong evidence supports an intervention's effectiveness, or a finding of insufficient evidence suggests areas for future research. For any finding, thorough and transparent methodological reporting contributes to the validity of the systematic review, increasing its usefulness to public health policy-makers and professionals.

INTRODUCTION

Evidence-based health care has in recent years had an increasingly larger impact on the direction of social, behavioral, and biomedical research. The principal strategy is to link high-quality scientific evidence to recommendations made in public health and clinical practice. A systematic review compiles all of the available evidence regarding a topic and presents a conclusion on the effect of the intervention being considered, thus reducing the time between the publication of primary evidence and the implementation of recommendations (2). A meta-analysis goes one step further to incorporate a quantitative synthesis of the findings obtained from the included studies, usually obtained via a systematic approach (6). Therefore, meta-analyses are often conducted in conjunction with systematic reviews. Unless otherwise specified, the term *review* will be used to refer to any systematic review, with or without a corresponding meta-analysis.

For any given topic, the body of scientific literature is often extensive, varying in quality, and varying in the results, making it difficult for practitioners to arrive at any one conclusion (19). However, these conclusions are valuable in the process of planning, funding, and implementing health care policies, both at the population level and at the individual level. An example of the role of research synthesis in formulating clinical guidelines involves a meta-analysis conducted over three decades ago, which concluded that corticosteroids given to mothers delivering prematurely are effective in reducing morbidity and mortality in the infants. This same conclusion could have been reached using evidence that was available as early as ten years prior to publication of the meta-analysis (2).

In order to ensure the reliability of the results, systematic reviews and meta-analyses must be conducted according to a precise methodology. Certain components in

particular, if poorly conducted, can affect the validity of the review. For instance, the search strategy might not incorporate all relevant studies, including non-English language studies and studies that were never published (29). Another common limitation in the methodology of reviews is that the quality of the individual studies is not assessed. Biases in these primary studies can in turn compromise the results of the review (5).

The convention within the public health community of relying on expert opinion and individual judgments has gradually been replaced by an increasing demand for evidence-based decision making, thus placing a new emphasis on the methodological quality of systematic reviews (30). The overall quality of a review depends on both the quality of the evidence and the quality of the methodological approach (5). Clear documentation of the methods undertaken in a review allows for a higher degree of transparency when examining the strengths and weaknesses of the review (14). Therefore, it is necessary that investigators are thorough in the reporting of their review methodologies. The *Guide to Community Preventive Services: Systematic Reviews and Evidence-Based Recommendations* (the Community Guide) was developed to present recommendations on the effectiveness of population-based public health interventions based on results from systematic reviews. For each intervention, the Task Force on Community Preventive Services (Task Force) issues a finding that depends on the strength of the evidence obtained in the review (18).

Objective

In a cross-sectional study design, we used a selection of reviews conducted by the Community Guide to evaluate the effect that the publication year, topic area, or the type of intervention considered in the review have on the quality of reporting of the methods. We also sought to determine if there is any association between the Task Force finding that was issued and the quality of reporting of the methodology in those reviews, and to identify the effect, if any, that the three factors mentioned above have on this association.

METHODS

A. Selection of Systematic Reviews

For this analysis, systematic reviews were selected from five of the topic areas on which reviews have been conducted by the Community Guide: preventing excessive alcohol consumption, cancer prevention and control, motor vehicle-related injury prevention, tobacco use, and vaccinations to prevent disease. On March 18, 2012, the website of the Community Guide (www.thecommunityguide.org) was accessed in order to obtain reports of the systematic reviews that have been conducted on these topics. The systematic reviews selected were published between the years of 2000 and 2010 in the *American Journal of Preventive Medicine*, due to the agreement between this journal and the CDC, which provides the principal staffing for all Community Guide activity (31).

The initial selection included all reports for which sufficient citation information was available on the website. Initially, three publications were selected within each topic area. Because some publications included reports of multiple systematic reviews, the number of reviews included in each topic could vary. In order to obtain a sufficient number of reviews within each topic, additional publications were randomly selected. When a saturation point was reached for a topic area, meaning that significantly more systematic reviews had been selected compared to other topics, no additional reviews in that topic area were considered for inclusion in the study.

B. Exclusion Criteria

The quality of execution of the individual studies in each review was assessed by the authors, according to the standardized data collection procedure used for all

Community Guide reviews (20). The data collection instrument consists of 23 questions related to quality assessment, such as "Was the intervention well described (what, how, who, where)?" and "Did the authors specify the screening criteria for study eligibility?" The questions address six categories of threats to validity: descriptions of the study population and interventions, sampling, measurement of the exposure and outcome, data analysis, interpretation of results, and other (18). Up to nine limitations can be assigned to each paper, based on failure to address the questions. A study that has 0 – 1 limitation is categorized as having good quality of execution, a study with 2 – 4 limitations is categorized as having fair quality of execution, and a study with 5 or more limitations is categorized as having limited quality of execution. According to Community Guide review procedures, studies with limited quality of execution are not included in the evidence used to issue findings. Therefore, any reviews that, following the assessment of quality of execution, resulted in zero studies qualifying for the review were also excluded from the analysis.

C. Data Extraction

All data were extracted by a single researcher who was not blinded with respect to the authors, institutions, or journal of publication. The following descriptive information was extracted from each review: the title of the review, the year of publication, and the topic area (one of the five topics listed above). Additional information was extracted regarding the intervention being evaluated, the type of intervention, the number of studies included in the analysis, including the number of studies with each level of execution quality (good, fair, or limited), and the resulting finding from the Task Force (recommendation with strong evidence, recommendation

with sufficient evidence, or insufficient evidence). The intervention type was classified according to the implementation mechanism of the intervention: whether it was directed towards an individual, towards an entire group or community, or implemented via a policy or law.

D. Definition of Variables

The primary outcome variable was defined as the finding issued by the Task Force regarding sufficiency of evidence in the systematic review. The predictor variables were two scores that together represent the overall quality of the review.

The first, called the MOOSE score, represents the quality of the methodological approach: it corresponds to the number of methodological characteristics that was addressed in the review. In order to evaluate the methodological quality of each review, five methodological characteristics were selected from the MOOSE statement, one from each section of the reporting checklist (see list below) (28). Each characteristic was selected based on its impact on the biases present in the review results and its impact on how conclusions are understood by the consumers of the review. These five characteristics are not specific to meta-analyses of observational studies only, but can apply to any systematic review. Each review was assigned a 'yes' rating for each characteristic if it was addressed to any degree in the review, and a 'no' rating otherwise.

The following five methodological characteristics were used to evaluate the quality of reporting for each review:

1. Type(s) of study designs used – from reporting of background

2. Qualifications of searchers (use of librarian in search) – from reporting of search strategy
3. Assessment of heterogeneity – from reporting of methods
4. Forest plot graphic summarizing the individual study estimates and the overall estimate – from reporting of results
5. Considerations of alternative explanations for observed results – from reporting of discussion/conclusions

The second predictor variable, study quality score, represents the quality of the evidence included in the review by incorporating the number of studies together with each study's corresponding quality of execution. The study quality score was calculated by summing the number of papers with good quality of execution, multiplied by a weight of two, with the number of papers with fair quality of execution, multiplied by a weight of one.

E. Statistical Analysis

Data analysis was completed using SAS software, version 9.3 of the SAS System for Windows, Copyright © 2010 (SAS Institute, Cary, NC).

Because the outcome variable, Task Force finding, had more than two categories, ordinal logistic regression was used to determine any association between the predictor variables, MOOSE score and study quality score, and the corresponding Task Force finding for that review. The use of the ordinal logistic regression model requires that the proportional odds assumption be met, and the Score test was used to validate this assumption. When this test statistic was found to be significant, indicating a violation of

the proportional odds assumption, polytomous logistic regression was used instead. Although this model does not reflect the ordinality of the category values in the outcome variable, it is still preferred to fitting multiple individual logistic models because it can incorporate all three categories into the same model (32).

