Introduction

While literature reviews can vary in their purpose, focus or procedure, the general goal in any literature review is to gain a thorough understanding of prior research on the topic of interest. Typically, the reviewer is concerned with the methods employed in prior research, the findings produced and their limitations, and gaps in knowledge about the topic that remain. The information garnered through the review process is typically synthesized to arrive at an overall conclusion about the effectiveness of an intervention, or to help justify, frame and plan a new study of interest.

Given the importance of the literature review in the research process, JRSA is publishing a series of brief reports on state-of-the-art literature review practice. The series is designed to introduce researchers to issues that are germane to conducting a literature review that effectively meets the review's purpose.

This first report in the series focuses on a type of literature review that is performed to review a body of research for the purpose of arriving at an overall conclusion about the effectiveness of an intervention: the systematic review. It describes the basic process for conducting a systematic review in this context, differentiates systematic from narrative reviews, and highlights two common features of systematic reviews that are employed to enhance the trustworthiness of the review findings: methodological quality considerations and meta-analysis. While entry-level researchers and others who lack experience in conducting or using the findings of a systematic review are the primary audience for this report, more experienced researchers may benefit from the content as well.

Primary Purpose for Conducting a Systematic Review

The Systematic Review is one of the primary vehicles researchers use today to arrive at a bottom-line conclusion about the effectiveness of an intervention. While a systematic review shares common features with other literature review processes, a properly designed and executed systematic review is arguably unique in its capacity to objectively assess and synthesize disparate findings from many individual studies. As a result, the systematic review has become a staple of the evidence-based movement and the state-of-the-art process for reviewing and synthesizing a body of evidence in order to produce a trustworthy finding about whether an intervention works.

The Systematic Review Process

Like other literature review processes, a systematic review identifies, retrieves, appraises and synthesizes information on a specific topic from many empirical studies. In practice, studies to be potentially included in the review are located using methods such as internet searches, reviewing reference pages and bibliographies from both online and print documents,

This project was supported by Grant No. 2018-86-CX-K038 awarded by the Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the author and do not necessarily represent the official position or policies of the US Department of Justice.

¹Systematic review processes also are being used to review a body of research for the purpose of determining the association between two variables, such as economic insecurity and child maltreatment. See, for example, Jolliffe, D., Farrington, P., Piquero, A., Loeber, R., and Hill, K. (2017.) Systematic review of early risk factors for life-course-persistent, adolescence-limited, and late-onset offenders in prospective longitudinal studies. Aggression and Violent Behavior: 33: 15-23.
and conducting keyword searches of relevant abstract databases.² Systematic reviews, however, markedly differ from other literature review processes in that they adhere to a pre-established protocol regarding the selection of research studies for the review, and the manner in which studies are reviewed and findings from multiple studies are synthesized. Simply put, the parameters for conducting all aspects of a systematic review are set in advance and the assessment and synthesis methods employed are quantitative and based upon scientific criteria.

One of the most distinguishing features of a systematic review is the incorporation of methodological quality considerations. Indeed, there is widespread agreement in the scientific community that some studies are more trustworthy than others, and both the quality and consistency of the evidence must be considered to arrive at a valid conclusion about an intervention's effectiveness.

Methodological Quality and the Maryland Scientific Methods Scale (SMS)

The practice of incorporating methodological quality considerations in a literature review arguably can be traced to work conducted by Larry Sherman and his colleagues at the University of Maryland in developing the landmark 1997 report to Congress titled Preventing Crime: What Works, What Doesn’t, What's Promising. To objectively assess and account for variations in the scientific rigor of studies to be included in their review, Sherman and his colleagues developed the Maryland Scientific Methods Scale (SMS). The SMS was used to score and numerically rank the methodological quality of a study on several different dimensions, including the following:

- The scientific rigor of the study’s research design, and particularly the study’s ability to control for extraneous variables, eliminate rival hypotheses for study findings, and minimize measurement error.
- The study's statistical power to detect meaningful differences, and the adequacy of the study sample size.³
- Other considerations, such as attrition of cases from the study, the study’s response rate (if applicable), and the use of appropriate statistical tests.⁴

Using these criteria, the Maryland team rated each study being considered for their review. Generally, studies that received the highest ratings of scientific rigor received the greatest weight and consideration in the review. Additionally, to be considered effective or promising, a program must have been shown to be effective in at least one evaluation that employed some type of control or comparison group, with the preponderance of all other available evidence supporting the same conclusion.⁵

Methodological quality considerations patterned on the SMS are a standard feature of systematic reviews today. While synthesis research designed to determine whether an intervention works theoretically should examine the entire body of evaluation evidence on the intervention, conclusions drawn from a review of studies that lack scientific rigor are generally untrustworthy and prone to bias.⁶ Thus, contemporary systematic review protocols typically exclude studies that fail to reach a specified level of methodological rigor from further analysis, and they rely only on studies that employed well designed and executed experimental or quasi-experimental research designs to draw conclusions about what works.⁷

Meta-analysis

In recent years, more and more systematic reviews are incorporating a statistical procedure called meta-analysis to synthesize findings from multiple studies in an objective and quantitative manner. “In practice, meta-analysis combines the results of many evaluations into one large study with many subjects. This is important, because single studies based on a small number of subjects can produce distorted findings

---

² Such as the National Criminal Justice Reference Service, Criminal Justice Abstracts, Google Scholar, and JSTOR.
³ The smaller the anticipated effects of the intervention, the larger the sample size must be.
⁵ Ibid. Also, note that a higher standard of evidence was required for a program to be rated as “working” or effective. Programs deemed to be “working” were found to be effective in at least two evaluations that employed some type of control or comparison group, with the preponderance of all other available evidence supporting the same conclusion.
⁶ See for example, Rice & Harris (2003).
about a program’s effectiveness.

