



Published in final edited form as:

*Adm Policy Ment Health*. 2011 January ; 38(1): 54–63. doi:10.1007/s10488-010-0315-y.

## Design Elements in Implementation Research: A Structured Review of Child Welfare and Child Mental Health Studies

**John Landsverk,**

Child and Adolescent Services Research Center (CASRC), Rady Children's Hospital San Diego, 3020 Children's Way, MC 5033, San Diego, CA 92123, USA

**C. Hendricks Brown,**

Department of Epidemiology and Public Health, School of Medicine, University of Miami, Miami, FL, USA

**Jennifer Rolls Reutz,**

Child and Adolescent Services Research Center (CASRC), Rady Children's Hospital San Diego, 3020 Children's Way, MC 5033, San Diego, CA 92123, USA

**Lawrence Palinkas,** and

School of Social Work, University of Southern California and Child and Adolescent Services, Los Angeles, CA, USA

**Sarah McCue Horwitz**

Department of Pediatrics and the Centers for Health Policy and Primary Care and Outcomes Research, Stanford University, Stanford, CA, USA

John Landsverk: [jlandsverk@casrc.org](mailto:jlandsverk@casrc.org)

### Abstract

Implementation science is an emerging field of research with considerable penetration in physical medicine and less in the fields of mental health and social services. There remains a lack of consensus on methodological approaches to the study of implementation processes and tests of implementation strategies. This paper addresses the need for methods development through a structured review that describes design elements in nine studies testing implementation strategies for evidence-based interventions addressing mental health problems of children in child welfare and child mental health settings. Randomized trial designs were dominant with considerable use of mixed method designs in the nine studies published since 2005. The findings are discussed in reference to the limitations of randomized designs in implementation science and the potential for use of alternative designs.

### Keywords

Child welfare; Child mental health; Implementation research; Research design

### Introduction

Dissemination and implementation research is moving toward the status of an emerging field characterized as implementation science, exemplified by the launching of the journal

Implementation Science in 2006, and the annual Dissemination and Implementation conferences sponsored by the Office of Behavioral and Social Science (OBSSR) in collaboration with other institutes and centers at the NIH that were initiated in 2007. Not surprisingly, progress in this emerging science is uneven, with a greater volume of empirical studies mounted in the physical health care field than in the fields of mental health or social services, due largely to medicine's focus first on quality of care and, more recently, on comparative effectiveness. The Cochrane Collaboration ([www.cochrane.org](http://www.cochrane.org)) has been an important resource for evaluating the evidence base for clinical interventions in medicine and, since 1994, has developed systematic reviews of studies of implementation strategies for health care settings through the Effective Practice and Organization of Care Group (EPOC, [www.epoc.cochrane.org](http://www.epoc.cochrane.org)). However, despite these important scientific efforts, there remains a lack of consensus on methodological approaches to the study of the implementation processes and tests of implementation strategies (Proctor et al. 2009). To begin to address this deficiency, this paper reviews design issues for implementation research and illustrates these issues through a structured review of implementation studies in U.S. child welfare and child mental health service systems.

A useful organizing heuristic is to conceptualize dissemination and implementation studies relative to two other phases of research, efficacy and effectiveness. Nicely captured in the 2009 National Research Council and Institute of Medicine report on *Preventing Mental, Emotional, and Behavioral Disorders Among Young People* (National Research Council and Institute of Medicine 2009, p. 326) and shown here in Fig. 1, dissemination and implementation (D&I) studies are the last stage of research in the science to practice continuum, preceded by effectiveness and efficacy studies that are distinct from and address different questions from D&I studies. The figure also demonstrates that distinct phases (albeit somewhat overlapping) exist within the D&I stage, which are characterized as adoption, sustainability, and moving to scale, similar to those proposed by Aarons et al. (in press). Given the very different purposes of efficacy, effectiveness, and D&I research as well as different purposes for phases within the D&I stage, it is likely that they may require different research designs or different emphases in the classic research design tension between internal and external validity. Indeed, two papers have argued that it is possible to combine several different designs into one overall study so that efficacy and implementation research questions can be addressed within a single study (Brown et al. 2009; Poduska et al. 2009).

This research model typology is also reflected in the NIH Roadmap initiative for re-engineering the clinical research enterprise that is currently driving the translational research initiative at the NIH (Culliton 2006; Zerhouni 2003, 2005, also see the website <http://nihroadmap.nih.gov>). The Roadmap initiative has identified three types of research leading to improvements in the public health of our nation, namely, basic research that informs the development of clinical interventions (e.g., biochemistry, neurosciences), treatment development that crafts the interventions and tests them in carefully controlled efficacy trials, and what has come to be known as service system and implementation research, where treatments and interventions are brought into and tested in usual care settings (for an interesting discussion of this interplay, see Westfall et al. 2007). Based on this tripartite division, the Roadmap further identified two translation steps that would be critical for moving from the findings of basic science to improvements in the quality of health care delivered in community clinical and other delivery settings. The first translation step brings together interdisciplinary teams which integrate the science work being done in the basic sciences and treatment development science, such as translating neuroscience and basic behavior research findings into new treatments. The focus of the second translation phase is to translate evidence-based treatments into service delivery settings and sectors in

local communities and it is this second step that we would identify as the D&I research enterprise.