Univariate analyses considered the association between the MOOSE score and the study quality score separately, with the Task Force finding outcome. Multivariate logistic models were also fit to examine the extent of confounding involving the following covariates: the year the review was published, the topic area of the review, and the type of intervention evaluated in the review. A backwards elimination modeling strategy was used to generate the models, beginning with the full model containing both independent variables and the three covariates (Table 1). For each predictor variable, additional multivariate models were generated by dropping the covariate that was least significant in the previous model. Similar models were fit with each the five individual methodological characteristics considered separately as the independent variable of interest.

These five individual characteristics were also modeled as the outcome variables in a second set of logistic regression analyses. The three covariates previously considered were incorporated into the models as independent variables of interest, in order to evaluate the association between these factors and the reporting of the methods in the reviews. The year of publication was grouped into three categories, along the natural divisions in the values for the systematic reviews that were included in the analysis: 2000-2001, 2004-2005, and 2008-2010. All variables were considered in both univariate models and multivariate models, which were generated using an all-possible-regression-models strategy.

RESULTS

Descriptive Statistics

A total of 117 systematic reviews of public health interventions conducted by the Community Guide were identified in the initial selection process. From this group, 32 reviews were excluded due to over-representation of reviews in certain topic areas the original data set. An additional 13 reviews were excluded from analysis due to having reported that zero studies qualified for the review, leaving a set of 72 reviews for analysis (Table 2). Regarding the intervention types considered in each systematic review (individual, group or community, and policy or law), almost half of the reviews evaluated interventions directed at the individual, and the next highest proportion evaluated group or community-level interventions (Table 3). One review evaluated a combination of interventions including all three intervention types (33). Regarding the topic areas of the included reviews, the largest proportions dealt with the topic areas of vaccinations (30.6%) and cancer control and prevention (29.1%). Reviews in the topic area of preventing excessive alcohol consumption comprised the smallest proportion of those considered in this analysis (Table 3).

Reporting of the five variables relating to methodological characteristics included in this analysis was low (Figure 1). The mean number of characteristics addressed in the reviews was 1.72 ± 1.26. Only two of the 72 reviews addressed all five characteristics of interest and another four reviews addressed any four of the five characteristics, while 14 reviews did not address any of the characteristics. The most frequently reported characteristic was assessment of heterogeneity via subgroup analysis, while the least often reported methodological characteristic was the use of a librarian for the search.

The systematic reviews included were published in seven different years. Because some reviews were reported in the same publication, some of those years have more reviews published than other years (Figure 2). There were 16, 19, and 20 reviews published in 2000, 2001, and 2008, respectively; in contrast, only one of the reviews included in this analysis was published in 2004. The average number of MOOSE methodological criteria reported in the reviews published in each year showed increasing trends from 2000 to 2004 and again from 2008 to 2010 (Figure 3). The findings issued by the Task Force regarding sufficiency of evidence in these reviews also varied by year (Figure 4). However, in the three years mentioned above which had comparable numbers of reviews published (2000, 2001, and 2008), the findings issued were split approximately evenly between recommendations with strong or sufficient evidence and the finding of insufficient evidence.

Associations with Task Force Finding

Univariate Analyses

The study quality score, which represents the quality of evidence included in the review, had a significant association with the Task Force finding (Table 4). The odds ratio (OR) representing the association between the study quality score and a review being issued a recommendation with strong evidence compared to being found with insufficient evidence was 2.33 ($p=0.0038$). When considering the effect of the study quality score on a review receiving a finding of sufficient vs. insufficient evidence, the OR was 2.32 ($p=0.0040$).

The MOOSE score, which corresponds to the number of MOOSE criteria that were addressed in the review, also had a significant positive association with the Task

Force finding. For each additional methodological characteristic that was addressed in the report, a review was more likely to receive a finding of strong evidence versus insufficient evidence (OR=2.31, p=0.0021), and also more likely to receive a finding of sufficient evidence versus insufficient evidence (OR=2.24, p=0.0126).

Two of the individual methodological characteristics also had significant associations with the Task Force finding: reporting on assessment of heterogeneity and including a forest plot graphic of the study estimates. For reviews that received a recommendation with strong evidence, the odds of having addressed heterogeneity in the report were 13.2 times those odds for reviews that were found to have insufficient evidence ($p < 0.0001$). For reviews that received a recommendation with sufficient evidence, this same OR for having considered an assessment of heterogeneity was 7.04 (p=0.0098). The OR for the association between a review including the forest plot graphic and being issued a recommendation with strong evidence compared to being found with insufficient evidence was 11.69 (p=0.0005). This same OR comparing reviews with recommendations of sufficient evidence and reviews found with insufficient evidence was 6.86 (p=0.0200).

Multivariate Analyses

Three different models were used to examine the effect of potential confounders on any association between the number of MOOSE criteria addressed and the study quality score jointly with the Task Force finding. The three potential confounders considered were: the year of publication, the intervention type, and the topic area of the review. The final model selected to represent this association controlled only for the year of publication (Model C, Table 1). In this model, the study quality score was

significantly associated with the Task Force finding: with each additional point on the study quality score, a review was more likely to receive a recommendation with strong evidence (OR=1.49, p=0.0101) or sufficient evidence (OR=1.49, p=0.0102), as opposed to being found with insufficient evidence (Table 5). The MOOSE score was also significantly associated with the Task Force finding, but only when considering the likelihood of receiving a recommendation with strong evidence, versus being found with insufficient evidence (OR=2.09, p=0.0210). However, when the study quality score and the MOOSE score were considered individually as independent variables, none of the three covariates considered (the publication year, the intervention type, and the topic area of the review) had any significant effect on either association.

As mentioned previously, two methodological characteristics were independently associated with the Task Force finding. The following three MOOSE characteristics did not have a significant association with the Task Force finding when considered independently: reporting of the types of study designs included in the review, use of a librarian in the search, and consideration of alternative explanations for the results. When potential confounding by the covariates was examined in all possible regression models considering the five methodological characteristics independently as the predictor variables, the following result was true: there was no significant association observed between any of these characteristics and the Task Force finding.

Associations with Individual Methodological Characteristics

Univariate Analyses

We examined the association of each covariate as an independent predictor of the reporting of each methodological characteristic (Table 6). The three categories for the

year of publication were each significantly associated with different characteristics. Reviews published in 2000 or 2001 were less likely to include a forest plot graphic of the study effect estimates (OR=0.190, p=0.0020) and less likely to consider alternative explanations for the results (OR=0.212, p=0.0138) than reviews published in other years. Reviews published in 2004 or 2005 were more likely to report using a librarian for the search when compared to reviews published in other years (OR=15.999, p=0.0006). Finally, reviews published most recently, in 2008 through 2010, were less likely to include the types of study designs used in the review (OR=0.273, p=0.0183), but more likely to include a forest plot graphic (OR=6.120, p=0.0007) and more likely to report on alternative explanations for the results (OR=4.750, p=0.0074).

The three intervention types categorized in this analysis were also each significantly associated with different methodological characteristics. A review of an intervention targeted towards the individual, when compared to other reviews, was less likely to include alternative explanations for the results (OR= 0.148, p=0.0057). A review of an intervention targeted towards a group or community was less likely to describe the types of study designs included in the review (OR=0.327, p=0.0427), while a review of a policy or law intervention was more likely to do so (OR=5.466, p=0.0199). Reviews considering group or community interventions were also less likely to include an assessment of heterogeneity (OR=0.340, p=0.0350) or to include a forest plot graphic (OR=0.238, p=0.0131).

As with the different intervention types, the five topic areas included in this analysis were also each significantly associated with different methodological characteristics. Reviews dealing with preventing excessive alcohol consumption were more likely than non-alcohol-related reviews to include alternative explanations for the

results (OR=10.600, p=0.0475). Reviews dealing with cancer prevention and control were less likely than other reviews to report the types of study designs included in the review (OR=0.044, p=0.0034), but were more likely to include a forest plot graphic (OR=3.899, p=0.0124). In contrast, reviews dealing with vaccinations were less likely to include a forest plot than non-vaccination-related reviews (OR=0.241, p=0.0218) and also less likely to include an assessment of heterogeneity (OR=0.230, p=0.0087). Reviews in the topic area of motor vehicle injury prevention were more likely than other reviews to report using a librarian in the search (OR=8.594, p=0.0052), while reviews in the topic area of tobacco use were more likely to report on assessment of heterogeneity (OR=3.720, p=0.0391).

