

Approaches to Mixed Methods Dissemination and Implementation Research: Methods, Strengths, Caveats, and Opportunities

Carla A. Green · Naihua Duan · Robert D. Gibbons ·
Kimberly E. Hoagwood · Lawrence A. Palinkas ·
Jennifer P. Wisdom

Published online: 11 April 2014
© Springer Science+Business Media New York 2014

Abstract Limited translation of research into practice has prompted study of diffusion and implementation, and development of effective methods of encouraging adoption, dissemination and implementation. Mixed methods techniques offer approaches for assessing and addressing processes affecting implementation of evidence-based interventions. We describe common mixed methods approaches used in dissemination and implementation research, discuss strengths and limitations of mixed methods approaches to data collection, and suggest promising methods not yet widely used in implementation research. We review qualitative, quantitative, and hybrid approaches to mixed methods dissemination and implementation studies, and describe methods for integrating

multiple methods to increase depth of understanding while improving reliability and validity of findings.

Keywords Dissemination and implementation research · Mixed methods · Qualitative methods

Introduction

Lack of translation of research findings into practice, and significant lags in translation time for those that are translated, have prompted health services researchers to study natural processes of diffusion of innovative findings and to develop more effective methods of encouraging adoption, dissemination and implementation (D & I) (Berwick 2003; Proctor et al. 2009; Westfall et al. 2007). These efforts have led to more nuanced understandings of the processes and agents involved in diffusion and implementation, and what was once viewed as a vexing failure among clinicians and organizations to implement what was “evidence-based” is now more appropriately viewed as a failure to design implementation strategies that take into account the organizational, clinical, and social environments that affect uptake of research.

What is emerging is a more complex picture of the ways in which research findings and implementation processes are situated within organizational cultures and processes, within communities, and in concert with regional, state, and national policies. There is also increasing recognition that if care and health are to be improved, research must be designed, disseminated, and implemented in concert with stakeholders. This means learning about the experiences, perspectives, and needs of a full range of players, from policy-makers to agency directors, supervisors to front-line clinical staff, and from patients to their families. To

C. A. Green (✉)
Center for Health Research, Kaiser Permanente Northwest,
3800 N. Interstate Avenue, Portland, OR 97227, USA
e-mail: carla.a.green@kpchr.org

N. Duan
Columbia University Medical Center, New York, NY, USA

R. D. Gibbons
Center for Health Statistics, University of Chicago,
Chicago, IL, USA

K. E. Hoagwood
Department of Child and Adolescent Psychiatry, New York
University Langone Medical Center, New York, NY, USA

L. A. Palinkas
Albert G. and Frances Lomas Feldman Professor of Social
Policy and Health, School of Social Work, University of
Southern California, Los Angeles, CA, USA

J. P. Wisdom
George Washington University, Washington, DC, USA

achieve these goals, researchers have increasingly turned to mixed methods approaches to understand, collaborate with, and respond to stakeholders in the communities in which they intend their work to be disseminated and implemented (Shortell 1999). Mixed methods designs—those which systematically integrate qualitative and quantitative data—are intrinsically suited to the work of D & I research: They provide an array of methods and opportunities for collecting, triangulating, and analyzing information gathered from different stakeholder constituencies, and for developing a deeper understanding of the full range of perspectives and processes that affect adoption and implementation. Formative, process, and evaluative questions are all fair game (Stetler et al. 2006), and mixed methods designs capitalize on the strengths of each method used while attempting to reduce each method's weaknesses. That is, they address the limited generalizability that results from most qualitative approaches and the limited depth of understanding typical of findings derived from quantitative data by combining techniques from both approaches.

Integrating Qualitative and Quantitative Data

In mixed methods studies, qualitative and quantitative data can be integrated at multiple stages—at the time of data collection, during analysis, or during interpretation. Data are integrated differently depending on whether the study collects qualitative and quantitative data sequentially or simultaneously, and on the extent to which the study places emphasis on each technique (Creswell and Plano Clark 2007). In some D & I mixed methods designs, for example, qualitative data can be analyzed to inform later quantitative data collection processes (sequential, exploratory models) or qualitative data collection that follows quantitative data collection can be analyzed to explain quantitative results (sequential, explanatory models). When both types of data are collected simultaneously, they may be analyzed together, each to inform the other, or one type of data may be transformed for use in analyses of the other data (e.g., qualitative data converted to categorical data for inclusion in quantitative analysis; quantitative data used to create classifications of individuals whose qualitative responses are then compared). Irrespective of the methods chosen, an important component of integration should be analyses of consistencies and inconsistencies in findings (Creswell and Plano Clark 2007). This involves searching for and evaluating inconsistencies within and across data sources. For example, in thematic analyses, it is important to identify and report on cases that contradict what appear to be common themes in the data; when comparing quantitative results to qualitative findings, inconsistencies might be a

function of differential responses of subgroups to the intervention that can be further explored using existing data.

Because more detailed methods of analysis and reporting of qualitative and mixed methods studies are beyond the scope of this paper, we refer readers to existing comprehensive sources.¹ In the sections that follow, we review qualitative and quantitative approaches that can be integrated in different ways to produce strong mixed methods designs. We also cover hybrid methods—approaches that include, as essential components, multiple data sources and types, or analytic techniques that inherently integrate qualitative and quantitative approaches. Most hybrid methods have more recent origins, so have been used less frequently or not yet applied to D & I research. We include these methods because of their potential promise in this context.

Qualitative and Hybrid Approaches within Dissemination and Implementation Research

Creswell identifies five traditions of qualitative inquiry (biography, phenomenology, grounded theory, ethnography and case study) and five philosophical frameworks underlying these approaches (ontological, epistemological, axiological, rhetorical, methodological) (Creswell 1998). These traditions and approaches remain the underpinnings of qualitative inquiry within mixed methods D & I research. Within these frameworks, researchers have a wide range of mixed methods designs and data collection techniques from which to choose. Appropriately matching research and sampling design to research questions, data collection approaches, emphasis on qualitative versus quantitative data, and ordering of particular methods, are essential to producing interpretable and useful findings (Creswell and Plano Clark 2007; Palinkas et al. 2011a, 2011, 2013). In the sections that follow, we describe the qualitative methods most commonly used in D & I research, and describe some of the ways those methods can be integrated with, or augment, quantitative approaches.

¹ For analyzing and reporting qualitative and mixed methods data, see Miles and Huberman (1994), Creswell (1998), Creswell and Plano Clark (2007), Denzin and Lincoln (1998, 2005), Bernard (2011), and Bourgeault et al. (2010). For focus group interviews, see Morgan and Krueger (1998). Those interested in grounded theory and constant comparative analyses should refer to: (Charmaz 2001, 2006; Creswell 1998; Glaser and Strauss 1967). See (Hsieh and Shannon 2005; Krippendorff 2004) for detailed explanations of content analysis. For discussions of rigor and threats to validity in qualitative research, including of reliability, validity, and trustworthiness, see (Davies and Dodd 2002; Krefting 1991; Morse et al. 2002; Poland 1995; Whittemore et al. 2001).

Most qualitative inquiry in D & I research revolves around the collection and analysis of text or observational data. Text may be generated using interviews, result from notes taken during observations, or be drawn from existing documents, such as meeting minutes, correspondence, training materials, bylaws, standard practice manuals, organizational reports and websites, and books, magazines or newspapers. Analysis of text can include the following: (1) testing hypotheses (e.g., by way of content analysis); (2) identifying common meanings and interconnections such as among clinicians providing team-based care (e.g., through hermeneutic analysis); (3) discovering commonalities in the ways individuals talk or tell stories about an event such as an implementation process (e.g., using narrative analysis), or (4) identifying categories and concepts and linking those concepts into a formal theory of implementation roll-outs (e.g., using grounded theory) (Bernard 2011). Mixed methods D & I projects typically pair one or more of these qualitative methods with one or more quantitative methods to triangulate findings and improve validity, to aid understanding of quantitative results, or to include measures derived from qualitative data in quantitative analyses.

Interviews

Interviews are among the most commonly employed qualitative data collection methods used in D & I research. They can be conducted individually or in groups, and can be semi-structured or structured in nature. Interviews have a place in all phases of D & I research, from formative and developmental assessments through implementation, process, and evaluative components.