Historically, basic science and treatment research as partners in the first translation step have relied heavily on what has come to be known as the “gold standard” of designs; namely, the randomized control trial (RCT) involving randomization at the person level. While effectiveness research also has been portrayed as part of the science to practice continuum, it has generated far less discussion and detailing. In one of the few comprehensive discussions of the distinction between efficacy and effectiveness trials, Flay (1986) notes that “whereas efficacy trials are concerned with testing whether a treatment or procedure does more good than harm when delivered under optimum conditions, effectiveness trials are concerned with testing whether a treatment does more good than harm when delivered via a real-world program” (Flay 1986, p. 455). Pertinent to this design paper, Flay also introduced the concept of “implementation effectiveness” trials and indicates such trials “can be uncontrolled in the sense that variations in program delivery are not controlled, or they can be controlled experiments in which different approaches to the delivery or dissemination of an efficacious technology are tested” (Flay 1986, p. 458).

This issue of control groups or conditions and use of randomized trials for studies about translating research to practice has been recently discussed by Glasgow et al. around the concept of practical clinical trials (2005) with a recommendation that such translation requires as much if not more attention to external validity as internal validity. While the authors acknowledge that RCT designs provide the highest level of evidence, they argue that “in many clinical and community settings, and especially in studies with underserved populations and low resource settings, randomization may not be feasible or acceptable” (Glasgow et al. 2005, p. 554) and propose alternatives to the randomized design such as “interrupted time series,” “multiple baseline across settings” or “regression-discontinuity” designs. Brown and colleagues have argued that the gap between some of these designs and true RCTs can be diminished by incorporating randomization across time and place instead of person-level randomization (Brown et al. 2006, 2008, 2009; National Research Council and Institute of Medicine/IOM 2009).

The use of classic RCT designs in the second translation step has been even more vigorously and provocatively addressed by Berwick, the well-known leader of the Institute for Healthcare Improvement (IHI). In a recent JAMA commentary, he argues there is an unhappy tension between “2 great streams of endeavor with little prospect for merging into a framework for conjoint action: improving clinical evidence and improving the process of care” (Berwick 2008, p. 1182). He proposes resolving this tension with a radical research design position (based in epistemology), namely, that the randomized clinical trial design—so important in evidence-based medicine—is not useful for studying health care improvement in complex social and organizational environments. Asserting the need for essentially new evaluation methods, he points to Pawson and Tilley’s (1997) position that the study of social change in complex systems requires a wider range of scientific methodologies, those that can capture change mechanisms within social and organizational contexts. Berwick extends this argument, suggesting that “many assessment techniques developed in engineering and used in quality improvement—i.e., statistical process control, time series analysis, simulations, and factorial experiments—have more power to inform about mechanisms and context than do RCTs, as do ethnography, anthropology, and other qualitative methods” (Berwick 2008, p. 1183).

In summary, a robust debate ongoing for D&I efforts focuses on whether randomized designs, found so useful for efficacy and effectiveness research, can or should be applied to D&I research. Even if one rejected Berwick’s conclusion that randomization designs are not

useful for testing improvement in complex environments, it is useful to address Berwick's challenge to search non-traditional sources for alternative designs in the third stage of research, namely dissemination and implementation.

In this paper, we take a first step in addressing this question by examining the use of randomization and other design elements in empirical studies that can be characterized as implementation studies. These are studies that test an implementation strategy and that take place in the service systems that would meet Glasgow and colleagues' definition of serving disadvantaged populations and also have low resources, namely child welfare and child mental health. The objective of this paper is to identify similarities and variation in design elements within dissemination and implementation research, drawing on approaches to studying social change processes in real world contexts.

## Methods

### Searches

The structured literature review focused on empirical research that addressed the implementation of an evidence-based program addressing mental health problems in children, either prevention or treatment, within a child welfare or mental health service system. The search parameters permitted another service system such as schools if the primary intervention was for mental health problems and was delivered by an extension of the child mental health system. Searches included peer-reviewed articles reporting on an implementation study, as defined below. The unit of analysis was the study since a single study could be described in more than one paper. Studies could include one or more stage of implementation as defined in this series (exploration, adoption/preparation, active implementation/or sustainment as described by Aarons et al. (in press)).