Multivariate Analyses

Various combinations of the three predictor variables were also incorporated simultaneously into multivariate models, with each methodological characteristic as the outcome. However, none of these models demonstrated a significant association between any combination of predictor variables and any outcome variable (results not shown).

DISCUSSION

Principal Findings

Overall, the methodological reporting quality of these reviews did not satisfy the five MOOSE criteria that were selected for evaluation. The mean number of methodological characteristics addressed was 1.72 out of a possible 5. This result follows a trend of weak reporting in systematic reviews and meta-analyses observed in other studies, and not just within the Community Guide. One survey of 164 reviews published between 1955 and 1990 found that out of 23 reporting items, the mean number addressed was 8.94, although this average increased over time (34).

As expected, the study quality score, which incorporates the number of studies included in the review with the quality of execution of those studies, was associated with the Task Force finding of sufficiency of evidence. The Task Force makes its recommendation based on three factors: the number of available studies, the strength of the design and execution of those studies, and the degree of consistency in the reported effects (18). Therefore, we would expect to see a positive association between these two variables.

We also observed a significant association of similar magnitude between the MOOSE score, representing the reporting quality of the review, and the Task Force finding. Reviews that addressed more of the methodological characteristics were more likely to have a finding of strong or sufficient evidence. Specifically, reviews that included a forest plot graphic of the study effect estimates or that reported an assessment of heterogeneity were more likely to have a finding of strong or sufficient evidence. This result could be explained by the possibility that these reviews tended to include more studies, and therefore the report might have been thought to be improved

by a forest plot. The availability of more studies for the review would also have increased the likelihood of that review receiving a Task Force finding of strong or sufficient evidence. These analyses did not exclude reviews with a small number of studies, nor did they control for the number of studies in each review.

Nonetheless, inclusion of a forest plot does not necessarily indicate having included many studies in the review. A recent assessment of current practice of the use of forest plots in systematic reviews found that 71% of the 639 forest plots in the study contained data on three or fewer studies (35). Interestingly, some of the forest plots even had no studies, perhaps to emphasize the paucity of research that had been conducted in that particular topic area.

In these Community Guide reviews, the inclusion of a forest plot seems to have increased in recent years compared to earlier years. Although the Community Guide methodology guidelines do not require quantitative synthesis (meta-analysis) of results from individual studies, the increased prevalence of forest plots might indicate an increasing trend in the utilization of this technique. Forest plots may be used as a graphical method for summarizing the results of the meta-analysis, or alternatively, if no quantitative synthesis is performed, may display only the effect estimates and confidence intervals from the individual studies (35).

One area in which reporting seemed to worsen over time was in the reporting of information regarding the designs of the individual studies included in the review. This item is particularly significant to the report because the recommendation made by the Task Force depends on the strength of the study design. Furthermore, some study designs are more appropriate than others when considering certain types of public health questions (10). A recommendation based on a systematic review that

incorporates unsuitable types of study designs might weaken the strength and impact of the recommendation. Information on the types of study designs included would allow the readers of a systematic review to better analyze the quality of the review and evaluate its applicability to different circumstances and situations.

The five topic areas considered in this study were associated with good reporting quality for different methodological characteristics. This result may be due to an interaction of the topic area with the investigators involved in preparing the report. Authors have different reporting styles and may find certain aspects of the review methodology more important to report than others. Authors may also focus their work on reviews in certain topic areas. These variations may explain the differences observed in the methodological reporting quality of reviews across different topic areas.

Strengths and Limitations

To our knowledge, this is the first analysis of the reporting quality of the systematic reviews conducted by the Community Guide. The use of methodological characteristics from the MOOSE guidelines statement is particularly relevant when considering systematic reviews conducted by the Community Guide. Because the Community Guide focuses on evaluations of population-based interventions, many of the individual studies included in the reviews are observational studies. Therefore, the MOOSE guidelines, which deal primarily with synthesis of observational studies, can be confidently applied to these reviews.

Moreover, this study considered five methodological characteristics that, when inadequately addressed, have been shown to impact the conclusions of a systematic review. For example, the inclusion of a librarian as part of the systematic review team

has been shown to contribute to the formulation and reporting of a search strategy that is effective and reproducible (36). Additionally, an assessment of heterogeneity is important in establishing the consistency of the results across the individual studies in order to determine the appropriate method for combining these results and the appropriate application of the findings of the review (7).

There were several limitations to this study. First, the data from each systematic review were abstracted by only one investigator, allowing for mistakes in data collection. Decisions regarding whether or not a methodological characteristic was addressed in a review were not validated by a second investigator, also allowing for inadvertent inclusion or omission in the identification of methodological characteristics. Secondly, the investigators did not contact the authors of the reviews to obtain missing information. For example, approximately half of the reviews did not report the number of studies that were rated with each quality of execution (good, fair, or limited). Therefore, the study quality score could only be calculated for 38 of the 72 reviews, and any analyses involving the study quality score were completed with this decreased sample.

Lastly, the MOOSE guidelines statement was not intended to be used as an instrument to assess the quality of systematic reviews, although other studies have used it for that purpose (37). Our study did not implement the entire MOOSE statement as a tool to evaluate the quality of the Community Guide reviews; instead, we selected key characteristics that are relevant to the validity of systematic review findings. However, our results might have been more robust if we had used one complete, validated method for evaluating reporting quality. In doing so, the analyses would have included a full picture of the methodological reporting quality for each review. Additionally, we could

have found significant associations between other methodological characteristics and the Task Force findings regarding sufficiency of evidence.

Recommendations for Future Research

In addition to assessing the reporting quality of other methodological characteristics beyond the five that were considered in this study, there are several other aspects of this study that could benefit from further research. An interesting finding was that the reporting quality of reviews that evaluated interventions targeting a group or community appeared to be the lowest of all three intervention types evaluated (individual, group or community, and policy or law). These reviews were least likely to report the types of study designs included in the review, least likely to include an assessment of heterogeneity in the report, and least likely to include a forest plot showing the study effect estimates. The number of studies in these reviews did not differ meaningfully from that of reviews considering other intervention types: the average number of studies included in reviews evaluating group- or community-level interventions was 9.2, while the average number of studies included in reviews evaluating individual and policy/law interventions was 9.3 and 9.0, respectively.

This finding could be investigated further by repeating the analysis and incorporating additional covariates in the multivariate models. One potential covariate is the lead author of the systematic review team. As mentioned earlier, reporting styles may differ from author to author, and authors may specialize in the evaluation of interventions in specific topic areas or of certain types. The number of studies included in the review could also be included as a potential covariate because, although it might

not have affected the results regarding the intervention type, this factor could play a role in other associations.

Finally, expanding the data set to include all of the systematic reviews that have been conducted by the Community Guide since its inception would present a more complete picture of the methodology employed in these reviews. The results of such an analysis would also more accurately represent the relationship between the findings of sufficiency of evidence issued by the Task Force and the reporting quality of the reviews.

REFERENCES

1. PLoS Medicine Editors. Many reviews are systematic but some are more transparent and completely reported than others. *PLoS Med.* 2007;4(3):399-400.
2. Moher D, Olkin I. Meta-analysis of randomized controlled trials: a concern for standards. *JAMA.* 1995;274(24):1962-1964.
3. Anonymous. Open learning material for reviewers. Alderson P, Green S, editors. Cochrane Collaboration; 2002.
4. Littell JH, Corcoran J, Pillai V. Systematic reviews and meta-analysis. New York, NY: Oxford University Press; 2008.
5. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? *Health Technol Asses.* 2003;7(1).
6. Finckh A, Tramer MR. Primer: strengths and weaknesses of meta-analysis. *Nat Clin Pract Rheum.* 2008;4(3):146-152.
7. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327(7414):557-560.
8. Maier W, Moller H-J. Meta-analyses: a method to maximise the evidence from clinical studies? *Eur Arch Psy Clin N.* 2010;260:17-23.
9. Egger M, Schneider M, Smith GD. Spurious precision? Meta-analysis of observational studies. *BMJ.* 1998;316(7125):140-144.
10. Shrier I, Boivin J-F, Steele RJ, Platt RW, Furlan A, Kakuma R, et al. Should meta-analyses of interventions include observational studies in addition to

randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol.* 2007;166(10):1203-1209.

