

Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research

Lawrence A. Palinkas · Sarah M. Horwitz ·
Carla A. Green · Jennifer P. Wisdom ·
Naihua Duan · Kimberly Hoagwood

Published online: 6 November 2013
© Springer Science+Business Media New York 2013

Abstract Purposeful sampling is widely used in qualitative research for the identification and selection of information-rich cases related to the phenomenon of interest. Although there are several different purposeful sampling strategies, criterion sampling appears to be used most commonly in implementation research. However, combining sampling strategies may be more appropriate to the aims of implementation research and more consistent with recent developments in quantitative methods. This paper reviews the principles and practice of purposeful sampling in implementation research, summarizes types and categories of purposeful sampling strategies and provides a set of recommendations for use of single strategy or multistage strategy designs, particularly for state implementation research.

Keywords Mental health services · Children and adolescents · Mixed methods · Qualitative methods implementation · State systems

Introduction

Recently there have been several calls for the use of mixed method designs in implementation research (Proctor et al. 2009; Landsverk et al. 2012; Palinkas et al. 2011a; Aarons et al. 2011). This has been precipitated by the realization that the challenges of implementing evidence-based and other innovative practices, treatments, interventions and programs are sufficiently complex that a single methodological approach is often inadequate. This is particularly true of efforts to implement evidence-based practices (EBPs) in statewide systems where relationships among key stakeholders extend both vertically (from state to local organizations) and horizontally (between organizations located in different parts of a state). As in other areas of research, mixed method designs are viewed as preferable in implementation research because they provide a better understanding of research issues than either qualitative or quantitative approaches alone (Palinkas et al. 2011a, b, c). In such designs, qualitative methods are used to explore and obtain depth of understanding as to the reasons for success or failure to implement EBP or to identify strategies for facilitating implementation while quantitative methods are used to test and confirm hypotheses based on an existing conceptual model and obtain breadth of understanding of predictors of successful implementation (Teddlie and Tashakkori 2003).

Sampling strategies for quantitative methods used in mixed methods designs in implementation research are

L. A. Palinkas (✉)
School of Social Work, University of Southern California,
669 W. 34th Street, Los Angeles, CA 90089-0411, USA
e-mail: palinkas@usc.edu

S. M. Horwitz · K. Hoagwood
Department of Child and Adolescent Psychiatry, New York
University, New York, NY, USA

C. A. Green
Center for Health Research, Kaiser Permanente Northwest,
Portland, OR, USA

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

N. Duan
Department of Psychiatry and New York State Neuropsychiatric
Institute, Columbia University, New York, NY, USA

generally well-established and based on probability theory. In contrast, sampling strategies for qualitative methods in implementation studies are less explicit and often less evident. Although the samples for qualitative inquiry are generally assumed to be selected purposefully to yield cases that are “information rich” (Patton 2002), there are no clear guidelines for conducting purposeful sampling in mixed methods implementation studies, particularly when studies have more than one specific objective. Moreover, it is not entirely clear what forms of purposeful sampling are most appropriate for the challenges of using both quantitative and qualitative methods in the mixed methods designs used in implementation research. Such a consideration requires a determination of the objectives of each methodology and the potential impact of selecting one strategy to achieve one objective on the selection of other strategies to achieve additional objectives.

In this paper, we present different approaches to the use of purposeful sampling strategies in implementation research. We begin with a review of the principles and practice of purposeful sampling in implementation research, a summary of the types and categories of purposeful sampling strategies, and a set of recommendations for matching the appropriate single strategy or multistage strategy to study aims and quantitative method designs.

Principles of Purposeful Sampling

Purposeful sampling is a technique widely used in qualitative research for the identification and selection of information-rich cases for the most effective use of limited resources (Patton 2002). This involves identifying and selecting individuals or groups of individuals that are especially knowledgeable about or experienced with a phenomenon of interest (Cresswell and Plano Clark 2011). In addition to knowledge and experience, Bernard (2002) and Spradley (1979) note the importance of availability and willingness to participate, and the ability to communicate experiences and opinions in an articulate, expressive, and reflective manner. In contrast, probabilistic or random sampling is used to ensure the generalizability of findings by minimizing the potential for bias in selection and to control for the potential influence of known and unknown confounders.