Databases searched were PubMed MEDLINE, PsychInfo, and Social Services (see Table 1 for websites). The authors considered including the NIH RePORTER database as well (formerly known as CRISP) to widen the search to funded implementation studies but determined the abstracts provided insufficient detail on design elements. "Gray" literature, such as government and foundation reports, was originally searched as well but was not included because almost no studies were found and the study designs results, if found in a report, were not usually presented in the format found in peer-reviewed journals. Further, these reports often had not been peer reviewed.

Search terms included combinations of the following: (1) implementation, (2) dissemination, (3) child or children, (4) child welfare, (5) foster care, (6) mental health, (7) welfare services, and (8) social services. MeSH terms were used when appropriate. Titles and abstracts of all potentially applicable sources were obtained and all items that appeared to meet the search criteria were obtained in full text.

Inclusion and exclusion criteria were used to determine which of the studies identified through our broad filters could be excluded from further consideration. The purpose of this structured review was to examine studies that tested a strategy or strategies for implementing an intervention that is considered as an "evidence-based" program. For this purpose, implementation strategies are considered "focal interventions" while the evidence-based program being implemented is considered the "target intervention". Thus, studies were required to involve both a focal intervention as well as a target intervention. The search only included empirical studies at any stage of implementation involving human subjects who are involved in the delivery of the intervention, thus excluding model based studies without data. In addition, studies were required to have some form of comparison design testing the implementation strategy and the design had to be longitudinal with at least two

data points. Finally, studies were included if the article reporting on the study was published in English from January, 1995 to April, 2010. Other specific exclusion criteria included: (1) case or descriptive studies, (2) physical health studies, and (3) implementation of policy or practice target intervention studies(s), and (4) substance use treatment or prevention studies. Papers reporting qualitative findings from a study were included if the study employed mixed methods with both qualitative and quantitative measures.

## Coding

Study design was coded using the categories established by the EPOC Review Group within the Cochrane Collaborative that has been included in the website (<http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/datacollectionchecklist.pdf>) as the document Data Collection Checklist. The EPOC four design categories are described below as presented in the EPOC website:

*Randomized controlled trial (RCT)* i.e. a trial in which the participants (or other units) were definitely assigned prospectively to one or two (or more) alternative forms of health care using a process of random allocation (e.g. random number generation, coin flips).

*Controlled clinical trial (CCT)* may be a trial in which participants (or other units) were: (a) definitely assigned prospectively to one or two (or more) alternative forms of intervention using a quasi-random allocation method (e.g. alternation, date of birth, patient identifier) or; (b) possibly assigned prospectively to one or two (or more) alternative forms of intervention using a process of random or quasi-random allocation.

*Controlled before and after study (CBA)* i.e. involvement of intervention and control groups other than by random process, and inclusion of baseline period of assessment of main outcomes. There are three minimum criteria for inclusion of CBAs: (a) *Contemporaneous data collection*: Pre and post intervention periods for study and control sites are the same. If it is not clear in the paper, e.g. dates of collection are not mentioned in the text, the paper should be discussed with the investigators before data extraction is undertaken. (b) *Appropriate choice of control site*: Studies using second site as controls: if study and control sites are comparable with respect to site characteristics, level of care and setting of care. If not clear from the paper whether study and control sites are comparable, the paper should be discussed with the investigators before data extraction is undertaken. (c) *Minimum number of sites*: a minimum of two intervention sites and two control sites.

*Interrupted time series (ITS)* i.e. a change in trend attributable to the intervention. There are two minimum criteria for inclusion of ITS designs: (a) clearly defined point in time when the intervention occurred and (b) at least three data points before and three after the intervention.

In addition to study design, studies were coded for study methods (quantitative only, mixed quantitative and qualitative), level at which randomization occurred (if appropriate), stage of implementation (see above), and what implementation theory or model was specified.

All searches were conducted by the third author with specification consultation from the other authors. Selection of papers for coding was conducted by the third and first author. All papers were coded independently by three of the five authors, with coding disagreements discussed via e-mail and resolved through consensus where possible.

## Results

As shown in Fig. 2, the structured literature search yielded 338 articles that were selected for full text review after exclusion for non-empirical articles and articles on either physical health or substance abuse interventions. Nine studies, represented by one paper each, met the full inclusion criteria. The majority of exclusions were for uncontrolled case studies or descriptive studies ( $N = 329$ ). This supports the general perception that the use of controlled designs testing implementation strategies was scarce for implementation research related to the two service systems of child welfare and child mental health. It also suggests that the terms “dissemination” and “implementation” are widely found in the published literature but are not specific in terms of locating empirical literature in the emerging implementation science, given the high volume of hits associated with these terms.