Semi-Structured Interviews

Semi-structured interviews are typically exploratory, while structured interviews are more likely to be quantitative and confirmatory—that is, structured interviews typically have fixed responses deriving from conceptual models with clear hypotheses to be tested (see section on Formal Ethnographic Methods below, for an exception). In structured interviews, participants are asked the same questions in the same order and provided with the same set of responses. Semi-structured interviews allow the flexibility of qualitative data collection while at the same time providing more standardization than in naturalistic or unstructured interviews. Interview guides provide a set of questions and prompts to guide the interviewer, but the interviewer is allowed to follow the flow of conversation, asking questions as they occur naturally, and following-up with unanticipated questions when interviewees raise topics of

particular interest or importance. In some cases, structured and semi-structured questions are included in the same interview allowing easy integration and triangulation of results, as the sample is the same for both qualitative and quantitative data collection.

In addition to the type of interview approach chosen, researchers must make choices about how they will frame semi-structured interviews. The questions that are asked, and the consequences of those choices, depend on what data are desired, and how those data will be integrated with other analyses, including whether responses will be coded for inclusion in quantitative analyses. Questions asking interviewees to generalize and compare their situation or experiences to those of others, will produce sociological, often abstract, answers in response, whereas when researchers seeking to understand the specifics of peoples' experiences, interviewers need to ask questions that elicit the particulars of those individual experiences (Chase 2005). For this reason, if the goal is to understand the results of quantitative analyses, researchers may choose questions that lead to generalizations, while those developing questions as part of formative work that precedes and informs implementation of interventions may be more likely to use questions that result in detailed responses that will help to identify obstacles to implementation or opportunities for smoothing intervention roll-out. If the goal is to code qualitative responses so that they can be included in statistical analyses, interviewers must be sure to ask all participants these questions and probe for responses that can be clearly coded in either a binary or scalar fashion.

Key Informant Interviews

Key informant interviews can range from loosely organized conversations to semi-structured interviews—the distinction between these interviews and other approaches is that they are conducted with individuals who have extensive and important information (Gilchrist and Williams 1999) needed to carry out and understand processes targeted by D & I projects. That is, they are interviews with experts (Marshall 1996) who are selected because they have comprehensive knowledge because of their roles or because of their ability to translate, interpret, teach or mentor the researcher in the setting of interest (Dicicco-Bloom and Crabtree 2006). Although historically used in anthropology in lieu of broader sampling procedures (Tremblay 1957), in D & I research, they are most commonly used early in developmental evaluations, or after implementation, to take advantage of the informant's in-depth knowledge of the setting and how its characteristics may affect or have affected implementation. Key informant interviews can also be used during other phases of

evaluation as a relatively quick and simple method for assessing effects of context on interventions, or on intervention processes, progress, and outcomes. Such interviews, though extremely helpful in obtaining an “insider’s” view, also provide unique perspectives that may not be representative of other stakeholders. Nevertheless, the best key informants are keen observers who often understand and report a range of stakeholder perspectives, even if they do not agree with those perspectives. They can help guide data collection and generate hypotheses in addition to providing insight and aiding understanding at different project phases. Corroboration and examination of hypotheses resulting from key informant interviews are important methods of integrating findings using multiple methods (Gilchrist and Williams 1999).

Individual in-Depth Interviews

Compared to key informant interviews, individual in-depth interviews are typically designed to obtain deeper understandings of commonalities and differences among groups of individuals that share important characteristics or experiences, or to understand the perspectives of individuals at different points along a continuum of interest (Miller and Crabtree 1999). In-depth interviews, in particular, are intended to elicit personal, intimate, and detailed narratives (Dicicco-Bloom and Crabtree 2006). Their most important use in D & I projects is to shed light on the ways in which implementation processes interact with organizations and stakeholders to produce outcomes—both expected and unexpected. Recognizing that stakeholders’ primary responsibilities are rarely research focused, interview length and guides are constructed to address key research questions *and* be mindful of the exigencies experienced by those being interviewed. Therefore, interview guides for busy clinicians or administrators are often shorter and more narrowly focused; interviews with users of clinical services may be longer and, correspondingly, include questions that delve more deeply, with prompts to encourage additional exploration of interviewees’ experiences.

Semi-structured interview guides are often adapted over time as data are analyzed and more is learned about the research question and the strengths and weaknesses of the guide (Charmaz 2006). This adaptability makes semi-structured interviews—whether group or individual—extremely useful in mixed methods D & I research. Various designs are common, including interviews in the formative phase of a quantitative D & I project, explanatory interviews used to explain results obtained using other methods (typically quantitative) or to understand processes and implementation during rollout of a program, an intervention or a randomized controlled trial (Creswell et al. 2011;

Palinkas et al. 2011b, 2011; Stetler et al. 2006). Flexibility allows researchers to change or add questions in response to findings from interviews as well as other data sources. Similarly, findings can inform implementation while it is still in process, providing opportunities to alter approaches and increase the likelihood of successes in ever-changing clinical and social environments. Thus, most interview-based qualitative D & I research is flexible and iterative in nature, and opportunities for integration are many and varied. The increased rigor obtained from triangulation of interview and quantitative data increases confidence when results converge across data-collection methods, and this is a major benefit of this mixed-methods pairing (Torrey et al. 2012).

Focus Group Interviews

Focus groups are collective conversations or group interviews that have at their core the assumption that group interaction will stimulate thoughts and ideas that might not be elicited in an individual interview (Kamberelis and Dimitriadis 2005). Typically, a group of individuals sharing common experiences or states (e.g., parents of children with mental health problems), or exposure to specific services, are asked about their perspectives, beliefs, or attitudes regarding their shared experiences. Like individual interviews, focus groups have a place in formative, process, implementation, and explanatory phases of projects. They have advantages over individual interviews in that they can be more cost-effective (more participants interviewed in the same time period) and because the group structure can be more stimulating, and thus may elicit a wider range of perspectives and ideas than individual interviews (Morgan 1993). Group interviews also have disadvantages compared to individual interviews. They are more difficult to coordinate, convene, and conduct; participants may be less likely to share sensitive information in group settings; and it may not be possible to explore topics in as in-depth a manner as in individual interviews (Bernard 2011). Moreover, in D & I research, focus group interviews are more likely to include stakeholders who know one another when compared to other research applications. This is particularly true when interviews target staff involved in service delivery or project implementation. In such situations, power relations become important, because truthful or complete responses may not be forthcoming from participants who feel that full disclosure might put them at-risk in some way (e.g., when supervisors are participants in the same group interview). If such situations cannot be avoided, alternative techniques that protect confidentiality, such as individual interviews or surveys may provide more accurate data. Focus group interview data can be integrated with D & I data from other sources in most of the ways that

individual interview data can be integrated. An exception to this is the ability to convert qualitative data to binary or scalar indicators for use in quantitative analyses. Unless group perspectives can be characterized for composite measures, this is a limitation of group over individual interviews for mixed methods integration.

Observational Approaches: Participant Observation and Ethnographic Methods

Observation is fundamental to all scientific inquiry, though the types of observation differ substantially from observation that follows experimental interventions to non-interventionist techniques that seek to examine the natural course of events as they would occur without the presence of the observer (Adler and Adler 1998). Participant observation and ethnography are qualitative observational techniques, developed primarily in anthropology and sociology, that have significant value in D & I research. Observational research of this type has been evolving over time, with a shift in focus from the researcher as dispassionate observer to that of a participant observer interacting as a member of the community s/he is studying (Angrosino 2005).

Ethnography refers to both the process *and* the outcome of the research venture, which includes interpretations and descriptions of a particular group, organization, or system, including processes and behaviors at the levels studied, and details about the customs, norms, roles, and methods of interaction in the setting (Creswell 1998). In D & I research, ethnography is typically carried out through participant (or sometimes non-participant) observation and interviews, with the researcher immersing him/herself in the regular, daily, activities of the people involved in the setting while recording observations that document interactions, processes, tasks and outcomes, as well as personal reactions to these observations. In most cases, this is a long-term investment of time and energy, with regular observation occurring over weeks, months or years (though see the section on rapid ethnographic assessment (REA) for an alternative model). Goals are to (a) produce a full picture of the ways in which a project was implemented, (b) describe the extent of fidelity to the intervention, and (c) identify and understand barriers and facilitators of implementation. Researchers often use key informant interviews, in-depth interviews and focus group interviews. Combined with text from other sources and available quantitative data, to create detailed accounts of the implementation process and its context. Taking careful, detailed, field notes is a critical component of ethnography, as is recording of interviews, review of relevant documents and quantitative data, and working to identify any personal

biases that might affect conclusions. Searching for information that might contradict conclusions is also critical to good producing good ethnography.