In addition, meta-analysis focuses on the direction and magnitude of program effects found across studies, and it produces a summary statistic called the average effect size. This statistic essentially indicates in a quantitative manner whether and to what degree the intervention is effective at producing a particular outcome, such as reducing recidivism, making it easier for the reviewer to synthesize the evidence from multiple, often disparate studies. For example, meta-analysis is especially useful in evaluations of recidivism reduction programs. In his 2008 report, Przybylski called attention to the various ways recidivism has been measured in the research (e.g., rearrest, return to prison) and the different follow-up periods across studies as methodological barriers that can be accounted for with meta-analysis. Unlike traditional literature review approaches where statistically significant findings are tallied and used to determine the proportion of studies that found a program effect, meta-analysis produces a quantitative measure that accounts for differences in measurement and more accurately summarizes the strength and consistency of findings from multiple studies.

**Differentiating the Systematic Review from the Narrative Review**

A systematic review of the literature fundamentally differs from a narrative review, the most common form of synthesis research years ago. Like any literature review, the narrative review process involves identifying and retrieving numerous individual studies, but the parameters of the review – including the criteria for including or excluding an individual study - are determined by the reviewer using professional judgement. The process for synthesizing information and drawing conclusions about the bottom-line effectiveness of an intervention is also based on the reviewer’s professional judgement. While narrative reviews provide a rudimentary mechanism for assessing the general quality and consistency of the research evidence on the effectiveness of an intervention, they have been criticized for their high level of subjectivity and lack of transparency.

**Summary/Closing**

The Systematic Review is widely used today to assess and synthesize a body of evidence on the effectiveness of an intervention. Within this context, the primary goal of the literature review is to arrive at a bottom-line conclusion about whether an intervention works. The average magnitude of an intervention’s effects is often determined through the review process too.

While a systematic review shares common features with other literature review processes, a properly designed and executed systematic review that is based on sound methodological quality considerations for individual studies (such as the SMS), objective exclusionary criteria, and meta-analysis techniques for the synthesis of findings from individual studies, arguably is unique in its capacity to sum up evidence and arrive at a trustworthy conclusion about the effectiveness of an intervention. As a result, systematic reviews have become a staple of the evidence-based move-

---

8 Ibid, pages 13-14. By pooling the original studies, meta-analysis counteracts a common methodological problem in evaluation research – small sample size – thereby helping the analyst draw more accurate and generalizable conclusions about an intervention’s effects. See, for example, Lipsey (2002).

9 An effect size can be calculated in a number of ways. See, for example, Lipsey and Wilson (2001). The odds-ratio effect size is commonly calculated when the outcome of interest is measured dichotomously. The mean difference effect size is commonly calculated when the outcome of interest is measured continuously.

10 Przybylski (2008), page 14.

11 Przybylski (2017).

12 It is important to note that the effectiveness of a well-defined criminal justice intervention can be determined through direct evaluation, provided the evaluation is sufficiently rigorous, and positive findings are replicated in subsequent evaluations of the intervention using different population samples. However, as Lipsey et al. (XXXX; pages 19-20) point out, this demonstration of effectiveness typically relies on only a few studies focused on a very specific and highly-protocolized (structured) program model, and the research demonstrating effectiveness is not generalizable beyond high-fidelity implementations of that model. Most interventions used in criminal or juvenile justice settings, however, are not so well-defined, but rather are locally developed or customized variants of generic program types, and a much larger body of research on the effectiveness of these intervention types tends to exist. This larger body evidence typically involves more variation in program and program setting characteristics, as well as more variation in study findings. Systematic reviews, particularly those that incorporate meta-analyses where appropriate, are particularly adept at synthesizing evidence when the knowledge base consists of disparate findings or more than a handful of studies.
ment. Systematic reviews, for example, are the primary mechanism for determining what works used by the Campbell Collaboration, a well-known and widely used source of information on the effectiveness of social and educational interventions.\(^\text{13}\)

A systematic review is more transparent, objective and rigorous than a narrative review, the primary approach used for reviewing and synthesizing evidence from multiple studies years ago. Because it adheres to a pre-established protocol for finding, assessing and synthesizing the results of many individual studies, a systematic review can be more easily assessed and/or replicated by other researchers. It’s pre-designed procedures, including the use of methodological quality considerations and meta-analysis when appropriate, also reduce bias and subjectivity, thereby enhancing the trustworthiness of the review’s findings.

References


\(^{13}\) The Campbell Collaboration is named after the late Donald Campbell, and esteemed social scientist considered by many to be a founding father of modern program evaluation who is arguably best known for his seminal texts on experimental and quasi-experimental design (see Petrosino, 2013; pages 9 and 11).