11. Reeves B, van Binsbergen J, van Weel C. Systematic reviews incorporating evidence from nonrandomized study designs: reasons for caution when estimating health effects. *Eur J Clin Nutr.* 2005;59(Supplement 1):S155-S161.
12. Delgado-Rodriguez M. Glossary on meta-analysis. *J Epidemiol Community Health.* 2001;55:534-536.
13. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ.* 2009;339:b2700.
14. Altman DG, Simera I, Hoey J, Moher D, Schulz KF. EQUATOR: reporting guidelines for health research. *Lancet.* 2008;371:1149-1150.
15. Bero LA, Jadad AR. How consumers and policymakers can use systematic reviews for decision making. *Ann Intern Med.* 1997;127(1):37-42.
16. JARS Working Group. Reporting standards for research in psychology: Why do we need them? What might they be? *Am Psychol.* 2008;63(9):839-851.
17. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med.* 2007;4(3):447-455.
18. Briss PA, Zaza S, Pappaioanou M, Fielding J, Wright-de Aguero L, Truman BI, et al. Developing an evidence-based *Guide to Community Preventive Services* - methods. *Am J Prev Med.* 2000;18(1S):35-43.

19. Truman BI, Smith-Akin K, Hinman AR, Gebbie KM, Brownson R, Novick LF, et al. Developing the *Guide to Community Preventive Services*—Overview and Rationale. *Am J Prev Med.* 2000;18(1S):18-26.

20. Zaza S, Wright-de Aguero L, Briss PA, Truman BI, Hopkins DP, Hennessy MH, et al. Data collection instrument and procedure for systematic reviews in the *Guide to Community Preventive Services*. *Am J Prev Med*. 2000;18(1S):44-74.

21. Guide to Community Preventive Services. Systematic review methods. Available at: http://www.thecommunityguide.org/about/methods.html. Accessed March 24, 2012.

22. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA*. 1996;276(8):637-639.

23. Moher D, Jones A, Lepage L, for the CONSORT Group. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*. 2001;285:1992-1995.

24. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*. 2010;63:834-840.

25. Plint AC, Moher D, Morrison A, Schulz KF, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Australia.* 2006;185(5):263-267.

26. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet.* 1999;354(1896-1900).

27. Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 2009;339:b2535.

28. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA.* 2000;283(15):2008-2012.

29. Roundtree AK, Kallen MA, Lopez-Olivo MA, Kimmel B, Skidmore B, Ortiz Z, et al. Poor reporting of search strategy and conflict of interest in over 250 narrative and systematic reviews of two biologic agents in arthritis: a systematic review. *J Clin Epidemiol.* 2009;62:128-137.

30. Omenn GS, Clark NM. The *Guide to Community Preventive Services* will be influential in academic health centers: education, research and links with practice. *Am J Prev Med.* 2000;18(1S):12-14.

31. Briss PA, Brownson RC, Fielding JE, Zaza S. Developing and using the *Guide to Community Preventive Services*: lessons learned about evidence-based public health. *Annu Rev Publ Health.* 2004;25:281-302.

32. Kleinbaum DG, Klein M. Logistic regression: a self-learning text. 3rd ed. New York, NY: Springer; 2010.

33. Ndiaye SM, Hopkins DP, Shefer AM, Hinman AR, Briss PA, Rodewald L, et al. Interventions to improve influenza, pneumococcal polysaccharide, and hepatitis B vaccination coverage among high-risk adults: a systematic review. *Am J Prev Med.* 2005;28(5S):248-279.

34. Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: an update. *Mt Sinai J Med.* 1996;63(3):216-224.

35. Schriger DL, Altman DG, Vetter JA, Heafner T, Moher D. Forest plots in reports of systematic reviews: a cross-sectional study reviewing current practice. *Int J Epidemiol*. 2010;39:421-429.

36. Weller AC. Mounting evidence that librarians are essential for comprehensive literature searches for meta-analyses and Cochrane reports. *J Med Libr Assoc.* 2004;92(2):163-164.

37. Mallen C, Peat G, Croft P. Quality assessment of observational studies is not commonplace in systematic reviews. *J Clin Epidemiol.* 2006;59:765-769.

38. Baron RC, Rimer BK, Coates RJ, Kerner J, Mullen PD, Chattopadhyay S, et al. Methods for conducting systematic reviews of evidence on effectiveness and economic efficiency of interventions to increase screening for breast, cervical, and colorectal cancers. *Am J Prev Med*. 2008;35(1S):S26-S33.

39. Ndiaye SM, Hopkins DP, Smith SJ, Hinman AR, Briss PA, Task Force on Community Preventive Services. Methods for conducting systematic reviews of targeted vaccination strategies for *The Guide to Community Preventive Services*. *Am J Prev Med*. 2005;28(5S):238-247.

TABLES

Table 1. Multivariate polytomous logistic regression models, where the outcome variable was the rating of sufficiency of evidence issued by the Task Force†.

Predictor Variable(s)	Model	Covariates
Number of MOOSE† criteria addressed and Study quality score	A Full model	Year of publication Topic area of review Intervention type
	B	Year of publication Intervention type
	C	Year of publication
Number of MOOSE† criteria addressed	D Full model	Year of publication Topic area of review Intervention type
	E	Year of publication Intervention type
	F	Intervention type
	G	Year of publication
Study quality score	H Full model	Year of publication Topic area of review Intervention type
	I	Topic area of review Intervention type
	J	Intervention type

† Abbreviations: Task Force, Task Force on Community Preventive Services;
MOOSE, Meta-analysis of Observational Studies in Epidemiology reporting proposal

Table 2. Summary of the 72 selected Community Guide† systematic reviews.

Reference	Topic	Intervention Type	MOOSE† Score Range: 0 - 5	Study Quality Score‡	Task Force† Finding
Baron 2008	Cancer	Group or	2	10	Strong evidence
		Group or	1	2	Insufficient evidence
		Group or	3	11	Strong evidence
		Group or	2	9	Sufficient evidence
		Group or	1	1	Insufficient evidence
		Individual	2	16	Strong evidence
		Individual	2	12	Strong evidence
		Individual	2	4	Sufficient evidence
		Group or	0	2	Insufficient evidence
		Individual	2	22	Strong evidence
		Individual	2	14	Strong evidence
		Individual	1	7	Strong evidence
		Group or	0	8	Insufficient evidence
		Group or	0	2	Insufficient evidence
		Group or	0	2	Insufficient evidence
		Individual	2	30	Strong evidence
		Individual	2	7	Strong evidence
		Individual	0	2	Insufficient evidence
Baron 2010	Cancer	Individual	3	31	Strong evidence
Briss 2000	Vaccinations	Individual	2	45	Strong evidence
		Group or	2	-	Strong evidence
		Policy or law	1	-	Sufficient evidence
		Group or	1	1	Insufficient evidence
		Individual	1	-	Insufficient evidence
		Individual	2	3	Insufficient evidence
		Individual	0	-	Insufficient evidence

Table continues on next page.

Table 2, cont. Summary of the 72 selected Community Guide† systematic reviews.