The review of articles/studies indicated that the development of controlled studies of implementation strategies is a very recent phenomenon. Despite use of a search timeframe that began with 1995, none of the articles meeting full inclusion criteria were published before 2005—in fact, eight of the nine coded articles were published in 2008–2009. Furthermore, included articles reported on the design and conceptualization of the study since the study quantitative data were not yet available (Glisson and Schoenwald 2005) or reported on a qualitative analysis that could be completed prior to the full completion of the study with published quantitative findings (Palinkas et al. 2008, 2009). These indicators all point to implementation research, as defined in this paper, being a very recent initiative in the empirical research in these two service sectors.

Examination of these early papers indicates that randomized trial designs dominated. Eight of the nine studies used a RCT as defined by EPOC review criteria while the remaining study employed a design with mixed features that coders categorized as a CCT by EPOC criteria (Herschell et al. 2009). The paper for that study also suggested that the researchers would have preferred a full RCT design but were constrained by the policy that was controlling the implementation up-take study. None of the nine studies indicated any partial use or consideration of a quasi-experimental design. In addition, the study papers were coded for the level at which randomization was executed since the conceptual frame of implementation as being a multi-level phenomenon is so pervasive (Aarons et al. in press). Only one of the nine studies reported randomization at two levels, namely the study of MST implementation linked to the ARC organizational development intervention (Glisson and Schoenwald 2005), where randomization was carried out at the county level for the ARC organizational implementation intervention and at the youth level for the selection of participants into the MST intervention. None of the other eight studies used more than one level for randomization. In general, randomization was carried out at the organization or geographic entity in five of nine studies with randomization at the individual or provider level in three of the studies.

The additional design feature coded in this study was whether the data collection design used a quantitative only or a mixed method design with both quantitative and qualitative data collection. Five of the nine studies used a quantitative only design while four of the nine used a mixed methods approach with both quantitative and qualitative methods for data collection.

As noted in the introduction, the lead article in this series proposes a conceptual model of evidence-based practice implementation with four categories; exploration, adoption/preparation, implementation, and sustainment. The studies were coded for which stage was addressed. As seen in Table 2, only one study addressed the exploration stage (Chamberlain et al. 2008), two involved the adoption-preparation stage (Barwick et al. 2009; Chamberlain

et al. 2008), all nine addressed the implementation stage, and one was seen as conducting research on the sustainment stage (Chamberlain et al. 2008). Only two of the studies addressed more than one stage with the Chamberlain study unusual for addressing all four stages in a single study.

Finally, as shown in Table 2, there were a range of implementation theories or models used for these nine studies. Two of the studies explicitly linked the research to Roger's well-known diffusion of innovations theory (Atkins et al. 2008;Palinkas et al. 2009). Glisson and Schoenwald (2005) specified the ARC implementation intervention at the organizational level that has supporting effectiveness data from prior studies. Three linked their study to implementation interventions that did not have prior research while the coders could find no explicit implementation theory or model in three of the studies.

## Discussion

The results from this structured review demonstrate that the very recent publication history of controlled studies identified in child mental health and child welfare in the search process clearly supports the continuing emergence of the implementation science field. Furthermore, implementation science in child mental health and child welfare systems is considerably less well developed compared to certain areas of evidence-based medicine as indicated by the dominance of medical and nursing implementations being examined in Implementation Science and other journals. Additionally, the small number of studies that have used controlled designs to test implementation strategies demonstrates the unevenness of the field as we contrast the number of studies related to child mental health and child welfare with the number of studies that have been systematically reviewed by the EPOC group within the Cochrane Collaborative. A review of the Cochrane Collaborative website as of June 30, 2010 found 66 systematic reviews and 26 protocols of reviews in progress by the EPOC review group, including 49 reviews of specific types of interventions, 34 of interventions to improve specific types of practice, and nine broad reviews or protocols. While this review points to the paucity of implementation research being carried out in child welfare and child mental health, it also suggests that the EPOC reviews may constitute a useful repository of information regarding implementation strategies and the use of a fuller range of design options, some of which may be transferable to the more nascent child welfare and child mental health areas.

The dominance of randomized designs in the nine studies coded for the structured review might be thought to be an outlier due to the constrained scope of the search. However, an examination by the authors of a 20% sample ( $N = 14$ ) of the 66 EPOC systematic reviews shows that RCT designs are also the majority in the studies included in the reviews, with 57% of the reviews (eight of 14 reviews examined) having every study meeting the review criteria as an RCT, 36% of the reviews (five of 14) with a mix of RCTs, CCTs, ITS, and CBA designs (*note*: in four of these five reviews, RCTs made up the majority of the study designs –78, 73, 66, and 60%), and the remaining review on interventions to improve antibiotic prescribing practices for hospital inpatients involved 60% ITS design studies, with only 15% of the studies being RCTs.

