

Chapter 1

Background

This chapter describes the differing ways that meta-analyses and literature reviews summarize or integrate information from a collection of individual studies and how those respective methods shed light on somewhat different as well as overlapping aspects of intervention. The chapter also describes the methods used in this book's overview and integration of numerous meta-analyses and literature reviews, and how the authors of those works generally organized them by "type" of intervention. The chapter concludes by reviewing the main methods for measuring program effectiveness.

LITERATURE REVIEWS AND META-ANALYSES

Integrating Information

Literature reviews and meta-analyses provide different but complementary ways of integrating information and drawing conclusions. If one asks, Does intervention X work?, the conclusion that is usually drawn in a literature review has been largely based on an integration of all-or-none judgments about each of the several individual studies that comprise X. For example, say that ten separate family intervention programs have each been categorically judged to be effective or "working," based on their having reached the .05 level of statistical significance, and say that ten other such programs have been categorized as "not working" because they did not attain that level (even though several may have had a mildly positive outcome, such as $p < .20$). By then tallying—in this respect integrating—the categorical judgments regarding all twenty programs, it might reasonably be concluded that X yielded mixed results, that is, some positive and some negative, but did not reliably or even usually "work." (Throughout this volume "the .05 level" will usually be presented as " $p < .05$." In either case it will mean there were five or fewer chances in 100 that the results of a study could be accounted for by chance alone.)

In contrast, meta-analysis focuses more directly and primarily on the *group* or collection of studies. Its conclusion about the group's effectiveness is not derived from all-or-none judgments as to whether each individual study has reached a particular significance level; instead, it reflects the *degree* of difference that exists between each treatment or experimental program, on the one hand, and its control or comparison program, on the other (Light and Pillemer, 1984).

The statistic that is used to reflect this difference is called the "effect size" (ES). This statistic indicates the number of standard deviation units by which one program outperforms the other. (An ES of .10 corresponds to a difference of about 10 percent in the recidivism rates of the experimental as compared to the control/traditional programs, and ESs of .20 and .30 correspond to differences of approximately 20 percent and 30 percent respectively. This relatively straight-line relationship also exists whether one uses, as a *baseline*, a control/traditional-program recidivism rate that is 40 percent, 50 percent, or even 60 percent. However, the straight-line relationship disappears when, for example, one's control/traditional baseline drops well *below* 40 percent, especially when one's ES is .40 or more [Cohen, 1988, p. 181].) An *average* ES can then be obtained for the group of individual programs that was studied collectively. As a result, by focusing on the same twenty studies mentioned above, meta-analysis could have possibly justified a conclusion that intervention X was moderately effective, that is, generally positive. This conclusion could have reflected the finding that X's programs, collectively, had reduced recidivism by, say, an average of 12 percent and that the variability across its studies was not very large (thereby reducing the likelihood that the difference was due to chance). Such a conclusion could have been drawn even if no individual study had reached the .05 level *per se* (Fiske, 1983; Glass, McGaw, and Smith, 1981; Hedges and Olkin, 1985; Leviton and Cook, 1981; Lipsey, 1992; Strube and Hartmann, 1983; Wolf, 1986). This outcome could have occurred independently of the fact that—not infrequently—individual studies miss the .05 level (a) mainly because the sample-sizes of their respective experimental and control/traditional programs are relatively small (e.g., 50 each), (b) despite the fact that a sizable difference may simultaneously exist between the recidivism rates of those experimental and control/traditional programs.

Meta-analysis proceeds from the valid premise that the significance level which is ordinarily obtained for any individual study is produced not only by differences that may *actually* exist between the experimental and control programs that have been compared (e.g., differences in recidivism), but by several extraneous factors as well. The latter—often called

“noise” or “error variance”—includes both sampling and measurement error; together with other extraneous factors and with spin-offs from methodological shortcomings, they create erroneous variability among the studies being examined. Meta-analysis recognizes and tries to deal with the fact that results from individual studies, for instance, results which have traditionally been expressed as “significance levels,” are probabilistically distributed and contain an unknown degree of error—therefore unreliability.

To increase reliability, meta-analysis depends essentially on across-study statistics—central tendency and variability—and then draws its conclusion directly from the entire set of studies. Though meta-analysis utilizes not only large but also relatively *small* differences or fluctuations that exist between given experimental programs and their controls (or other types of comparisons), the conclusion derived via this method is—overall—less influenced by the above-mentioned error variance than is the conclusion from a traditional literature review. In traditional reviews, this variance, especially if substantial, can sometimes make a program either reach or miss the .05 level when, in fact, the program should not—or not quite—have done so.