Ethical concerns that are particular to participant observation must also be addressed. For example, difficulties can arise if key individuals do not consent to be observed, particularly when they interact with others who have consented. Ethnography is not for the faint of heart, but when done well, it can provide invaluable, comprehensive, information about implementation and dissemination that, when combined with quantitatively measured outcomes, can provide a complete picture of the processes and outcomes associated with D & I projects. Gabbay and Le May's ethnographic work on clinical decision making in two primary care settings clearly shows how implementation of evidence based practices in routine clinical settings compares to expectations among researchers and administrations about the ways clinicians consume research and become aware of and use guidelines. Over 2 years of observations and interviews carried out in two small group practices, the authors found that clinicians relied on trusted sources such as colleagues, and free magazines, rather than directly accessing and appraising information and evidence from original sources or guidelines (Gabbay and le May 2004). Clinicians referred to guidelines to confirm existing practices, and when they had patients with challenging or unfamiliar problems. Guidelines were not routinely used, and little attention was paid to them when they were disseminated (Gabbay and le May 2004). The findings outlined in this report represent the kind of information essential to researchers developing frameworks designed to increase adoption of evidence based practices.

Rapid Ethnographic Assessment (REA)

Rapid ethnographic assessment is hybrid method and one of a group of rapid evaluation and assessment methods (REAM) that have significant potential for use in dissemination, implementation, and evaluation studies, particularly when time is of the essence and rigorous research results are needed (Beebe 2001; McNall and Foster-Fishman 2007). REAM and REA offer real-time evaluations that can provide quick assessments of local conditions that can be used to inform the design and implementation of effective interventions (McNall and Foster-Fishman 2007). Some projects can be completed in as little as 8 weeks (McNall and Foster-Fishman 2007); methods typically include key informant and focus group interviews, targeted rapid quantitative assessment surveys, and intensive direct observation (Trotter et al. 2001). Speed is gained by rapid data collection using multiple modalities, including quantitative data, with less complicated analytic approaches

used for qualitative data (e.g., coding and analysis of interview notes rather than transcribed interviews). Advantages include the ability to obtain information about implementation and processes quickly, allowing modifications. A See Murray et al. (1994) and Needle et al. (2003) for examples.

Event Structure Analysis (ESA)

To our knowledge, this promising hybrid method has yet to be applied in D & I research. It offers a systematic, uniform, computer-assisted method (Heise 2012) of analyzing and interpreting narrative and observational data derived from ethnographic studies (Corsaro and Heise 1990; Heise 1989). It appears particularly relevant for analyzing the kinds of organizational processes (Pentland 1999; Stevenson and Greenberg 1998; Trumpy 2008) that are often critical to D & I research. Event structure analysis (ESA) breaks down the constituent parts of event sequences to develop graphical models that allow causal interpretations and explanations of processes that can then be tested and further refined. The strength of the method is that analysts, through the process of specifying the model, are forced to carefully consider contextual factors, causal ordering of events, the processes leading to each event, and the understanding, and interpretation of all events in the model (Griffin and Korstad 1998).

Formal Ethnographic Methods

Formal ethnographic methods are hybrid approaches that involve structured qualitative data collection and analytic techniques that are quasi-statistical in nature. Unlike semi-structured approaches, formal ethnographic methods require that the same stimuli (i.e., task or set of questions) be asked of all study participants. This is often referred to as structured interviewing (Bernard 2011) or systematic data collection (Weller and Romeny 1988). Tasks might include pile sorts, triads, rank ordering, semantic frames, or free listing. Data from tasks usually fall into one of three categories: Similarity data, in which participants provide estimates of how alike two or more items are; ordered data, in which participants provide an ordinal rating of items on a single conceptual scale; and performance data, in which responses provided by participants can be graded as “correct” or “incorrect” (Bernard 2002).

Concept Mapping

Perhaps the most common form of formal ethnographic methods used in implementation research is “concept mapping.” Developed by William Trochim (Trochim 1989), this technique blends focus group interviewing and

rank ordering with the quantitative techniques of multidimensional scaling and hierarchical cluster analysis. Concept mapping is a participatory qualitative research method that yields a conceptual framework for how a group views a particular topic. It uses inductive and structured group data collection processes to produce illustrative cluster maps depicting relationships among ideas in cluster form. It includes six distinct stages of activity: In the preparation stage, focal areas for investigation are identified and criteria for participant selection/recruitment are determined. In the generation stage, participants address the focal question and generate a list of items to be used in subsequent data collection and analysis. Qualitative data at this stage is obtained through “brainstorming” sessions. In the structuring stage, participants independently organize the list of generated items by sorting the items into piles based on perceived similarity. Each item is then rated in terms of its importance or usefulness to the focal question. In the representation or mapping stage, data are entered into specialized concept-mapping computer software (Concept Systems 2014), which is used to analyze participant data. Results include quantitative summaries of individual concepts, and visual representations or concept maps based on multidimensional scaling and hierarchical cluster analysis. In the interpretation stage, participants collectively process and qualitatively analyze the concept maps. This includes an assessment and discussion of cluster domains, evaluation of items that form each cluster, and discussion of content within each cluster. Based on this discussion, investigators may reduce the number of clusters. Finally, in the utilization stage, findings are discussed by investigators and study participants to determine how they best inform the original focal question.

Concept mapping has been used in several D & I projects. Aarons and colleagues (Aarons et al. 2009) used the technique to solicit information on factors likely to affect implementation of evidence based practices in public sector mental health settings. Providers and consumers participated in focus groups and generated a series of 105 unique statements describing barriers and facilitators of evidence based practice implementation. Participants rated statements according to importance and changeability, and real-time multidimensional scaling and hierarchical cluster analysis were used to generate a visual display of how statements clustered. Participants assigned meanings to, and identified appropriate names for, each of the 14 clusters identified (Aarons et al. 2009). This analysis uncovered a complex implementation process and multiple leverage points where change efforts would be most likely to improve implementation. Other examples of concept mapping in projects with D & I foci or D & I components include: Jabbar and Abelson (2011), Arrington et al. (2008) and Behar and Hydaker (2009).

Case Study Research

Case study research is, in most cases, a hybrid method that has long been used when there are needs to understand complex conditions and contextual factors using multiple sources of data that can be integrated to aid understanding (Yin 2003a). Sources of data may include documents, archival records, interviews, direct observation, participant observation, physical artifacts, survey and other quantitative data (Yin 2003b). Data are combined from multiple sources to create a clear and comprehensive picture of the context and demands of the research setting, the processes involved in intervention roll-out and how they change over time, and the ways the intervention affects clinical and organizational practices and outcomes among service users. Single case designs are useful as tests of theoretical or conceptual models when the case is (1) unique, extreme, or revelatory; (2) thought to be representative or typical; or (3) because there is a need for longitudinal study (Yin 2003b). Multiple case designs, sometimes called comparative case study designs, have different goals: (1) to predict similar results across cases (replication), or (2) to predict contrasting results across cases based on a particular theory or conceptual model (theoretical replication) (Yin 2003b). The rationale for multiple case studies is considered analogous to conducting multiple experiments on the same topic using the same conceptual model to replicate results (Yin 2003b). Multiple case studies require more resources and time than single case studies, but may be particularly useful in the context of practical clinical trials and other projects with multiple implementation sites.