Neither the review for this paper nor the consideration of the EPOC systematic reviews suggests strong evidence for the position Berwick has laid out in his "The Science of Improvement" paper (Berwick 2008) or wide-spread use of quasi-experimental designs, such as those suggested in the Glasgow et al. "Practical clinical trials for translating research to practice" (Glasgow et al. 2005). However, we do not believe that the relatively minor role of non-randomized designs in implementation research to-date should preclude consideration of quasi-experimental and observational designs. We would argue that the

issue of external validity is not well addressed by standard randomized designs and we agree with Glasgow that external validity is critical for making the science to practice translation. In addition, randomized designs involving small numbers of units may still fail to control for unmeasured variation across the experimental and control groups. However, since implementation processes inherently take place in a multilevel context [see Shortell (2004) and the Aarons et al. (in press) paper in this series], and the nine studies in the structured review largely used randomization at an organizational or higher level, the problem of sufficient power to carry out randomized designs at these levels becomes problematic, both in terms of number of units available to achieve adequate power, and cost for carrying out large scale implementation trials. Useful analytic strategies for improving statistical power with small numbers of randomized units include the use of covariates at the level of randomization and the addition of multiple outcomes that can be used in growth modeling (Brown and Liao 1999).

We note the considerable use of mixed methods designs in the nine studies reviewed and suggest this is a useful trend to increase the context specificity of implementation studies as well as allow for convergence, complementarity, or expansion of findings obtained from the combination of qualitative and quantitative measures of implementation processes and outcomes. We would also argue that there have been recent developments in two design areas coming from the experimental tradition that warrant serious consideration for implementation study designs.

First, there is clear evidence that investigators of efficacy studies are increasingly considering employment of alternative randomized designs. Informed by the limitations of such studies (particularly for generalizability/external validity), this field is developing a range of adaptive (Brown et al. 2009), “encouragement,” (Duan et al. 2002), and “smart” designs (Lavori et al. 2000). Some of these designs, such as “roll-out” and “cumulative trial” designs provide additional precision by combining outcomes from replicate cohorts or related trials (Brown et al. 2009). Others have elements that, for implementation research, mimic the element of choice by consumers and providers in community service settings targeted for implementation of evidence-based practices. Over the past decade, efficacy researchers have developed an intriguing set of randomized designs that are considerably more complex than traditional RCTs, but also more sensitive to issues of external validity. For example, Duan and colleagues (Duan et al. 2002) have proposed a “randomized encouragement trial (RET)” as an extension of Zelen’s randomized consent design. Unlike RCT designs that randomize participants to mandated and possibly blinded treatment assignment, which suppresses consumers’ preferences and choices, RET randomizes consumers to encouragement strategies for the targeted treatment and facilitates their preferences and choices under naturalistic clinical practice settings. RET designs can also be characterized as a special case of what are termed “practical clinical trials” (previously known as pragmatic clinical trials) that have arisen out of the distinction between “explanatory and pragmatic attitudes in therapeutic trials” (Schwartz and Lellouch 2009). An influential 2003 JAMA article (Tunis et al. 2003) reviews the development of practical clinical trials and elaborates the needed infrastructure to support this important direction. The utility of practical clinical trials in psychiatry has been robustly demonstrated in the CATIE network studies of schizophrenia. However, cognizant of the high costs of practical clinical trials in large scale networks and need for increasingly efficient designs, we believe more efficient and less costly methods are needed for implementation research.

Related design developments are evident in the use of Sequential Multiple Assignment Randomized Trial (SMART) designs, a clinical trial design that experimentally examines strategy choices and allows multiple comparison options (Murphy et al. 2007, 2008; Ten Have et al. 2003a, b). Developed from the seminal work of Lavori et al. (2000, 2001),



SMART designs can address the need to accommodate patient and provider preferences for treatment while using adaptive randomization strategies to support interpretation of the outcomes from the research. Another example of complex designs emerging from the RCT tradition is the randomized fractional factorial design which screens more efficiently and tests multiple treatment components with less cost (Collins et al. 2005; Strecher et al. 2008).

Duan and colleagues encourage implementation researchers to explore the potential implications of designs used in quality engineering, a field facing similar marketing concerns albeit for the “uptake” of consumer products. Noting that “this work uses experimental and statistical methods such as robust parameter design to achieve good performance under representative user conditions,” Duan suggests that “research on mental health interventions can benefit from such techniques” (Duan et al. 2001, p. 413). Quality engineering design enhancements include screening experiments and fractional factorial designs from agricultural applications, now being used in public health intervention research. Marketing research further exemplifies use of conjoint analysis, which also has promise for implementation research (Green and Krieger 1993; Green et al. 2001). Designs from the quasi-experimental tradition are used in quality improvement research. Both Walshe (2007) and Cable (2001) call for use of quasi-experimental designs for the purpose of “enhancing causal interpretations of quality improvement interventions.” Similarly, while rooted in RCT’s, West and colleagues’ paper, “Alternatives to the Randomized Control Trial” (West et al. 2008), references the “Campbellian Perspective” with its seminal work on conceptualizing four types of threat to internal validity and how these might be addressed in quasi-experimental designs. West reviews classical approaches in this tradition, such as non-random quantitative assignment of treatment including the regression discontinuity design and interrupted time series analysis, along with techniques for use in observational studies to address internal validity concerns, such as propensity score and sensitivity analyses.