As Morse and Niehaus (2009) observe, whether the methodology employed is quantitative or qualitative, sampling methods are intended to maximize efficiency and validity. Nevertheless, sampling must be consistent with the aims and assumptions inherent in the use of either method. Qualitative methods are, for the most part, intended to achieve depth of understanding while quantitative methods are intended to achieve breadth of understanding

(Patton 2002). Qualitative methods place primary emphasis on saturation (i.e., obtaining a comprehensive understanding by continuing to sample until no new substantive information is acquired) (Miles and Huberman 1994). Quantitative methods place primary emphasis on generalizability (i.e., ensuring that the knowledge gained is representative of the population from which the sample was drawn). Each methodology, in turn, has different expectations and standards for determining the number of participants required to achieve its aims. Quantitative methods rely on established formulae for avoiding Type I and Type II errors, while qualitative methods often rely on precedents for determining number of participants based on type of analysis proposed (e.g., 3–6 participants interviewed multiple times in a phenomenological study versus 20–30 participants interviewed once or twice in a grounded theory study), level of detail required, and emphasis of homogeneity (requiring smaller samples) versus heterogeneity (requiring larger samples) (Guest et al. 2006; Morse and Niehaus 2009; Padgett 2008).

Types of Purposeful Sampling Designs

There exist numerous purposeful sampling designs. Examples include the selection of extreme or deviant (outlier) cases for the purpose of learning from an unusual manifestations of phenomena of interest; the selection of cases with maximum variation for the purpose of documenting unique or diverse variations that have emerged in adapting to different conditions, and to identify important common patterns that cut across variations; and the selection of homogeneous cases for the purpose of reducing variation, simplifying analysis, and facilitating group interviewing. A list of some of these strategies and examples of their use in implementation research is provided in Table 1.

Embedded in each strategy is the ability to compare and contrast, to identify similarities and differences in the phenomenon of interest. Nevertheless, some of these strategies (e.g., maximum variation sampling, extreme case sampling, intensity sampling, and purposeful random sampling) are used to identify and expand the range of variation or differences, similar to the use of quantitative measures to describe the variability or dispersion of values for a particular variable or variables, while other strategies (e.g., homogeneous sampling, typical case sampling, criterion sampling, and snowball sampling) are used to narrow the range of variation and focus on similarities. The latter are similar to the use of quantitative central tendency measures (e.g., mean, median, and mode). Moreover, certain strategies, like stratified purposeful sampling or opportunistic or emergent sampling, are designed to achieve both goals. As Patton (2002, p. 240) explains, “the

Table 1 Purposeful sampling strategies in implementation research

Strategy	Objective	Example	Considerations
Emphasis on similarity			
Criterion-i	To identify and select all cases that meet some predetermined criterion of importance	Selection of consultant trainers and program leaders at study sites to facilitate and barriers to EBP implementation (Marshall et al. 2008)	Can be used to identify cases from standardized questionnaires for in-depth follow-up (Patton 2002)
Criterion-e	To identify and select all cases that exceed or fall outside a specified criterion	Selection of directors of agencies that failed to move to the next stage of implementation within expected period of time	
Typical case	To illustrate or highlight what is typical, normal or average	A child undergoing treatment for trauma (Hoagwood et al. 2007)	The purpose is to describe and illustrate what is typical to those unfamiliar with the setting, not to make generalized statements about the experiences of all participants (Patton 2002)
Homogeneity	To describe a particular subgroup in depth, to reduce variation, simplify analysis and facilitate group interviewing	Selecting Latino/a directors of mental health services agencies to discuss challenges of implementing evidence-based treatments for mental health problems with Latino/a clients	Often used for selecting focus group participants
Snowball	To identify cases of interest from sampling people who know people that generally have similar characteristics who, in turn know people, also with similar characteristics	Asking recruited program managers to identify clinicians, administrative support staff, and consumers for project recruitment (Green and Aarons 2011)	Begins by asking key informants or well-situated people “Who knows a lot about...” (Patton 2002)
Extreme or deviant case	To illuminate both the unusual and the typical	Selecting clinicians from state agencies or mental health with best and worst performance records or implementation outcomes	Extreme successes or failures may be discredited as being too extreme or unusual to yield useful information, leading one to select cases that manifest sufficient intensity to illuminate the nature of success or failure, but not in the extreme
Emphasis on variation			
Intensity	Same objective as extreme case sampling but with less emphasis on extremes	Clinicians providing usual care and clinicians who dropped out of a study prior to consent to contrast with clinicians who provided the intervention under investigation. (Kramer and Burns 2008)	Requires the researcher to do some exploratory work to determine the nature of the variation of the situation under study, then sampling intense examples of the phenomenon of interest
Maximum variation	Important shared patterns that cut across cases and derived their significance from having emerged out of heterogeneity	Sampling mental health services programs in urban and rural areas in different parts of the state (north, central, south) to capture maximum variation in location (Bachman et al. 2009)	Can be used to document unique or diverse variations that have emerged in adapting to different conditions (Patton 2002)
Critical case	To permit logical generalization and maximum application of information because if it is true in this one case, it’s likely to be true of all other cases	Investigation of a group of agencies that decided to stop using an EBP to identify reasons for lack of EBP sustainment	Depends on recognition of key dimensions that make for a critical case. Particularly important when resources may limit the study of only one site (program, community, population) (Patton 2002)