Case study methods are sometimes underappreciated because of a perceived lack of rigor, but this may result from confusion between case study research and case study teaching (Yin 2003b). In case study teaching, characteristics of cases are altered or enhanced to facilitate learning, while such alterations are not acceptable in case study research (Yin 2003b). Lack of generalizability, particularly with single case studies, is a limitation of the case study approach, though Yin (2003b) argues that scientists rarely generalize from a single study or experiment and suggests that rigorous case studies should be viewed as generalizable to theoretical propositions rather than to populations or universes, and thus should be used to for analytic generalizations rather than statistical generalizations (Yin 2003b). In this context, rigorous case studies provide a thorough and deep understanding of the case or cases under study—the types of information needed to understand why a particular implementation process succeeded, failed, or had mixed results. A variety of resources are available to support design and analysis of rigorous case studies, and to assess the quality and rigor of such research (Caronna 2010; Creswell 1998; Stake 2005; Yin 1999, 2003a,

2003b). A recent case study of implementation of The Incredible Years parenting intervention in a residential substance abuse treatment program for women shows the value of such approaches in D & I research (Aarons et al. 2012). The focus of the case study was on how the intervention was adapted to fit the setting and the implications of those adaptations on fidelity. Some changes were consistent with the approach and intent of the model while others were not. The authors use the case study to illustrate the need to develop implementation models that allow for greater flexibility and adaptation while staying true to critical frameworks and core elements.

Qualitative Comparative Analysis (QCA)

Qualitative comparative analysis (QCA) is a special type of case study methodology based on principles of set theory and designed to elucidate cross-case patterns for studies with small sample sizes, using a “configurational” rather than a relationships-between-variables approach (Ragin 1997, 1999b; Ragin et al. 2003). That is, QCA provides a method of analyzing causal complexity by examining how different configurations of antecedent factors are necessary or sufficient for producing the outcomes of interest, rather than how a common set of antecedent conditions leads to a specific outcome (Ragin 1999a, 1999b). Researchers using QCA select a case and collect data describing that case (e.g., using case study research methods), then construct truth tables that define causally relevant characteristics. Each case is reviewed to complete a row of the truth table, indicating whether each characteristic is true or false for that case. Once all cases are included and the truth table is complete, each row of the table is reviewed to identify patterns in causal combinations and to simplify the table by combining rows that show common patterns leading to the same outcome. When the table is fully simplified, an equation or set of equations can be written to describe the causal pathway(s). QCA has been used increasingly in health services research, but has had little application in D & I research. See Ford and colleagues (Ford et al. 2005) for one D & I example.

Quantitative Designs and Considerations within Mixed Methods Dissemination and Implementation Research

As a result of the strict requirements necessary to produce reliable and valid results of statistical analyses, quantitative components of D & I research are more constrained than qualitative approaches. That is, the structures associated with “real-world” implementation settings, procedures necessary for implementation, and the composition and methods of the intervention, combined with the hypotheses

to be tested and the limits of specific statistical procedures, can significantly constrain study designs for quantitative outcomes. These limits suggest opportunities for mixed-methods integration: Quantitative requirements for valid and reliable measures that are used without adaptation can be tempered by qualitative data collection procedures that can be modified to explore unexpected findings or processes.

Efforts to conduct effectiveness research in routine clinical settings have also led to the development of less-rigid approaches and designs that are more acceptable to stakeholders, including nonrandomized designs, need or risk-based assignment, interrupted time series designs, and pragmatic clinical trials. In the sections that follow, we review quantitative methods of particular relevance to D & I research, and discuss mixed methods applications for each approach that can fill gaps or address weaknesses associated with each approach.

Nonrandomized Designs

The exigencies of particular settings or situations, and needs to improve participation and buy-in from different stakeholders, sometimes require the use of nonrandomized designs. Several of these approaches are well-suited to mixed methods D & I research and, when threats to internal validity can be managed, are advantageous because they are more likely to be generalizable (West et al. 2008).

Need- or Risk-Based Assignment to Intervention Conditions

Need-based assignment (NBA) is a potentially promising method for managing clinical trials implementation in settings where randomization is not acceptable or possible (Finkelstein et al. 1996a, 1996b; West et al. 2008). NBA tends to be compatible with routine practice because, when properly designed, it replicates what frontline practitioners already do when developing treatment plans. In this context, formative qualitative assessments can help researchers determine the design and approach that is most appropriate for the settings in which implementation will take place. Pre-intervention assessments, administered to all participants, provide baseline need scores. Participants with scores exceeding a pre-specified threshold are offered high-intensity services (the experimental condition), while those below the threshold are offered low-intensity services (the comparison condition). Follow-up assessments are compared across conditions to assess intervention effects. Since the groups differ at baseline, a direct comparison of follow-up outcomes across intervention conditions does not

provide a valid estimate of intervention effects. Rather, adjustment is made using statistical models applied to each group to account for the pre-existing differences in baseline needs and provide a more appropriate estimate of intervention effects.

A methodological challenge in application of NBA in multi-level service structures is accommodating need at different levels. For example, some agencies may have greater needs for an intervention than others (i.e., lower functionality, higher stress) and thus should be prioritized for agency-level interventions. Additional prioritization may be warranted at provider and consumer levels (greater training needs for providers; higher symptom severity among children). To date, methods for applying needs-based assignment at multiple levels have not yet been developed. As is often the case, however, limitations of one approach suggest opportunities for others. In this case, qualitative data collection might be used to help formulate the most appropriate approaches for particular settings, and to assess need at organizational or other levels.

Regression-Discontinuity and Interrupted Time Series Designs

These quasi-experimental designs present an alternative approach to analyses of data when randomization is not possible but existing data are available (e.g., through electronic medical records) or when data can be collected over time, prior to assessment of intervention outcomes (Cook and Campbell 1979; Imbens and Lemieux 2008; Lee and Lemieux 2010; Shadish et al. 2002; Thistlethwaite and Campbell 1960; West et al. 2008). Regression discontinuity analysis can be applied to data collected from need-based allocation assignment; fitting separate regression curves to those who fall above the threshold and receive the high-intensity intervention, and those who fall below the threshold and receive the low-intensity intervention. The gap (“discontinuity”) between the two regression curves at the threshold is used to assess the intervention effect. Interrupted time series analysis is a special type of regression discontinuity analysis, with time used as the thresholding device. This method uses data collected from periods prior to interventions to establish trends; changes in trends following interventions can then be examined to establish evidence of intervention effects. Results of these types of designs often integrate nicely with qualitative process and evaluation data collected over the course of the study. Changes in trends over time, discontinuities identified following interventions, lags in effects, or lack of intervention effects can often be explained when qualitative process evaluation data have been collected simultaneously with quantitative data.

Pragmatic Clinical Trials: Experimental Designs with Random Assignment in “Real World” Settings

Pragmatic or practical clinical trials (PCTs) (Schwartz and Lellouch 2009; Tunis et al. 2003) are designed to inform practical decision-making in routine clinical settings, and can be contrasted with explanatory clinical trials, the focus of which is to identify treatment effects under controlled laboratory conditions. Because of their practical focus, PCTs are often designed as comparative effectiveness trials of alternative interventions. Inclusion criteria tend to be minimally restrictive, data are collected for a range of health outcomes rather than a narrow few, and implementation is tested in a variety of care settings (Tunis et al. 2003).

PCTs and explanatory clinical trials are based on different paradigms and address distinct aims and objectives, some of which are well-suited to mixed methods approaches. Most importantly, in explanatory trials, contextual factors are usually considered confounders to be controlled, while the same factors are often considered integral components of implementation protocols in pragmatic trials. As an example, when comparing behavioral therapy versus medication for the treatment of adolescent depression, behavioral therapy invariably requires more contact between patient and provider. From the explanatory perspective, such a difference in the intensity of patient-provider contact is considered a confounding factor, and needs to be controlled in order to rule out the possibility that observed differences between therapy and medication patients are not a result of differences in the intensity of patient-provider contact. From the pragmatic perspective, however, the higher intensity of patient-provider contact is a natural component of the implementation of the therapy in its practical context (Schwartz and Lellouch 2009). The clinical decision that needs to be made for implementation is how the therapy “bundle,” including the imbedded higher intensity of patient-provider contact, differs from the medication “bundle,” including the imbedded lower intensity of patient-provider contact. Mixed methods approaches offer opportunities to study and describe contextual and other non-controlled factors at work in PCTs, and findings can be used to address implementation barriers.