Finally, there have been recent studies to compare the results between randomized experiments and observational studies. These have included both examples of strong divergence between observational studies and experimental studies, such as found in studies of hormone replacement therapy (Barrett-Connor et al. 2004) as well as examples where the results were remarkably consistent. Notable is the paper published in 2000 by Concato et al. (2000) that examined meta-analyses of RCTs and meta-analyses of either cohort or case-control studies on the same intervention. Across interventions addressing five very different medical conditions, the authors found remarkable similarity in the results between the two types of designs, which are perceived to be quite different in the hierarchy of evidence. The authors concluded that “The results of well-designed observational studies (with either a cohort or a case-control design) do not systematically overestimate the magnitude of the effects of treatment as compared to those in randomized, controlled trials on the same topic” (p. 1887). A later study published in 2008 by Cook et al. (2008) came to the same conclusion when comparing the results from randomized experiments and regression-discontinuity designs. In both papers, the authors argued that the quality of the observational studies had to be high to be comparable to the results in randomized designs. This line of research informs the emerging field of implementation science by suggesting that observational studies may also be seriously considered.

## Conclusion

The typology of research organized by the NIH Roadmap initiative and extended by the recent National Research Council/IOM prevention report is useful for locating the emerging science of dissemination and implementation. However, review of design element issues in the emerging science underscores the unresolved issues regarding the inherent tension between internal and external validity in designs under consideration across the science to

practice spectrum. Despite this tension, a structured review of implementation studies in child mental health and child welfare service settings demonstrated the dominant use of randomized designs which are strong on internal validity, and no use of alternative designs from the quasi-experimental tradition or the emerging alternative designs from the efficacy tradition, despite their potential to address problems in external validity. This dominance of experimental designs is also apparent in the work of the Cochrane Collaborative EPOC review group, which has championed the systematic review of research testing implementation strategies in physical medical settings. We would argue that the emergence of implementation science will be well served by extensive and critical discussion and development of design options that address the tension between internal and external validity.

## Acknowledgments

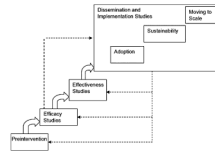
Support for this manuscript was provided by the following grants: P30 MH074678, DA019984, MH040859, DA024 370, P30 MH068579, and MH080916.

## References

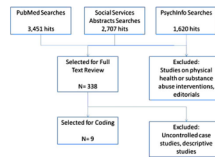
- Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*. in press.
- Atkins MS, Frazier SL, Leathers SJ, Graczyk PA, Talbott E, Jakobsons L, et al. Teacher key opinion leaders and mental health consultation in low-income urban schools. *Journal of Consulting and Clinical Psychology* 2008;76(5):905–908. [PubMed: 18837608]
- Barrett-Connor E, Grady D, Stefanick M. The rise and fall of menopausal hormone therapy. *Annual Review of Public Health* 2004;26:115–140.
- Barwick MA, Peters J, Boydell K. Getting to uptake: Do communities of practice support the implementation of evidence-based practice? *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 2009;18(1):16–29. [PubMed: 19270845]
- Bert SC, Farris JR, Borkowski JG. Parent training: Implementation strategies for adventures in parenting. *Journal of Primary Prevention* 2008;29(3):243–261. [PubMed: 18446440]
- Berwick D. The science of improvement. *JAMA* 2008;299(10):1182–1184. [PubMed: 18334694]
- Brown CH, Liao J. Principles for designing randomized preventive trials in mental health: An emerging developmental epidemiologic paradigm. *American Journal of Community Psychology* 1999;27:677–714.
- Brown C, Ten Have T, Jo B, Dagne G, Wyman P, Muthén B, et al. Adaptive designs for randomized trials in public health. *Annual Review of Public Health* 2009;30:1–25.
- Brown CH, Wang W, Kellam SG, Muthén BO, Petras H, Toyinbo P, et al. Methods for testing theory and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. *Drug and Alcohol Dependence* 2008;S95:S74–S104. [PubMed: 18215473]
- Brown CH, Wyman PA, Guo J, Peña J. Dynamic wait-listed designs for randomized trials: New designs for prevention of youth suicide. *Clinical Trials* 2006;3:259–271. [PubMed: 16895043]
- Cable G. Enhancing causal interpretations of quality improvement interventions. *Quality in Health Care* 2001;10(3):179–186. [PubMed: 11533426]
- Chamberlain P, Brown CH, Saldana L, Reid J, Wang W, Marsenich L, et al. Engaging and recruiting counties in an experiment on implementing evidence-based practice in California. *Administration and Policy in Mental Health* 2008;35(4):250–260. [PubMed: 18302015]
- Collins L, Murphy S, Nair V, Strecher V. A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine* 2005;30(1):65–73. [PubMed: 16097907]
- Concato J, Shah N, Horwitz R. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *The New England Journal of Medicine* 2000;342(25):1887–1892. [PubMed: 10861325]