Table 1 continued

Strategy	Objective	Example	Considerations
Theory-based	To find manifestations of a theoretical construct so as to elaborate and examine the construct and its variations	Sampling therapists based on academic training to understand the impact of CBT training versus psychodynamic training in graduate school of acceptance of EBPs	Sample on the basis of potential manifestation or representation of important theoretical constructs Sampling on the basis of emerging concepts with the aim being to explore the dimensional range or varied conditions along which the properties of concepts vary
Confirming and disconfirming case	To confirm the importance and meaning of possible patterns and checking out the viability of emergent findings with new data and additional cases	Once trends are identified, deliberately seeking examples that are counter to the trend	Usually employed in later phases of data collection. Confirmatory cases are additional examples that fit already emergent patterns to add richness, depth and credibility. Disconfirming cases are a source of rival interpretations as well as a means for placing boundaries around confirmed findings
Stratified purposeful	To capture major variations rather than to identify a common core, although the latter may emerge in the analysis	Combining typical case sampling with maximum variation sampling by taking a stratified purposeful sample of above average, average, and below average cases of health care expenditures for a particular problem	This represents less than the full maximum variation sample, but more than simple typical case sampling
Purposeful random	To increase the credibility of results	Selecting for interviews a random sample of providers to describe experiences with EBP implementation	Not as representative of the population as a probability random sample
Nonspecific emphasis			
Opportunistic or emergent	To take advantage of circumstances, events and opportunities for additional data collection as they arise		Usually employed when it is impossible to identify sample or the population from which a sample should be drawn at the outset of a study. Used primarily in conducting ethnographic fieldwork
Convenience	To collect information from participants who are easily accessible to the researcher	Recruiting providers attending a staff meeting for study participation	Although commonly used, it is neither purposeful nor strategic

purpose of a stratified purposeful sample is to capture major variations rather than to identify a common core, although the latter may also emerge in the analysis. Each of the strata would constitute a fairly homogeneous sample.”

Challenges to Use of Purposeful Sampling

Despite its wide use, there are numerous challenges in identifying and applying the appropriate purposeful sampling strategy in any study. For instance, the range of variation in a sample from which purposive sample is to be taken is often not really known at the outset of a study. To set as the goal the sampling of information-rich informants that cover the range of variation assumes one knows that range of variation. Consequently, an iterative approach of sampling and re-sampling to draw an appropriate sample is usually recommended to make certain the theoretical

saturation occurs (Miles and Huberman 1994). However, that saturation may be determined a priori on the basis of an existing theory or conceptual framework, or it may emerge from the data themselves, as in a grounded theory approach (Glaser and Strauss 1967). Second, there are differences in opinion about these approaches among qualitative researchers, with some resisting or refusing systematic sampling of any kind and rejecting the limiting nature of such realist, systematic, or positivist approaches. This includes critics of interventions and “bottom up” case studies and critiques. Nevertheless, even those who equate purposeful sampling with systematic sampling must offer a rationale for selecting study participants that is linked with the aims of the investigation (i.e., why recruit these individuals for this particular study? What qualifies them to address the aims of the study?). While systematic sampling may be associated with a post-positivist tradition of

qualitative data collection and analysis, such sampling is not inherently limited to such analyses and the need for such sampling is not inherently limited to post-positivist qualitative approaches (Patton 2002).

Purposeful Sampling in Implementation Research

Characteristics of Implementation Research

In implementation research, quantitative and qualitative methods often play important roles, either simultaneously or sequentially, for the purpose of answering the same question through (a) convergence of results from different sources, (b) answering related questions in a complementary fashion, (c) using one set of methods to expand or explain the results obtained from use of the other set of methods, (d) using one set of methods to develop questionnaires or conceptual models that inform the use of the other set, or (e) using one set of methods to identify the sample for analysis using the other set of methods (Palinkas et al. 2011a). A review of mixed method designs in implementation research conducted by Palinkas et al. (2011a) revealed seven different sequential and simultaneous structural arrangements, five different functions of mixed methods, and three different ways of linking quantitative and qualitative data together. However, this review did not consider the sampling strategies involved in the types of quantitative and qualitative methods common to implementation research, nor did it consider the consequences of the sampling strategy selected for one method or set of methods on the choice of sampling strategy for the other method or set of methods. For instance, one of the most significant challenges to sampling in sequential mixed method designs lies in the limitations the initial method may place on sampling for the subsequent method. As Morse and Niehaus (2009) observe, when the initial method is qualitative, the sample selected may be too small and lack the randomization necessary to fulfill the assumptions required in a subsequent quantitative analysis. On the other hand, when the initial method is quantitative, the sample selected may be too large for each individual to be included in qualitative inquiry and lack purposeful selection or information necessary to reduce the sample size to one more appropriate for qualitative research. The fact that potential participants were recruited and selected at random does not necessarily make them information rich.