Randomization in PCTs

Randomization can be extremely valuable in PCTs because it can be difficult to determine if differences are due to baseline differences in the groups that receive or do not receive an intervention, or whether the results can be attributed to the intervention (Hotopf 2002). For these reasons most PCTs include some form of randomization,

though this can sometimes be difficult in clinical settings if randomization distorts routine care delivery or clinician-patient relationships, or if the intervention targets a vulnerable population with reservations about research participation. Irrespective of randomization designs, PCT researchers must balance and understand the effects of conducting a study, and collecting data, on the clinical settings in which they are working (Thorpe et al. 2009) and the effects of those settings on intervention outcomes. Qualitative approaches have important applicability here, helping to identify barriers or facilitators of implementation, stakeholder perspectives, and adaptations that can increase the likelihood of success (Luce et al. 2009; Oakley et al. 2006). Qualitative data collection can also be used to monitor the effects of the research enterprise on organizational functioning and clinical processes so that negative effects can be mitigated to the greatest extent possible or, for those that cannot be mitigated, carefully described. Such descriptions can provide invaluable information for decision makers considering intervention adoption and for researchers designing alternative approaches.

Parallel Randomized and Nonrandomized Trial Designs

In situations where a large proportion of eligible individuals decline randomization, external validity is threatened. Instead of excluding these candidates, it is possible to use designs in which participants are retained and entered into a separate nonrandomized trial based on their treatment preferences. In this case, addition of the nonrandomized trial data to the randomized trial data can enhance generalizability of results. Parallel randomized and nonrandomized trial designs have considerable potential because they take advantage of the stronger internal validity of the RCT and enhanced generalizability from the quasi-experimental trial. Qualitative data collection with participants who refuse randomization can shed light on factors affecting willingness to be randomized and determine how those factors might be related to trial outcomes.

Selection Bias

Selection bias is a common challenge for implementation studies in which participants are allowed to self-select. Self-selection means that those receiving one intervention are likely to be different from those receiving the other intervention. For example, patients with severe conditions may be more likely to receive more intensive interventions, while patients with milder conditions may be more likely to receive less intensive interventions or no active intervention beyond “watch and monitor.” In such situations, direct comparisons of outcomes across intervention conditions may be misleading. Using qualitative data collection to

understand self-selection may help researchers to better target interventions.

Propensity scores, the conditional probability of receiving a specific intervention given a set of observed covariates (Rosenbaum 2002; Rosenbaum and Rubin 1983, 1984) are a promising approach for addressing selection bias resulting from imbalances between intervention and comparison groups on observed covariates. These include as weighting, stratification, and matching (Rosenbaum 2002; Rosenbaum and Rubin 1983, 1984). One limitation of the approach is that propensity score methods can only be used to address overt bias, namely selection bias due to observed confounding factors. If hidden bias resulting from unobserved confounding factors is present, propensity score methods are limited. That is, they can be used to balance the observed covariates and any components of hidden bias that are correlated with observed covariates, but additional methodologies such as instrumental variable analysis (Angrist et al. 1996), and sensitivity analyses (Rosenbaum 2002; Rosenbaum and Rubin 1983, 1984) are needed to more fully address these problems. Qualitative assessments can be used uncover unobserved confounders and identify factors that might be measured for inclusion in propensity score calculations.

Design and Analysis for Multi-level Interventions

Mental health service delivery is often multi-level in nature, with clients nested within providers, providers nested within agencies or clinics, and agencies nested within county and state policies. A common design used for multi-level interventions is the group or cluster randomized design, with randomized assignment at the highest level of the intervention, most often the agency or clinic level. This approach has two significant limitations, however. First, the evaluation is subject to variance inflation at the agency level; second, there is no information that allows us to untangle the impact of the various components of the intervention targeted at each level, nor to assess whether the interventions at those levels interact (Donner 1998; Donner and Klar 1994; Murray 1998). Split plot designs present an alternative that addresses the limits of cluster randomized designs (Fisher 1925; Yates 1935). These designs are particularly useful for state-level rollouts because they improve statistical efficiency and enable the unique contributions from interventions at each level to be disentangled. For example, agencies can be randomized to either receive an agency-level intervention or remain in usual care. Then, within agencies, providers are randomized to either receive a provider-level intervention, or remain in usual practice. Finally, within agencies and providers (with all combinations of agency and provider level interventions), consumers are randomized to either

receive consumer-level interventions (e.g. engagement strategies), or remain in usual care. Combining the 3 phases of randomization, we can focus on main-effects analyses to separately assess the impacts of the three different intervention components. Under the assumption of additivity, each of the 3 intervention components can be estimated and tested using the entire sample, achieving full statistical efficiency. Moreover, each of the intervention effects is free from design effects (variance inflation) from the higher levels. Disadvantages to the split plot design include the need to have clearly defined interventions at each level, and adequate sample sizes. Mixed methods approaches to these designs typically include qualitative data collection for process and implementation evaluations to ensure understanding of critical factors affecting processes and outcomes at different levels. Such evaluations might include focus group interviews with consumers; individual or focus group interviews with clinicians, and key informant interviews with executive directors or other administrative staff. Participant observation can also be of great value in identifying and describing how processes play out at each level, and how they interact across levels.

Quantitative Approaches to Data Collection and Integration within Mixed Methods D & I Studies

Survey Methods

Survey methods are widely-used, cost-effective, methods of collecting large amounts of data that are representative of populations of interest. They can be particularly useful to D & I researchers conducting multi-level implementation projects, and often are developed and administered using mixed methods approaches (Beatty and Willis 2007; Fowler 2009). Formative qualitative work may be used to identify key themes and constructs to be assessed in a survey and cognitive interviewing used to develop, refine and validate survey items (Beatty and Willis 2007). Surveys can also include open-ended questions that allow respondents to answer using their own words. When such mixed methods techniques are employed, a successful survey can be characterized as an integrated mixed methods approach that used qualitative methods to develop and ascertain the meaning of questions, quantitative methods to collect the structured data required for the study, and open-ended qualitative questions to explore areas that are not appropriate for close-ended responses or for which adequate information is not available to create fixed response categories.

Target Populations and Sample Selection

While most surveys target data collection from individual respondents in a specified population (e.g., clients served by

an agency), many D & I projects also seek data at the agency or organizational level (e.g., health care facilities or business entities). In either case, researchers must define the population, specify how members will be identified and approached, and tailor questions to the population. For D & I in state systems, for example, respondents may include state policymakers, such as commissioners, deputy directors, or other executive leadership, organizational administrators, as well as clinicians, patients, and families. Because most projects cannot afford to administer surveys to the entire target population, sampling is necessary and the sampling strategy must allow population-level inferences. When the population is small (e.g., state policymakers) key informant or other individual interviews may be more useful and cost-effective than surveys. Whether for qualitative or quantitative approaches, sample selection methods depend on the research questions, the expected ranges of responses, and the mechanisms available for accessing members in the target population. A number of excellent resources exist for survey sampling approaches and methods (Babbie 1990; Fowler 2009; Frankel et al. 1999; Kish 1995; Marsden and Wright 2010; Rossi et al. 1983). Similar resources are available for sampling in qualitative research (Blankertz 1998; Draucker et al. 2007; Morse 2000; Palinkas et al. 2013; Strauss and Corbin 1998).

Questionnaires

Survey methods are typically implemented using a questionnaire (or instrument) that includes a collection of questions inquiring about specific behaviors or attributes. A simple questionnaire presents the same list of questions sequentially, in the same order, to all respondents. More complex questionnaires can be constructed that are customized to present a set of questions selected according to the characteristics of the specific respondent (e.g., a survey about adolescent mental health services would skip questions about pregnancy for male respondents). Such use of branching logic is facilitated by information technology in administering surveys (e.g., computer assisted interviewing or CAI) (Couper et al. 1998). Surveys can be conducted either in person, by telephone, using the web (Couper 2008), or via ecological momentary assessment (EMA) using mobile devices (Shiffman et al. 2008).