- Cook T, Shadish W, Wong V. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management* 2008;27(4):724–750.
- Culliton B. Extracting knowledge from science: A conversation with Elias Zerhouni. *Health Affairs* 2006;25(3):w94–w103. [PubMed: 16527825]
- Duan N, Braslow J, Weisz J, Wells K. Fidelity, adherence, and robustness of interventions. *Psychiatric Services* 2001;52(4):413. [PubMed: 11274482]
- Duan N, Wells K, Braslow J, Weisz J. Randomized encouragement trial: A paradigm for public health intervention evaluation. 2002 Unpublished Manuscript.
- Flay B. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine* 1986;15(5):451–474. [PubMed: 3534875]
- Glasgow R, Magid D, Beck A, Ritzwoller D, Estabrooks P. Practical clinical trials for translating research to practice: Design and measurement recommendations. *Medical Care* 2005;43(6):551–557. [PubMed: 15908849]
- Glisson C, Schoenwald SK. The ARC organizational and community intervention strategy for implementing evidence-based children's mental health treatments. *Mental Health Services Research* 2005;7(4):243–259. [PubMed: 16320107]
- Green P, Krieger A. Conjoint analysis with product-positioning applications. *Handbooks in Operations Research and Management Science* 1993;5:467–515.
- Green P, Krieger A, Wind Y. Thirty years of conjoint analysis: Reflections and prospects. *Interfaces* 2001;31(3):56–73.
- Herschell AD, McNeil CB, Urquiza AJ, McGrath JM, Zebell NM, Timmer SG, et al. Evaluation of a treatment manual and workshops for disseminating parent-child interaction therapy. *Administration and Policy in Mental Health and Mental Health Services Research* 2009;36(1):63–81. [PubMed: 19016321]
- Lavori P, Dawson R, Rush A. Flexible treatment strategies in chronic disease: Clinical and research implications. *Biological Psychiatry* 2000;48(6):605–614. [PubMed: 11018231]
- Lavori P, Rush A, Wisniewski S, Alpert J, Fava M, Kupfer D, et al. Strengthening clinical effectiveness trials: Equipoise-stratified randomization. *Biological Psychiatry* 2001;50(10):792–801. [PubMed: 11720698]
- Lochman JE, Boxmeyer C, Powell N, Qu L, Wells K, Windle M. Dissemination of the Coping Power Program: Importance of intensity of counselor training. *Journal of Consulting and Clinical Psychology* 2009;77(3):397–409. [PubMed: 19485582]
- Murphy S, Lynch K, Oslin D, McKay J, Ten Have T. Developing adaptive treatment strategies in substance abuse research. *Drug and Alcohol Dependence* 2007;88:S24–S30. [PubMed: 17056207]
- Murphy, S.; Oslin, D.; Ten-Have, T. Convergence: Five committees, one dialogue. Paper presented at the NIMH conference; Crystal City, VA. 2008 Jun.
- National Research Council and Institute of Medicine. Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities. Washington, DC: National Academy Press; 2009.
- Palinkas LA, Aarons GA, Chorpita BF, Hoagwood K, Landsverk J, Weisz JR. Cultural exchange and the implementation of evidence-based practices: Two case studies. *Research on Social Work Practice* 2009;19(5):602–612.
- Palinkas LA, Schoenwald SK, Hoagwood K, Landsverk J, Chorpita BF, Weisz JR. An ethnographic study of implementation of evidence-based treatments in child mental health: First steps. *Psychiatric Services* 2008;59(7):738–746. [PubMed: 18586990]
- Pawson, R.; Tilley, N. Realistic evaluation. London: Sage Publications Ltd; 1997.
- Poduska J, Kellam SG, Brown CH, Ford C, Keegan N, Windham A, et al. A group randomized control trial of a classroom-based intervention aimed at preventing early risk factors for drug abuse: Integrating effectiveness and implementation research. *Implementation Science* 2009;4(56):1–11. [PubMed: 19123945]
- Proctor E, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: An emerging science with conceptual, methodological, and training