Reference	Topic	Intervention Type	MOOSE† Score Range: 0 - 5	Study Quality Score‡ Range: 1 - 45	Task Force† Finding
Briss 2000, *cont.*	Vaccinations	Group or community	0	-	Strong evidence
		Group or community	1	-	Strong evidence
		Group or community	0	-	Sufficient evidence
		Individual	1	-	Sufficient evidence
		Group or community	0	-	Insufficient evidence
		Individual	1	-	Strong evidence
		Individual	1	-	Strong evidence
		Policy or law	1	-	Strong evidence
		Individual	0	-	Insufficient evidence
Campbell 2009	Alcohol	Policy or law	2	44	Sufficient evidence
Ditter 2005	Motor vehicle injury	Group or community	1	-	Insufficient evidence
		Individual	2	-	Insufficient evidence
Elder 2004	Motor vehicle	Group or community	3	-	Strong evidence
Elder 2010	Alcohol	Policy or law	3	-	Strong evidence
Hahn 2010	Alcohol	Policy or law	3	10	Sufficient evidence
		Policy or law	4	6	Insufficient evidence
Hopkins 2001	Tobacco	Policy or law	3	15	Strong evidence
		Group or community	1	-	Insufficient evidence
		Policy or law	2	-	Strong evidence
		Group or community	1	-	Strong evidence
		Policy or law	0	-	Strong evidence
		Group or community	1	-	Strong evidence
		Group or community	1	-	Insufficient evidence
		Group or community	0	-	Insufficient evidence

Table continues on next page.

Table 2, cont. Summary of the 72 selected Community Guide† systematic reviews.

Reference	Topic	Intervention Type	MOOSE† Score Range: 0 - 5	Study Quality Score‡ Range: 1 - 45	Task Force† Finding
Hopkins 2001, *cont.*	Tobacco	Individual	2	-	Sufficient evidence
		Individual	2	-	Insufficient evidence
		Individual	2	-	Strong evidence
		Individual	0	-	Insufficient evidence
		Group or community	1	5	Sufficient evidence
		Individual	2	-	Strong evidence
Hopkins 2010	Tobacco	Policy or law	5	35	Sufficient evidence
Leeks 2010	Tobacco	Group or community	4	15	Strong evidence
Ndiaye 2005	Vaccinations	Individual	2	2	Insufficient evidence
		Individual	2	1	Insufficient evidence
		Individual	2	1	Insufficient evidence
		Individual	3	7	Strong evidence
		Individual	2	1	Insufficient evidence
		N/A	5	24	Strong evidence
Sabatino 2008	Cancer	Individual	3	10	Sufficient evidence
		Individual	0	3	Insufficient evidence
Shults 2009	Motor vehicle	Group or community	3	6	Strong evidence
Zaza 2001	Motor vehicle injury	Policy or law	3	-	Strong evidence
		Group or community	3	-	Sufficient evidence
		Individual	4	-	Strong evidence
		Individual	2	-	Sufficient evidence
		Individual	4	-	Insufficient evidence

† Abbreviations: Community Guide, *Guide to Community Preventive Services*; MOOSE, Meta-analysis of Observational Studies in Epidemiology reporting proposal; Task Force, Task Force on Community Preventive Services

‡ Study quality score was calculated only for reviews that reported the number of studies with each quality rating.

Table 3. Descriptive characteristics of the Community Guide† systematic reviews (*n* = 72) included in this analysis.

Characteristic	*n* (%)	MOOSE† Score Mean (SD)	Study Quality Score Mean (SD)
Publication Year			
2000-2001	35 (48.6)	1.37 (1.11)	1.97 (7.95)
2004-2005	9 (12.5)	2.44 (1.13)	4.00 (7.81)
2008-2010	28 (38.9)	1.93 (1.36)	11.46 (11.27)
Topic Area			
Alcohol	4 (5.6)	3.00 (0.82)	15.00 (19.77)
Cancer	21 (29.2)	1.43 (1.08)	9.76 (8.80)
Motor vehicle injury	9 (12.5)	2.78 (0.97)	0.67 (2.00)
Tobacco	16 (22.2)	1.69 (1.40)	4.38 (9.64)
Vaccinations	22 (30.6)	1.36 (1.18)	3.86 (10.55)
Intervention Type (*n* = 71‡)			
Individual	34 (47.9)	1.76 (1.02)	6.41 (10.82)
Group or community	26 (36.6)	1.23 (1.18)	2.85 (4.29)
Policy or law	11 (15.5)	2.45 (1.44)	10.00 (15.56)

† Abbreviations: Community Guide, *Guide to Community Preventive Services*; MOOSE, Meta-analysis of Observational Studies in Epidemiology reporting proposal

‡ One review evaluated a combination of interventions including all three intervention types.

Table 4. Effect estimates for univariate associations, assessed independently, with the Task Force† findings regarding sufficiency of evidence in 72 systematic reviews conducted by Community Guide†.

Independent Variable	Regression Model	OR†	95% CI†	p-value
Study quality score	Polytomous	**Strong evidence: 2.33**	**(1.31 , 4.12)**	**0.0038***
		Sufficient evidence: 2.32	**(1.31 , 4.12)**	**0.0040***
Number of MOOSE† criteria addressed	Polytomous	**Strong evidence: 2.31**	**(1.35, 3.94)**	**0.0021***
		Sufficient evidence: 2.24	**(1.19, 4.21)**	**0.0126***
Methodological characteristics				
Types of study designs used	Ordinal	1.62	(0.66, 3.95)	0.2900
Use of librarian in search	Ordinal	1.20	(0.39, 3.67)	0.7576
Assessment of heterogeneity	Polytomous	**Strong evidence: 13.20**	**(3.75, 46.43)**	**< 0.0001***
		Sufficient evidence: 7.04	**(1.60, 30.92)**	**0.0098***
Forest plot graphic of effect estimates	Polytomous	**Strong evidence: 11.69**	**(2.91, 47.05)**	**0.0005***
		Sufficient evidence: 6.86	**(1.36, 34.71)**	**0.0200***
Alternative explanations for results	Ordinal	0.49	(0.17, 1.37)	0.1716

† Abbreviations: Task Force, Task Force on Community Preventive Services; Community Guide, *Guide to Community Preventive Services*; OR, odds ratio; CI, confidence interval; MOOSE, Meta-analysis of Observational Studies in Epidemiology reporting proposal

*** Indicates statistical significance at a 95% significance level.**

Table 5. Effect estimates for significant multivariate associations with the Task Force† findings of sufficiency of evidence in the 72 selected Community Guide† reviews. All models were polytomous logistic regression models, where the finding of insufficient evidence was the reference category.

Independent Variable(s)	Covariates	Adjusted OR†	95% CI†	p-value
Study quality score MOOSE† score	Year of publication (Model C)	<u>Study quality score</u> **Strong evidence: 1.49** **Sufficient evidence: 1.49** <u>MOOSE score</u> **Strong evidence: 2.09** Sufficient evidence: 1.96	 **(1.10, 2.03)** **(1.10, 2.03)** **(1.12, 3.92)** (0.94, 4.09)	 **0.0101*** **0.0102*** **0.0210*** 0.0742
Study quality score	None	Same as univariate model containing this variable only.		
MOOSE† score	None	Same as univariate model containing this variable only.		

† Abbreviations: Task Force, Task Force on Community Preventive Services; Community Guide, *Guide to Community Preventive Services*; OR, odds ratio; CI, confidence interval; MOOSE, Meta-analysis of Observational Studies in Epidemiology reporting proposal

*** Indicates statistical significance at a 95% significance level.**

Table 6. Odds ratio estimates and 95% CI† for univariate associations of three independent variables (year of publication, intervention type, and topic area of the review) describing Community Guide† systematic reviews, where the reporting of the individual MOOSE† methodological characteristics is the outcome.