The design of a good survey questionnaire usually follows a back-engineering approach, starting with the ultimate goal of data collection—the aims of the study and the hypotheses to be tested. Many experienced investigators begin their design process by drafting an outline of the final report and detailing how they will answer their fundamental analysis questions (Scheuren 2013). This pinpoints which pieces of information will be required and leads to construction of an analysis plan that connects data

collection objectives to specific questions and specifies the ways questions should be asked (Scheuren 2013). Similar back-engineering is beneficial for qualitative questions, even if the research is exploratory and theory-generating. That is, development of the approach as well as materials, such as interview guides, should be clearly tied to the desired end-product, including expectations for how the approach and materials might change over time. The draft final report then helps the researcher identify the information needed to describe all study participants, includes a clear sampling and data analysis plan, details opportunities for evolutions in approach, and specifies the key questions that are to be answered.

Survey Administration

Surveys can be administered in various ways, including paper-and-pencil, computer-assisted personal interviews (CAPI), computer-assisted telephone interviews, web-based surveys, and surveys using mobile devices (Couper 2008; Couper et al. 1998; Shiffman et al. 2008). While interviewer-administered surveys provide a high level of accuracy and more complete data, self-administered surveys are less costly and can provide greater confidentiality and improved respondent comfort (Tourangeau and Smith 1996). Information technology-based approaches can increase accuracy and reduce human error, though they may require programming expertise and can be vulnerable to technology failures. Different modes of administration can be particularly useful in D & I research, with mode selected to optimize comfort for and response from the target population. Here too, qualitative data can provide information to researchers who are making decisions about which survey modalities are best for particular topics and participants.

Survey Modalities and Mode Effects

Using a combination of survey administration modes can optimize response rates while containing survey costs. For example, if formative work suggests that significant proportions of the target population are comfortable with self-administered web surveys, this approach might be attempted first, followed by interviewer-administered telephone surveys for nonrespondents. A third mode might also be deployed if needed, with an interviewer traveling to the respondent to administer a face-to-face survey. When multiple modes of administration are combined, however, responses may vary across modes of administration. For example, participants may be more willing to accurately respond to sensitive questions in self-administered modes than in face-to-face modes (Tourangeau and Smith 1996). Such mode effects may require statistical adjustments (de

Leeuw 2005; Fowler et al. 2002) or alternatively, the use of random response techniques (Lensvelt-Mulders et al. 2005) to improve response validity.

Measurement Development for Dissemination and Implementation Research

Researchers are increasingly developing more rigorous methods of measurement development, and taking advantages of technological advances that make better measurement possible *and* less burdensome on participants. Such methods have not yet been widely used in D & I research, but their benefits, particularly as common outcome metrics are developed, suggest significant opportunities for application in this area. For example, in surveying agencies in a dissemination/implementation program, the methods described below can be used to customize questions for specific agencies or service users so that provide the most informative information for each unique situation, reduce respondent burden, and avoid the pitfalls of “one size fits all” approaches.

Item Response Theory (IRT)

Classical and item response theory (IRT) measurement methods differ dramatically in approach to administration and scoring. For example, consider a track and field meet in which athletes participate in a hurdles race and in high jump. Suppose that the hurdles are not all the same height and the score is determined by the runner’s time *and* the number of hurdles cleared. For the high jump, the cross bar is raised incrementally and athletes try to jump over the bar without dislodging it. The first of these two events is like a traditionally scored objective test: runners attempt to clear hurdles of varying heights, analogous to answering questions of varying difficulty. In either case, a specific counting operation measures ability to clear hurdles or answer questions. On the high jump, ability is measured by the highest position the athlete clears. IRT measurement uses the same logic as the high jump: Items are arranged on a continuum with fixed points of increasing difficulty of endorsement. Scores are measured by the location on the continuum of the most difficult item endorsed. In IRT, scores are obtained using a scale point rather than a count.

These methods of scoring hurdles and high jump, or their analogues in traditional and IRT measures, contrast sharply: If hurdles are arbitrarily added or removed, number of hurdles cleared cannot be compared across races run with different hurdles or different numbers of hurdles. Scores lose their comparability if item composition is changed. The same is *not* true, however, of the high jump or of IRT scoring. If one of the positions on the bar were omitted, height cleared is unchanged and only the precision

of the measurement at that point on the scale is affected. Thus, in IRT scoring, a certain number of items can be arbitrarily added, deleted or replaced without losing comparability of scores, thus reducing participant burden and costs of administration. This property of scaled measurement, compared with counts, is the most salient advantage of IRT over classical measurement.

Computerized adaptive testing (CAT) can be used to develop banks of items for specific populations, with a range of endorsement difficulties (Weiss 1985), for use in IRT-based outcomes measurement. Cognitive interviewing and other qualitative approaches can be used to understand participants’ experiences of endorsement difficulty for particular items, as well as factors associated with difficulty of endorsement. Once item banks are available, they can be used to build complex surveys that adapt to individual participants’ characteristics and response patterns (Gibbons et al. 2013). While use of CAT and IRT has been widespread in educational measurement, it has been less widely used in D & I research. In addition to cognitive interviewing, qualitative methods such focus groups and as concept mapping can be used to inform the item development necessary to use IRT approaches in D & I research.

Vertical Scaling

Vertical or developmental scaling is an IRT method frequently used in educational assessments to provide a single scale that is applicable across all grade levels so that growth in learning can be measured with a common yardstick (Tong and Kolen 2007). In the measurement of child outcomes following a D & I project, items that may be appropriate for a 14- or 15-year-old may not be appropriate for a 9- or 10- year-old. As long as there is a subset of common “anchor” items that can be used for adjacent developmental (age) groups, IRT-based vertical scaling can be used to provide a common assessment across different developmental levels. These techniques can be used to deliver lower-cost, less burdensome, outcome measures that can be compared across similar D & I projects.

Summary and Conclusions

Mixed methods approaches to D & I research hold great promise for unpacking the processes and factors that are often hidden within the black boxes that have been the hallmark of evidence-based practice implementation. A multitude of qualitative techniques are available to meet the needs of D & I researchers, ranging from traditional ethnographic techniques to REA, and from purely observational techniques to hybrid designs that inherently

combine both qualitative and quantitative methods. Conventional survey methods have their place as well, but newer technologies, combined with improvements in the underpinnings of measurement theory, make possible a new generation of more valid and less burdensome assessment processes. Together, the methods described in this paper provide a set of approaches that could be considered a toolkit for mixed methods D & I research.

Such a toolkit has particularly important application in multi-level state-related policy research that involves scaling up of evidence-based practices. These methods are useful for comparing the different perspectives of the various stakeholders and constituents—ranging from policy-makers to agency directors and management; from front-line clinical staff to patients and families—and for developing clear understandings of implementation successes and failures. Mixed methods provide the opportunity to produce enriched understandings of the complexities of implementation processes, and to tap into the nuances of vexing barriers and promising facilitators of implementation. Together, they provide necessary methods for improving strategies for effective, efficient, and sustainable roll-outs of evidence-based practices.

Acknowledgment This work was funded by an award from the National Institute of Mental Health (P30-MH090322; K. Hoagwood, PI).