- challenges. *Administration and Policy in Mental Health and Mental Health Services Research* 2009;36(1):24–34. [PubMed: 19104929]
- Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Clinical Epidemiology* 2009;62(5):499–505.
- Shortell SM. Increasing value: A research agenda for addressing the managerial and organizational challenges facing health care delivery in the United States. *Medical Care Research and Review* 2004;61:12S–30S. [PubMed: 15375281]
- Strecher V, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. Web-based smoking cessation components and tailoring depth: Results of a randomized trial. *American Journal of Preventive Medicine* 2008;34(5):373–381. [PubMed: 18407003]
- Ten Have T, Coyne J, Salzer M, Katz I. Research to improve the quality of care for depression: Alternatives to the simple randomized clinical trial\* 1. *General Hospital Psychiatry* 2003a;25(2):115–123.
- Ten Have T, Joffe M, Cary M. Causal logistic models for non-compliance under randomized treatment with univariate binary response. *Statistics in Medicine* 2003b;22(8):1255–1283. [PubMed: 12687654]
- Tunis S, Stryer D, Clancy C. Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003;290(12):1624–1632. [PubMed: 14506122]
- Walshe K. Understanding what works—and why—in quality improvement: The need for theory-driven evaluation. *International Journal for Quality in Health Care* 2007;19:57–59. [PubMed: 17337518]
- West S, Duan N, Pequegnat W, Gaist P, Des Jarlais D, Holtgrave D, et al. Alternatives to the randomized controlled trial. *American Journal of Public Health* 2008;98(8):1359–1366. [PubMed: 18556609]
- Westfall J, Mold J, Fagnan L. Practice-based research— “Blue Highways” on the NIH roadmap. *JAMA* 2007;297(4):403–406. [PubMed: 17244837]
- Zerhouni E. The NIH roadmap. *Science (Washington)* 2003;302(5642):63–72. [PubMed: 14526066]
- Zerhouni E. Translational and clinical science—time for a new vision. *New England Journal of Medicine* 2005;353(15):1621–1623. [PubMed: 16221788]



**Fig. 1.** Stages of research in prevention research cycle. *Source* Chapter 11: Implementation and Dissemination of Prevention Programs (2009) in National Research Council and Institute of Medicine (2009, p. 326)



**Fig. 2.**  
Structured literature review: searches and items identified

**Table 1**

Databases included in structured literature review

<b>Databases</b>	<b>Website</b>
PubMed MEDLINE	<a href="http://www.ncbi.nlm.nih.gov/sites/entrez">http://www.ncbi.nlm.nih.gov/sites/entrez</a>
PsychInfo	<a href="http://www.apa.org/pubs/databases/psycinfo/index.aspx">http://www.apa.org/pubs/databases/psycinfo/index.aspx</a>
Social Services Abstracts	<a href="http://www.csa.com/factsheets/ssa-set-c.php">http://www.csa.com/factsheets/ssa-set-c.php</a>

Table 2

Design features for nine implementation studies

Study	Brief description	Study design	Study methods	Level at which randomization occurred	Implementation stage	Implementation theory/model specified
Atkins et al. (2008)	Examines influence of key opinion leaders and mental health providers on classroom teachers' use of classroom practices for children with ADHD	RCT	Quantitative	School	Implementation	Diffusion of innovations theory
Barwick et al. (2009)	Examines whether practitioners in a community of practice changed their practice more readily than practitioners in the practice as usual group	RCT	Quantitative	Organization	Adoption, implementation	Community of practice model
Bert et al. (2008)	Assesses the differential impact of three intervention conditions on knowledge of parenting intervention principles	RCT	Mixed qualitative and quantitative methods	Individual parent	Implementation	None
Chamberlain et al. (2008)	Describes the initial phase of a randomized trial that tests two methods of implementing MTFC in 40 non-early adopting California counties	RCT	Mixed qualitative and quantitative methods	Counties within clusters	Exploration, adoption, implementation, sustainability	Community development teams (CDT)
Glisson and Schoenwald (2005)	Describes an on-going county level trial of the implementation of MST in rural Appalachian communities.	RCT	Quantitative	County	Implementation	Availability, responsiveness and continuity (ARC) model
Herschell et al. (2009)	Describes the evaluation of two different training programs for PCIT	CCT	Quantitative	NA	Implementation	None
Lochman et al. (2009)	Examines the impact of the intensity of training on the Coping Power provided to practitioners	RCT	Quantitative	Individual school counselors	Implementation	None
Palinkas et al. (2008)	Examines treatment implementation in the Clinic Treatment Project, a randomized effectiveness trial of evidence-based treatments	RCT	Mixed qualitative and quantitative methods	Individual clinician	Implementation	Innovation implementation model
Palinkas et al. (2009)	Describes a statewide effectiveness trial of SafeCare in Oklahoma compared to services as usual	RCT	Mixed qualitative and quantitative methods	Region and agency	Implementation	Diffusion of innovations theory, innovation implementation model