Presently, many researchers and academicians favor meta-analysis over literature reviews. Besides the above reasons, this preference especially reflects the fact that meta-analysis—via its “effect size” statistic—can more fully and directly reflect the *amount* of difference that exists between intervention programs and their respective controls, not just whether a statistically significant difference exists for a certain percentage of studies. In addition, meta-analysis can aggregate the findings from several individual studies on a quantitatively more refined basis.

Literature reviews, however, can more easily reveal and descriptively focus on particularly promising individual programs. This feature is important if one asks, not just (a) Does intervention X work? or How much, if any, does it improve performance? but (b) What elements and other factors characterize programs that seem to work? and, more specifically, What perhaps contributes to the latter’s apparent success? Though meta-analysis can itself provide clues regarding this second set of questions, literature reviews are better-suited for discussing the set, including possible clues to success. (Of course, few widely recognized answers yet exist to the latter questions.) In any event, both types of questions are important today. Also, error variance notwithstanding, literature reviews, we believe, can probably provide a reasonably reliable, straightforward answer to the practical questions, Did specified *individual* programs improve offenders’ performance? and, How often did certain *types* of

programs, such as those emphasizing family intervention or vocational training, provide improvement?

In sum, the meta-analysis and literature review methods each shed light on somewhat different aspects or levels of correctional intervention; they may also cast somewhat different light on similar aspects. As a result, each has its advantages relative to integrating given information and addressing various questions about individual programs or groups of programs. In addition, some overlap exists as to what these methods can produce. For instance, both can address simple but important questions about public protection, for example, by indicating the average amount of recidivism reduction observed in specified types of programs. Meta-analysis can do this by statistically transforming its effect sizes into equivalent percent-reductions (Lipsey, 1991; Cohen, 1988), and literature reviews can usually generate such reductions directly.

Though neither meta-analysis nor literature review constitutes the only valid or complete approach to describing and assessing correctional intervention or the only promising road to knowledge, each one has considerable merit. In any event it seems unlikely that either method necessarily or commonly contains so much error or uncertainty that its results should, *a priori*, be considered seriously flawed and therefore of questionable validity. Both methods have made useful, complementary contributions to the evaluation of intervention; and insofar as their respective results suggest similar or substantially overlapping conclusions, those conclusions should probably be considered more reliable than if they were based on either method alone.

METHODS

To obtain a broad overview of correctional research findings we examined thirty-two meta-analyses and literature reviews—the two basic forms of study conducted to date. Specifically, we reviewed the results from all delinquency centered American and Canadian meta-analyses that have been published thus far, namely: Andrews et al., 1990; Davidson et al., 1984; Garrett, 1985; Gensheimer et al., 1986; Gottschalk et al., 1987; Izzo and Ross, 1990; Lipsey, 1992; Mayer et al., 1986; and Whitehead and Lab, 1989.¹ In addition, we examined the findings from some fifteen general, that is, multitopic, literature reviews. (Multitopic means that several types of intervention were covered.) These were: Gendreau and Ross, 1979, 1987; Genevie, Margolies, and Muhlin, 1986; Gordon and Arbuthnot, 1987; Greenberg, 1977; Johns and Wallach, 1981; Lab and Whitehead, 1988; Lipton, Martinson, and Wilks, 1975; Lundman, 1984;

Panizzon, Olson-Raymer, and Guerra, 1991; Romig, 1978; Rutter and Giller, 1983; Van Voorhis, 1987; Whitehead and Lab, 1989; and Wright and Dixon, 1977. (Martinson, 1974; and Palmer, 1975, 1978, 1983, 1984 were included, although their observations often related to the Lipton, Martinson, and Wilks study sample.)

While the sample of general literature reviews may or may not have included all or almost all multitopic yet large-scale works, it did, in the aggregate, encompass a very wide range of views and outcomes published from 1955 to the present—quite possibly the full range. Though the sample emphasized but was not limited to reviews that have been perhaps the most widely and often quoted, the general literature reviews were otherwise unselected and every effort was made to represent their full range of findings and conclusions—whether “positive,” “neutral or mixed,” or “negative.” In any case, the sample of multitopic literature reviews was—when aggregated—very inclusive as to settings, types of offenders, and time periods.