References

- Aarons, G. A., Miller, E. A., Green, A. E., Perrott, J. A., & Bradway, R. (2012). Adaptation happens: A qualitative case study of implementation of the incredible years evidence-based parent training programme in a residential substance abuse treatment programme. *Journal of Children's Services*, 7(4), 233–245.
- Aarons, G. A., Wells, R. S., Zagsursky, K., Fettes, D. L., & Palinkas, L. A. (2009). Implementing evidence-based practice in community mental health agencies: A multiple stakeholder analysis. *American Journal of Public Health*, 99(11), 2087–2095. doi:10.2105/AJPH.2009.161711.
- Adler, P. A., & Adler, P. (1998). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp. 79–109). Thousand Oaks: Sage Publications.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variable. *Journal of the American Statistical Association*, 91, 444–455.
- Angrosino, M. V. (2005). Recontextualizing observation: Ethnography, pedagogy, and the prospects for a progressive political agenda. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (3rd ed., pp. 729–745). Thousand Oaks: Sage Publications.
- Arrington, B., Kimmey, J., Brewster, M., Bentley, J., Kane, M., Van, B. C., et al. (2008). Building a local agenda for dissemination of research into practice. *Journal of Public Health Management and Practice*, 14(2), 185–192. doi:10.1097/01.PHH.0000311898.03573.28.
- Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmont: Wadsworth.
- Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2), 287–311.
- Beebe, J. (2001). *Rapid assessment process: An introduction*. Lanham: AltaMira Press.
- Behar, L. B., & Hydaker, W. M. (2009). Defining community readiness for the implementation of a system of care. *Administration and Policy in Mental Health*, 36(6), 381–392. doi:10.1007/s10488-009-0227-x.
- Bernard, H. R. (2002). *Research methods in anthropology: Qualitative and quantitative approaches*. Walnut Creek: Alta Mira Press.
- Bernard, H. R. (2011). *Research methods in anthropology: Qualitative and quantitative approaches* (5th ed.). Lanham: AltaMira Press.
- Berwick, D. M. (2003). Disseminating innovations in health care. *The Journal of the American Medical Association*, 289(15), 1969–1975. doi:10.1001/jama.289.15.1969.
- Blankertz, L. (1998). The value and practicality of deliberate sampling for heterogeneity: A critical multiplist perspective. *American Journal of Evaluation*, 19(3), 307–324.
- Bourgeault, I., Dingwall, R., & de Vries, R. (2010). *Handbook of qualitative methods in health research*. Los Angeles: Sage Publications.
- Caronna, C. A. (2010). Why use qualitative methods to study health care organizations? Insights from multi-level case studies. In I. Bourgeault, R. Dingwall, & R. de Vries (Eds.), *Handbook of qualitative methods in health research* (pp. 71–87). Los Angeles: Sage Publications.
- Charmaz, K. (2001). Qualitative interviewing and grounded theory analysis. In J. F. Gubrium & S. Hutchinson (Eds.), *Handbook of interview research* (pp. 675–694). Thousand Oaks: Sage Publications.
- Charmaz, K. (2006). *Constructing grounded theory*. Thousand Oaks: Sage Publications Inc.
- Chase, S. E. (2005). Narrative inquiry: Multiple lenses, approaches, voices. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (3rd ed., pp. 651–679). Thousand Oaks: Sage Publications.
- Concept Systems. (2014). The concept system software, Global Max™. Concept Systems Incorporated. Retrieved 8 April 2014 from: <http://www.conceptsystems.com>.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Corsaro, W. A., & Heise, D. R. (1990). Event structure models from ethnographic data. In C. C. Clogg (Ed.), *Sociological methodology: 1990* (pp. 1–57). Cambridge: Basil Blackwell.
- Couper, M. P. (2008). *Designing effective web surveys*. Cambridge: Cambridge University Press.
- Couper, M. P., Baker, R., Bethlehem, J., Clark, C., Martin, J., Nicholls, W. I., et al. (1998). *Computer-assisted survey information collection*. New York: John Wiley.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks: Sage Publications.
- Creswell, J. W., Klassen, A. C., Plano Clark, V. L., & Smith, K. C. f. t. O. o. B. a. S. S. R. (2011). Best practices for mixed methods research in the health sciences. http://obssr.od.nih.gov/mixed_methods_research. Accessed on 8 Apr 2014.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks: Sage Publications Inc.
- Davies, D., & Dodd, J. (2002). Qualitative research and the question of rigor. *Qualitative Health Research*, 12(2), 279–289.
- de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 2, 233–255.
- Denzin, N. K., & Lincoln, Y. S. (1998). *Collecting and interpreting qualitative materials*. Thousand Oaks: Sage Publications.

- Denzin, N. K., & Lincoln, Y. S. (2005). *Handbook of qualitative research* (3rd ed.). Thousand Oaks: Sage Publications.
- Dicicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, 40(4), 314–321. doi:10.1111/j.1365-2929.2006.02418.x.
- Donner, A. (1998). Some aspects of the design and analysis of cluster randomization trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(1), 95–113.
- Donner, A., & Klar, N. (1994). Cluster randomization trials in epidemiology: Theory and application. *Journal of Statistical Planning and Inference*, 42(1), 37–56.
- Draucker, C. B., Martsof, D. S., Ross, R., & Rusk, T. B. (2007). Theoretical sampling and category development in grounded theory. *Qualitative Health Research*, 17(8), 1137–1148.
- Finkelstein, M. O., Levin, B., & Robbins, H. (1996a). Clinical and prophylactic trials with assured new treatment for those at greater risk: I. A design proposal. *American Journal of Public Health*, 86(5), 691–695.
- Finkelstein, M. O., Levin, B., & Robbins, H. (1996b). Clinical and prophylactic trials with assured new treatment for those at greater risk: II. Examples. *American Journal of Public Health*, 86(5), 696–705.
- Fisher, R. A. (1925). *Statistical methods for research workers* (14th ed.). Edinburgh: Oliver and Boyd.
- Ford, E. W., Duncan, W. J., & Ginter, P. M. (2005). Health departments' implementation of public health's core functions: An assessment of health impacts. *Public Health*, 119(1), 11–21. doi:10.1016/j.puhe.2004.03.002.
- Fowler, F. J. (2009). *Survey research methods* (4th ed.), Applied social research methods series Thousand Oaks: Sage Publications.
- Fowler, F. J., Jr, Gallagher, P. M., Stringfellow, V. L., Zaslavsky, A. M., Thompson, J. W., & Cleary, P. D. (2002). Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. *Medical Care*, 40(3), 190–200.
- Frankel, M. R., Shapiro, M. F., Duan, N., Morton, S. C., Berry, S. H., Brown, J. A., et al. (1999). National probability samples in studies of low-prevalence diseases. Part II: Designing and implementing the HIV cost and services utilization study sample. *Health Services Research*, 34(5 Pt 1), 969–992.
- Gabbay, J., & le May, A. (2004). Evidence based guidelines or collectively constructed “mindlines?” Ethnographic study of knowledge management in primary care. *BMJ*, 329(7473), 1013. doi:10.1136/bmj.329.7473.1013.
- Gibbons, R. D., Weiss, D. J., Pilkonis, P. A., Frank, E., Moore, T., Kim, J. B. et al. (2013). The CAT-DI: A computerized adaptive test for depression. *Archives of General Psychiatry*, in press.
- Gilchrist, V. J., & Williams, R. L. (1999). Key informant interviews. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (Second ed., pp. 71–88). Thousand Oaks: Sage Publications.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing Company.
- Griffin, L. J., & Korstad, R. R. (1998). Historical inference and event-structure analysis. *International Review of Social History*, 43(Supplement S6), 145–165. doi:10.1017/S0020859000115135.
- Heise, D. R. (1989). Modeling event structures. *The Journal of Mathematical Sociology*, 14(2–3), 139–169. doi:10.1080/0022250X.1989.9990048.
- Heise, D. (2012). Ethno. Retrieved from <http://www.indiana.edu/~socpsy/ESA/>. Accessed on 8 Apr 2014.
- Hotopf, M. (2002). The pragmatic randomised controlled trial. *Advances in Psychiatric Treatment*, 8(5), 326–333.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. doi:10.1177/1049732305276687.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615–635.
- Jabbar, A. M., & Abelson, J. (2011). Development of a framework for effective community engagement in Ontario. *Canada. Health Policy*, 101(1), 59–69. doi:10.1016/j.healthpol.2010.08.024.
- Kamberelis, G., & Dimitriadis, G. (2005). Focus groups: Strategic Articulations of Pedagogy, Politics, and Inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The sage handbook of qualitative research* (3rd ed.). Thousand Oaks: Sage Publications Inc.
- Kish, L. (1995). *Survey sampling*. New York: Wiley Classics Library John Wiley and Sons Inc.
- Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy*, 45(3), 214–222.
- Krippendorff, K. (2004). *Content analysis: An introduction to methodology*. Thousand Oaks: Sage Publications Inc.
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2), 281–355.
- Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods and Research*, 33(3), 319–348.
- Luce, B. R., Kramer, J. M., Goodman, S. N., Connor, J. T., Tunis, S., Whicher, D., et al. (2009). Rethinking randomized clinical trials for comparative effectiveness research: The need for transformational change. *Annals of Internal Medicine*, 151(3), 206–209.
- Marsden, P. V., & Wright, J. D. (2010). *Handbook of survey research* (2nd ed.). Bingley: Emerald Group Publishing Limited.
- Marshall, M. N. (1996). The key informant technique. *Family Practice*, 13(1), 92–97.
- McNall, M., & Foster-Fishman, P. G. (2007). Methods of rapid evaluation, assessment, and appraisal. *American Journal of Evaluation*, 28(2), 151–168.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks: Sage Publications.
- Miller, W. L., & Crabtree, B. F. (1999). Depth interviewing. *Doing qualitative research* (pp. 123–201). Thousand Oaks: Sage Publications.
- Morgan, D. L. (1993). *Successful focus groups: Advancing the state of the art*. Newbury Park: Sage Publications.
- Morgan, D. L., & Krueger, R. A. (1998). *The focus group kit*. Thousand Oaks: Sage Publications.
- Morse, J. M. (2000). Determining sample size. *Qualitative Health Research*, 10(1), 3–5.
- Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. Access on December 30, 2008 from http://www.ualberta.ca/~iiqm/backissues/1_2Final/html/morse.html. *International Journal of Qualitative Methods*, 1(2). Retrieved from http://www.ualberta.ca/~iiqm/backissues/1_2Final/html/morse.html.
- Murray, D. M. (1998). *Design and analysis of group-randomized trials* (29th ed.). Oxford: Oxford University Press.
- Murray, S. A., Tapson, J., Turnbull, L., McCallum, J., & Little, A. (1994). Listening to local voices: Adapting rapid appraisal to assess health and social needs in general practice. *BMJ*, 308(6930), 698–700.
- Needle, R. H., Trotter, R. T., Singer, M., Bates, C., Page, J. B., Metzger, D., et al. (2003). Rapid assessment of the HIV/AIDS crisis in racial and ethnic minority communities: An approach for timely community interventions. *American Journal of Public Health*, 93(6), 970–979.
- Oakley, A., Strange, V., Bonell, C., Allen, E., & Stephenson, J. (2006). Process evaluation in randomised controlled trials of complex interventions. *BMJ*, 332(7538), 413–416. doi:10.1136/bmj.332.7538.413.