	Methodological Characteristics				
Independent Variable(s)	Types of study designs	Use of librarian for search	Assessment of heterogeneity	Forest plot graphic of effect estimates	Alternative explanations for results
Year of publication					
2000-2001	0.87 (0.34, 2.24) p = 0.7676	0.60 (0.18, 2.06) p = 0.4215	1.25 (0.50, 3.17) p = 0.6326	**0.19** **(0.07, 0.55)** **p = 0.0020***	**0.21** **(0.06, 0.73)** **p = 0.0138***
2004-2005	OR† > 999.99 (<0.001, >999.99) p = 0.9562	**15.99** **(3.25, 78.70)** **p = 0.0006***	0.23 (0.04, 1.19) p = 0.0792	0.76 (0.17, 3.32) p = 0.7154	0.84 (0.16, 4.46) p = 0.8372
2008-2010	**0.27** **(0.09, 0.80)** **p = 0.0183***	0.23 (0.05, 1.13) p = 0.0710	1.46 (0.56, 3.79) p = 0.4366	**6.12** **(2.15, 17.42)** **p = 0.0007***	**4.75** **(1.52, 14.87)** **p = 0.0074***
Intervention type					
Individual	0.95 (0.37, 2.45) p = 0.9143	2.03 (0.59, 6.95) p = 0.2589	1.41 (0.56, 3.57) p = 0.4711	2.45 (0.93, 6.48) p = 0.0700	**0.15** **(0.04 0.57)** **p = 0.0057***
Group or community	**0.33** **(0.11, 0.96)** **p = 0.0427***	0.27 (0.05, 1.31) p = 0.1026	**0.34** **(0.13, 0.93)** **p = 0.0350***	**0.24** **(0.08, 0.74)** **p = 0.0131***	2.97 (0.99, 8.90) p = 0.0521
Policy or law	**5.47** **(1.31, 22.86)** **p = 0.0199***	1.01 (0.19, 5.34) p = 0.9905	2.94 (0.71, 12.16) p = 0.1360	1.64 (0.38, 5.03) p = 0.6284	1.92 (0.49, 7.52) p = 0.3498

Table continues on next page.

Table 6, cont. Odds ratio estimates and 95% CI† for univariate associations of three independent variables (year of publication, intervention type, and topic area of the review) describing Community Guide† systematic reviews, where the reporting of the individual MOOSE† methodological characteristics is the outcome.

	Methodological Characteristics				
Independent Variable(s)	Types of study designs	Use of librarian for search	Assessment of heterogeneity	Forest plot graphic of effect estimates	Alternative explanations for results
Topic area of review					
Alcohol	5.16 (0.51, 52.29) p = 0.1650	OR < 0.001 (<0.001, >999.99) p = 0.9804	3.00 (0.30, 30.30) p = 0.4711	5.16 (0.51, 52.29) p = 0.1650	**10.60** **(1.03, 109.45)** **p = 0.0475***
Cancer	**0.04** **(0.01, 0.36)** **p = 0.0034***	OR < 0.001 (<0.001, >999.99) p = 0.9538	0.81 (0.29, 2.24) p = 0.6816	**3.90** **(1.34, 11.33)** **p = 0.0124***	1.30 (0.41, 4.09) p = 0.6538
Motor Vehicle Injury	OR > 999.99 (<0.001, >999.99) p = 0.9562	**8.59** **(1.90, 38.88)** **p = 0.0052***	2.06 (0.47, 8.99) p = 0.3343	0.17 (0.02, 1.41) p = 0.1005	2.80 (0.66, 11.85) p = 0.1619
Tobacco	0.44 (0.13, 1.55) p = 0.2033	0.58 (0.12, 2.96) p = 0.5162	**3.72** **(1.07, 12.96)** **p = 0.0391***	1.29 (0.42, 4.00) p = 0.6515	0.36 (0.07, 1.75) p = 0.2047
Vaccinations	1.94 (0.70, 5.38) p = 0.2026	2.30 (0.67, 7.90) p = 0.1844	**0.23** **(0.08, 0.69)** **p = 0.0087***	**0.24** **(0.07, 0.81)** **p = 0.0218***	0.37 (0.10, 1.44) p = 0.1500

† Abbreviations: CI, confidence interval; Community Guide, *Guide to Community Preventive Services*; MOOSE, Meta-analysis of Observational Studies in Epidemiology reporting proposal; OR, odds ratio

*** Indicates statistical significance at a 95% significance level.**

FIGURES

Figure 1. Flow diagram showing the progress of studies through the stages of a systematic review, adapted from the QUOROM (Quality of Reporting of Meta-analyses) statement (26).

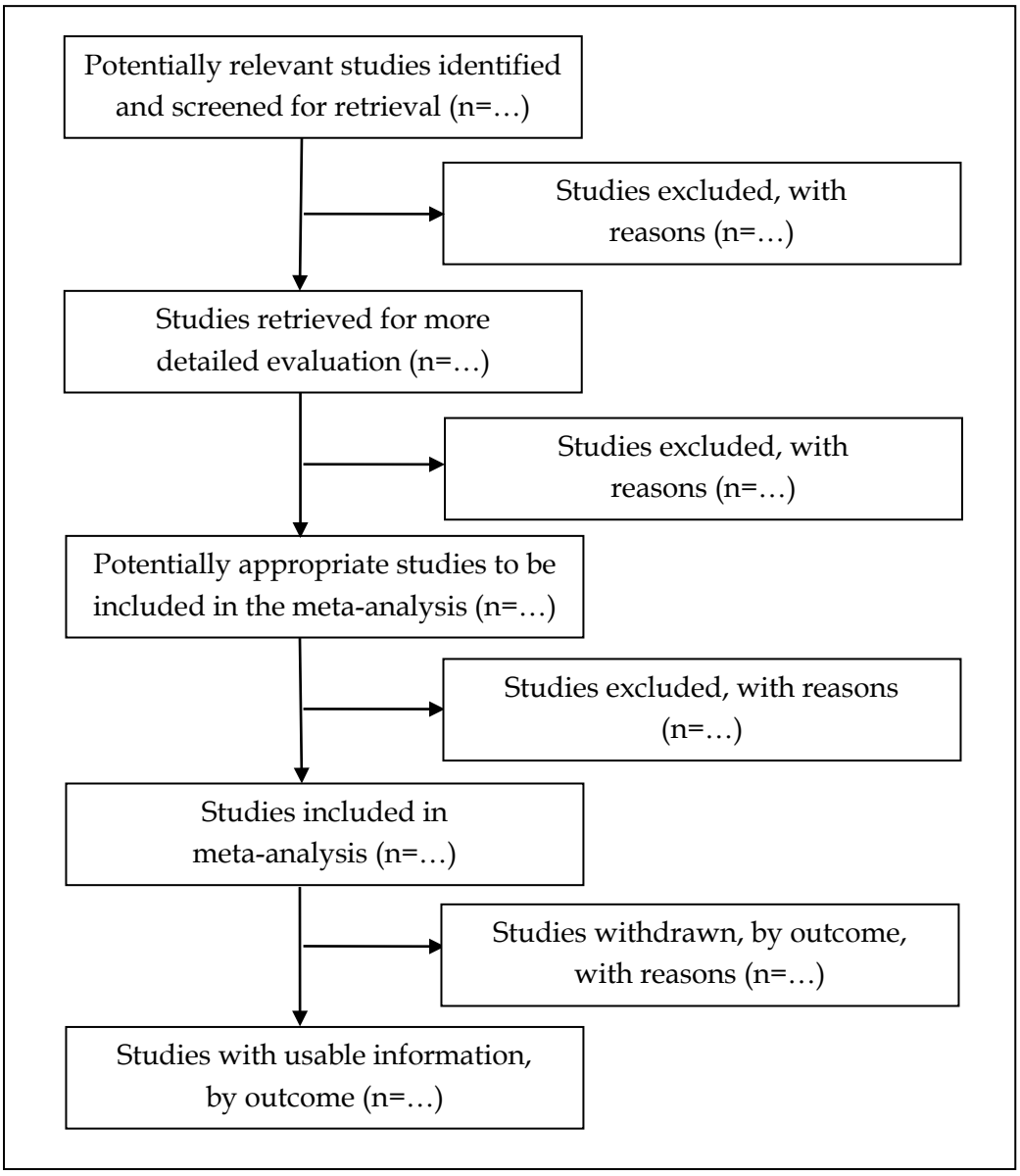
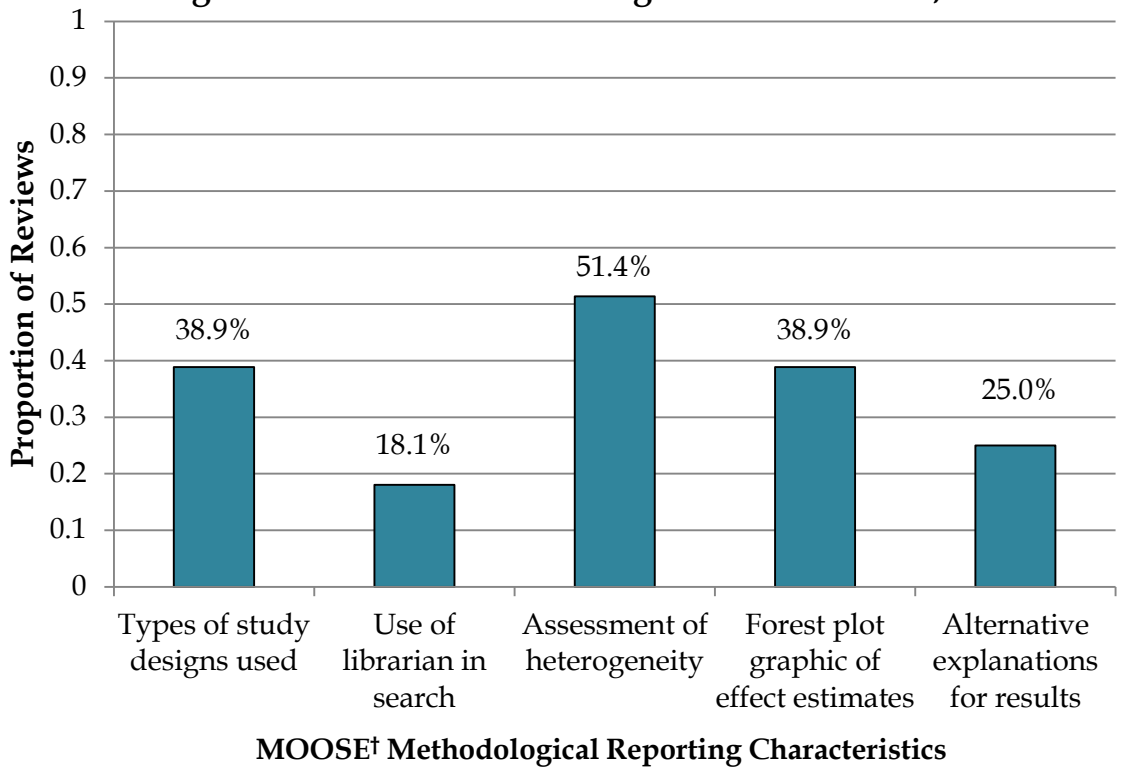