Finally, we examined the results from eight special-topic reviews, that is, literature surveys that focused mainly or exclusively on a single approach or intervention (such as family intervention). These were: Altschuler and Armstrong, 1990; Armstrong, 1988; Brody, 1976; Geismar and Wood, 1986; Graziano and Mooney, 1984; Krisberg et al., 1989; Parent, 1989; Schneider, 1986; and Ervin and Schneider, 1990. Many of the above-mentioned, *general* literature reviews had included the single approach or intervention that was focused on in these respective, special-topic reviews. The general reviews, of course, each included several other approaches as well.

Individually, the above-mentioned meta-analyses and literature reviews focused entirely or primarily on adjudicated youths—individuals in institutional settings, in community settings, or both. Each analysis and review purposely selected and emphasized individual studies in which a treatment or experimental program—referred to as “E”—was compared with a traditional or control program, labeled “C.” Each program’s performance measure, that is, its outcome or effectiveness criterion with respect to recidivism (see “Measuring Effectiveness,” which follows), typically involved a behavioral indicator such as arrests, convictions, detentions/incarcerations, or suspensions; however, revocation, unfavorable termination, or similar status criteria were not uncommon.

The authors of these meta-analyses and literature reviews generally organized them according to “type of intervention,” “method,” “general (or generic) approach,” and so forth—these terms being synonymous. In this context, “type” refers to the program feature, for example, vocational

training, that was used by an author (analyst or reviewer) to characterize the collection of experimental programs—that is, the treatments or interventions—that he or she had grouped together for purposes of analysis and discussion. For instance, a literature reviewer may have grouped fifteen separate studies that each emphasized vocational training and that differed from their control or comparison program in that respect. He or she may then have given those studies the generic label “vocational training” and may then have proceeded to determine whether—or, in the case of meta-analysts, to what degree—the various E programs within that group outperformed their respective C’s, or vice versa.

The types of intervention, that is, the subject areas that are included in the present overview of meta-analyses and literature reviews, are as follows: confrontation; area-wide strategies of delinquency prevention; social casework, social agency, or societal institution approaches to delinquency prevention; diversion; physical challenge; restitution; group counseling/therapy; individual counseling/therapy; family intervention; vocational training; employment; educational training; behavioral; cognitive-behavioral or cognitive; life skills (skill oriented; skill development); multimodal; probation enhancement and parole enhancement; intensive probation supervision; intensive aftercare (parole) supervision; and, community-based approaches vs. institutional intervention.

Not all individual meta-analyses and literature reviews touched on every subject area; in fact, many addressed fewer than half. Nevertheless, whenever the findings, the conclusions, or even the basic data-displays from any analysis or review bore on any given subject-area (e.g., confrontation), those findings, and so on, are mentioned in the results summarized in chapter 3 for that subject-area, that is, approach. More precisely, they are referred to in connection with the given approach *unless* the overall research quality of the individual studies which comprised that approach was very questionable. This occurred, for example, when the percentage of random-assignment studies and/or quasi-experimental studies that comprised the approach was either known to be, or seemed likely to have been, other than high. This situation, however, was uncommon across the analyses and reviews collectively.

MEASURING EFFECTIVENESS

The main indices of a program’s effectiveness should closely reflect the fact that intervention’s primary goal is increased public protection against illegal behavior. Such protection is directly reflected in *recidivism*, defined as any form of repeat offending. Recidivism, as measured by

arrests, parole revocation, incarceration, and so on, has long been used to assess the impact of rehabilitation, punishment, and incapacitation alike. Despite its complexities and the differing ways it is measured, this index is widely accepted by researchers, practitioners, policy makers, and the public itself, and it is usually considered a key element in any outcome evaluation (Maltz, 1984).

When evaluating rehabilitation in particular, additional measures are nevertheless common; sometimes, in fact, they appear instead of recidivism. These measures generally reflect intervention's secondary or offender-centered goal—one that involves attitude change, personality change, and skill development. Such indices often correlate with recidivism itself; beyond that, they are generally recognized as meaningful in their own right, even when the correlation is low. However, since intervention's primary goal—increased public protection—is society rather than offender centered, the main index of program effectiveness in this book will be considered the *reduction of recidivism*, especially as measured by arrests, convictions, and similar actions that reflect offender behavior. Without this index, program evaluation would not just be incomplete; it would miss the main point. This is independent of the fact that attitude change, skill development, and so on, doubtlessly contribute to recidivism reduction itself.