- Palinkas, L. A., Aarons, G. A., Horwitz, S., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011a). Mixed method designs in implementation research. *Administration and Policy in Mental Health*, 38(1), 44–53. doi:10.1007/s10488-010-0314-z.
- Palinkas, L. A., Aarons, G. A., Horwitz, S., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011b). Mixed method designs in implementation research. *Administration and Policy in Mental Health*, 38(1), 44–53.
- Palinkas, L. A., Horwitz, S. M., Chamberlain, P., Hurlburt, M. S., & Landsverk, J. (2011c). Mixed-methods designs in mental health services research: A review. *Psychiatric Services*, 62(3), 255–263. doi:10.1176/appi.ps.62.3.255.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2013). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health*
- Pentland, B. T. (1999). Building process theory with narrative: From description to explanation. *Academy of Management Review*, 24(4), 711–724.
- Poland, B. D. (1995). Transcription quality as an aspect of rigor in qualitative research. *Qualitative Inquiry*, 1(3), 290–310.
- Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health*, 36(1), 24–34. doi:10.1007/s10488-008-0197-4.
- Ragin, C. C. (1997). Turning the tables: How case-oriented research challenges variable-oriented research. *Comparative Social Research*, 16, 27–42.
- Ragin, C. C. (1999a). The distinctiveness of case-oriented research. *Health Services Research*, 34(5 Pt 2), 1137–1151.
- Ragin, C. C. (1999b). Using qualitative comparative analysis to study causal complexity. *Health Services Research*, 34(5 Pt 2), 1225–1239.
- Ragin, C. C., Shulman, D., Weinberg, A., & Gran, B. (2003). Complexity, generality, and qualitative comparative analysis. *Field Methods*, 15, 323–340.
- Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer Science+Business Media.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387), 516–524.
- Rossi, P. H., Wright, J. D., & Anderson, A. B. (1983). *Handbook of survey research: Quantitative studies in social relations* (First ed.). Waltham: Academic Press.
- Scheuren, F. (2013). Chapter 6, Designing a questionnaire. https://www.whatisasurvey.info/downloads/pamphlet_current.pdf. Retrieved from https://www.whatisasurvey.info/downloads/pamphlet_current.pdf. Accessed 8 Apr 2014.
- Schwartz, D., & Lellouch, J. (2009). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Clinical Epidemiology*, 62(5), 499–505. doi:10.1016/j.jclinepi.2009.01.012.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32.
- Shortell, S. M. (1999). The emergence of qualitative methods in health services research. *Health Services Research*, 34(5 Pt 2), 1083–1090.
- Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 443–466). Thousand Oaks: Sage Publications.
- Stetler, C. B., Legro, M. W., Wallace, C. M., Bowman, C., Guihan, M., Hagedorn, H., et al. (2006). The role of formative evaluation in implementation research and the QUERI experience. *Journal of General Internal Medicine*, 21(Suppl 2), S1–S8.
- Stevenson, W. B., & Greenberg, D. N. (1998). The formal analysis of narratives of organizational change. *Journal of Management*, 24(6), 741–762.
- Strauss, A. L., & Corbin, J. (1998). Theoretical sampling. *Basics of qualitative research: Techniques and procedures for developing grounded theory* (pp. 201–215). Thousand Oaks: SAGE Publications Inc.
- Thistlethwaite, D., & Campbell, D. (1960). Regression-discontinuity analysis: an alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51, 309–317.
- Thorpe, K. E., Zwarenstein, M., Oxman, A. D., Treweek, S., Furberg, C. D., Altman, D. G., et al. (2009). A pragmatic-explanatory continuum indicator summary (PRECIS): A tool to help trial designers. *CMAJ: Canadian Medical Association Journal*, 180(10), E47–E57.
- Tong, Y., & Kolen, M. J. (2007). Comparisons of methodologies and results in vertical scaling for educational achievement tests. *Applied Measurement in Education*, 20(2), 227–253. doi:10.1080/08957340701301207.
- Torrey, W. C., Bond, G. R., McHugo, G. J., & Swain, K. (2012). Evidence-based practice implementation in community mental health settings: The relative importance of key domains of implementation activity. *Administration and Policy in Mental Health*, 39(5), 353–364. doi:10.1007/s10488-011-0357-9.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60(2), 275–304.
- Tremblay, M. A. (1957). The key informant technique: A nonethnographic application. *American Anthropologist*, 59(4), 688–701. doi:10.1525/aa.1957.59.4.02a00100.
- Trochim, W. M. (1989). Introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, 12, 1–16.
- Trotter, R. T., Needle, R. H., Goosby, E., Bates, C., & Singer, M. (2001). A methodological model for rapid assessment, response, and evaluation: The RARE program in public health. *Field Methods*, 13(2), 137–159.
- Trumpy, A. J. (2008). Subject to negotiation: The mechanisms behind co-optation and corporate reform. *Social Problems*, 55, 519–536.
- Tunis, S. R., Stryer, D. B., & Clancy, C. M. (2003). Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *The Journal of the American Medical Association*, 290(12), 1624–1632.
- Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, 53(6), 774–789.
- Weller, S. C., & Romeny, A. K. (1988). *Systematic data collection* (12th ed.). Newbury Park: Sage.
- West, S. G., Duan, N., Pequegnat, W., Gaist, P., Des Jarlais, D. C., Holtgrave, D., et al. (2008). Alternatives to the randomized controlled trial. *American Journal of Public Health*, 98(8), 1359–1366.
- Westfall, J. M., Mold, J., & Fagnan, L. (2007). Practice-based research—“Blue Highways” on the NIH roadmap. *The Journal of the American Medical Association*, 297(4), 403–406. doi:10.1001/jama.297.4.403.
- Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research*, 11(4), 522–537.

- Yates, F. (1935). Complex experiments, with discussion. *Supplement to the Journal of the Royal Statistical Society, Series B* 2, 2(2), 181–247.
- Yin, R. K. (1999). Enhancing the quality of case studies in health services research. *Health Services Research*, 34(5 Pt 2), 1209–1224.
- Yin, R. K. (2003a). *Applications of case study research* (2nd ed.), Applied social research methods series Thousand Oaks: Sage Publications Inc.
- Yin, R. K. (2003b). *Case study research: Design and methods* (3rd ed.), Applied social research methods series Thousand Oaks: Sage Publications.