Figure 2. Methodological reporting checklist developed by the MOOSE (Meta-analyses of Observational Studies in Epidemiology) workgroup (28).

Table. A Proposed Reporting Checklist for Authors, Editors, and Reviewers of Meta-analyses of Observational Studies

- Reporting of background should include
  - Problem definition
  - Hypothesis statement
  - Description of study outcome(s)
  - Type of exposure or intervention used
  - Type of study designs used
  - Study population
- Reporting of search strategy should include
  - Qualifications of searchers (eg, librarians and investigators)
  - Search strategy, including time period included in the synthesis and keywords
  - Effort to include all available studies, including contact with authors
  - Databases and registries searched
  - Search software used, name and version, including special features used (eg, explosion)
  - Use of hand searching (eg, reference lists of obtained articles)
  - List of citations located and those excluded, including justification
  - Method of addressing articles published in languages other than English
  - Method of handling abstracts and unpublished studies
  - Description of any contact with authors
- Reporting of methods should include
  - Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested
  - Rationale for the selection and coding of data (eg, sound clinical principles or convenience)
  - Documentation of how data were classified and coded (eg, multiple raters, blinding, and interrater reliability)
  - Assessment of confounding (eg, comparability of cases and controls in studies where appropriate)
  - Assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results
  - Assessment of heterogeneity
  - Description of statistical methods (eg, complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient detail to be replicated
  - Provision of appropriate tables and graphics
- Reporting of results should include
  - Graphic summarizing individual study estimates and overall estimate
  - Table giving descriptive information for each study included
  - Results of sensitivity testing (eg, subgroup analysis)
  - Indication of statistical uncertainty of findings
- Reporting of discussion should include
  - Quantitative assessment of bias (eg, publication bias)
  - Justification for exclusion (eg, exclusion of non–English-language citations)
  - Assessment of quality of included studies
- Reporting of conclusions should include
  - Consideration of alternative explanations for observed results
  - Generalization of the conclusions (ie, appropriate for the data presented and within the domain of the literature review)
  - Guidelines for future research
  - Disclosure of funding source

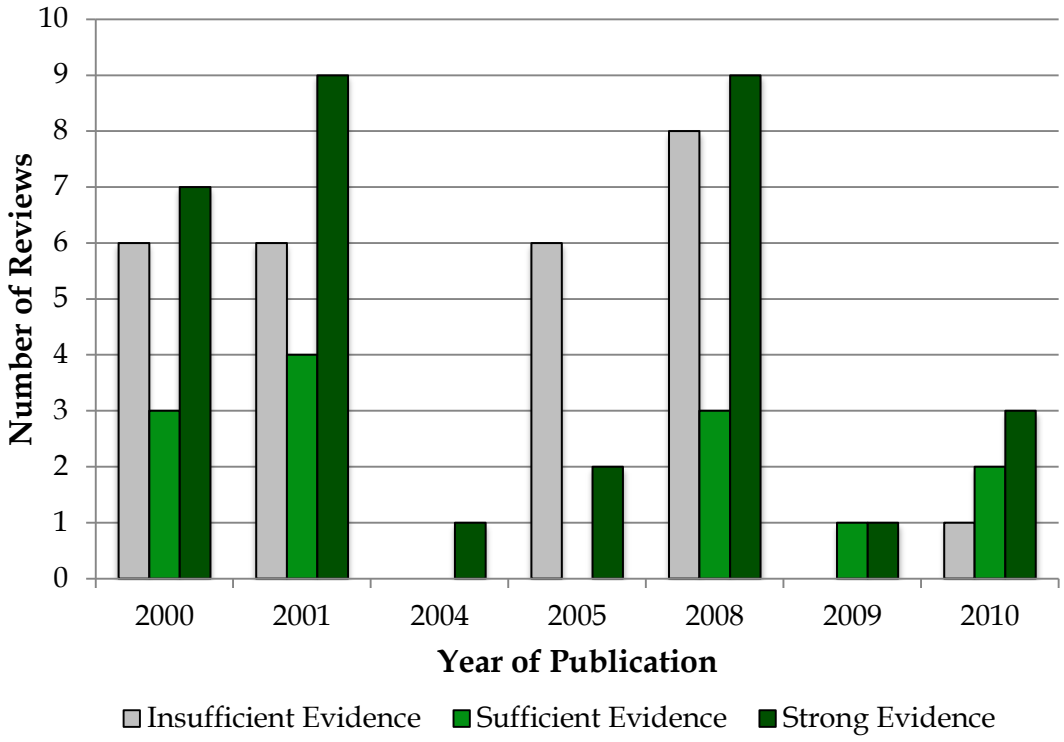
Figure 3. Proportion of the 72 Selected Community Guide* Reviews Addressing Each MOOSE† Methodological Characteristic, 2000-2010



* *Guide to Community Preventive Services*

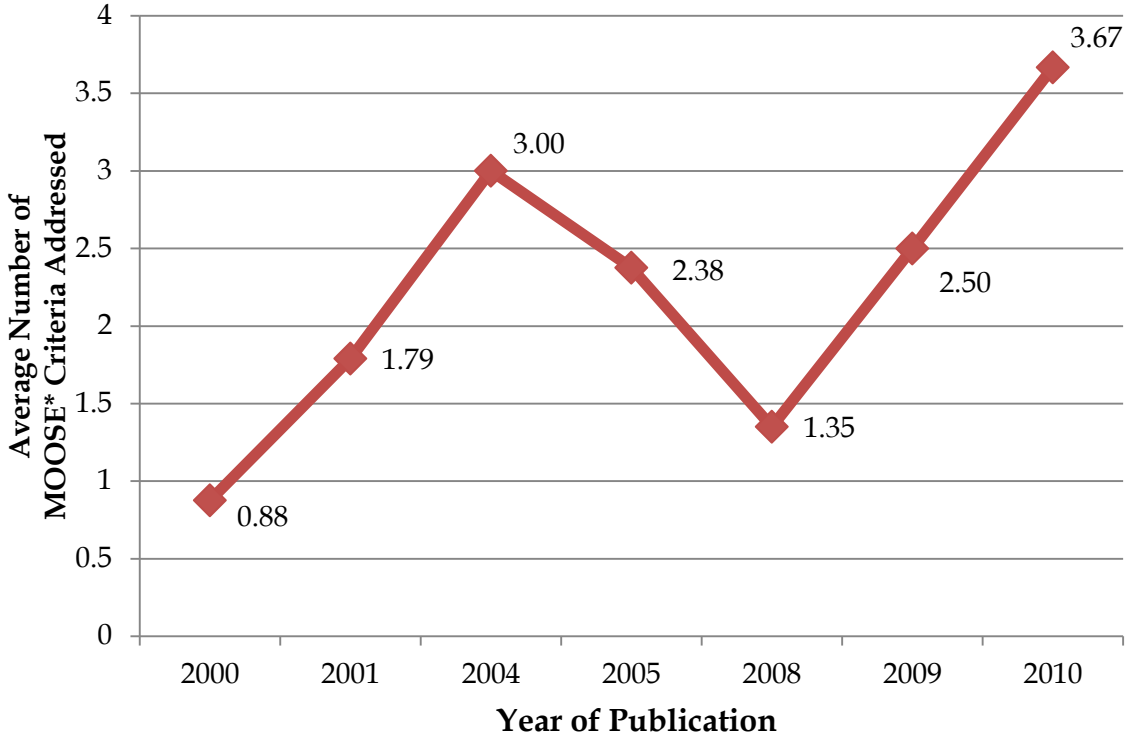
† Meta-analysis of Observational Studies in Epidemiology (MOOSE) reporting proposal

Figure 4. Task Force* Findings of Sufficiency of Evidence Issued to the 72 Selected Community Guide† Reviews by Year, 2000-2010



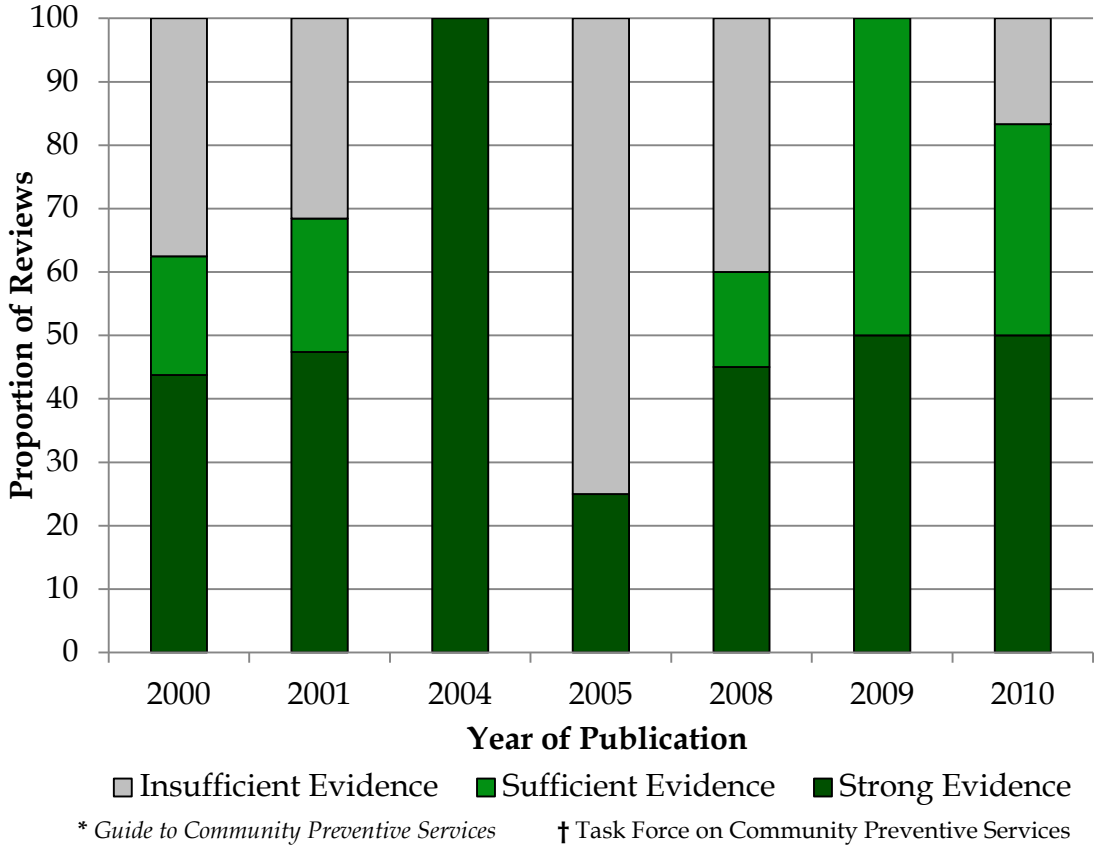
* Task Force on Community Preventive Services † *Guide to Community Preventive Services*

Figure 5. Average Number of Selected MOOSE* Criteria Addressed by the 72 Selected Community Guide† Reviews Each Year, 2000-2010



* Meta-analysis of Observational Studies in Epidemiology (MOOSE) reporting proposal.
† *Guide to Community Preventive Services*

Figure 6. Proportion of the 72 Selected Community Guide* Reviews by Year Receiving Each Task Force† Finding of Sufficiency of Evidence



* *Guide to Community Preventive Services* † Task Force on Community Preventive Services

CHAPTER III

PUBLIC HEALTH IMPLICATIONS

The results from our evaluation of the methodological reporting quality of selected systematic reviews conducted by the *Guide to Community Preventive Services* (Community Guide) revealed that there are many opportunities for improvement in the reporting of these reviews, based on the five methodological characteristics that were selected for evaluation. Over the last decade, the Community Guide has been a major contributor to the knowledge base that supports the transition within the public health community away from decision-making grounded solely on expert opinion or on professional consensus. Important advances have been made in research, policy, and programs to instead encourage the adoption of evidence-based recommendations in health care (30). In order for the Community Guide to continue to provide relevant and reliable information to support this new culture, it is important that the methodology used in the conduction of its systematic reviews is sound and that the descriptions of these methods are precise and easily reproducible.

Since the first publication of the Community Guide review methodology in 2000, the process by which its systematic reviews are conducted and reported has undergone several updates (18, 20, 31). For example, the methods used to conduct reviews of interventions to increase cancer screening were customized based on the topic area or the specific intervention being considered in the review. Therefore, these methods would have differed from those that were employed in reviews of targeted vaccination strategies (38, 39). As the procedures for conducting reviews continue to undergo

adjustments and improvements, the guidelines and policies for the reporting of these reviews should be considered as well.

A 2003 survey of state health departments indicated that recommendations from the Community Guide were being implemented in the modification of existing programs or in the development of new programs, but in less than half of the survey respondents (31). Determining quick and effective methods for the dissemination of these recommendations is an area of increasing importance for the future of the Community Guide. As the audience grows, so also will the expectation of confidence that the findings are valid, and this can be established through transparent and thorough reporting of the methods used in the conduction of the reviews. Such reporting should be consistent across all systematic reviews, regardless of the finding associated with the review. Taken together, all findings are of importance in public health, whether a recommendation with strong evidence supports the effectiveness of an intervention, or a finding of insufficient evidence suggests the need for future research.