A re-examination of the 22 studies and an additional 6 studies published since 2009 revealed that only 5 studies (Aarons and Palinkas 2007; Bachman et al. 2009; Palinkas et al. 2011c; Palinkas et al. 2012; Slade et al. 2008) made a specific reference to purposeful sampling. An additional three studies (Henke et al. 2008; Proctor et al. 2007; Swain

et al. 2010) did not make explicit reference to purposeful sampling but did provide a rationale for sample selection. The remaining 20 studies provided no description of the sampling strategy used to identify participants for qualitative data collection and analysis; however, a rationale could be inferred based on a description of who were recruited and selected for participation. Of the 28 studies, 3 used more than one sampling strategy. Twenty-one of the 28 studies (75 %) used some form of criterion sampling. In most instances, the criterion used is related to the individual's role, either in the research project (i.e., trainer, team leader), or the agency (program director, clinical supervisor, clinician); in other words, criterion of inclusion in a certain category (criterion-i), in contrast to cases that are external to a specific criterion (criterion-e). For instance, in a series of studies based on the National Implementing EBPs Project, participants included semi-structured interviews with consultant trainers and program leaders at each study site (Brunette et al. 2008; Marshall et al. 2008; Marty et al. 2008; Rapp et al. 2010; Woltmann et al. 2008). Six studies used some form of maximum variation sampling to ensure representativeness and diversity of organizations and individual practitioners. Two studies used intensity sampling to make contrasts. Aarons and Palinkas (2007), for example, purposefully selected 15 child welfare case managers representing those having the most positive and those having the most negative views of SafeCare, an evidence-based prevention intervention, based on results of a web-based quantitative survey asking about the perceived value and usefulness of SafeCare. Kramer and Burns (2008) recruited and interviewed clinicians providing usual care and clinicians who dropped out of a study prior to consent to contrast with clinicians who provided the intervention under investigation. One study (Hoagwood et al. 2007), used a typical case approach to identify participants for a qualitative assessment of the challenges faced in implementing a trauma-focused intervention for youth. Another study (Green and Aarons 2011) used a combined snowball sampling/criterion-i strategy by asking recruited program managers to identify clinicians, administrative support staff, and consumers for project recruitment. County mental directors, agency directors, and program managers were recruited to represent the policy interests of implementation while clinicians, administrative support staff and consumers were recruited to represent the direct practice perspectives of EBP implementation.

Table 2 below provides a description of the use of different purposeful sampling strategies in mixed methods implementation studies. Criterion-i sampling was most frequently used in mixed methods implementation studies that employed a simultaneous design where the qualitative method was secondary to the quantitative method or studies that employed a simultaneous structure where the

Table 2 Purposeful sampling strategies and mixed method designs in implementation research

Sampling strategy	Structure	Design	Function
Single stage sampling (n = 22)			
Criterion (n = 18)	Simultaneous (n = 17)	Merged (n = 9)	Convergence (n = 6)
	Sequential (n = 6)	Connected (n = 9)	Complementarity (n = 12)
		Embedded (n = 14)	Expansion (n = 10)
Maximum variation (n = 4)	Simultaneous (n = 3)	Merged (n = 1)	Development (n = 3)
			Sampling (n = 4)
	Sequential (n = 1)	Connected (n = 1)	Convergence (n = 1)
			Embedded (n = 2)
Intensity (n = 1)	Simultaneous	Merged	Expansion (n = 1)
	Sequential	Connected	Development (n = 2)
		Embedded	Convergence
Typical case study (n = 1)	Simultaneous	Embedded	Complementarity
Multistage sampling (n = 4)			
Criterion/maximum variation (n = 2)	Simultaneous	Embedded	Convergence
	Sequential	Connected	Complementarity
Criterion/intensity (n = 1)	Simultaneous	Embedded	Development
			Convergence
			Complementarity
Criterion/snowball (n = 1)	Sequential	Connected	Expansion
			Convergence Development

qualitative and quantitative methods were assigned equal priority. These mixed method designs were used to complement the depth of understanding afforded by the qualitative methods with the breadth of understanding afforded by the quantitative methods (n = 13), to explain or elaborate upon the findings of one set of methods (usually quantitative) with the findings from the other set of methods (n = 10), or to seek convergence through triangulation of results or quantifying qualitative data (n = 8). The process of mixing methods in the large majority (n = 18) of these studies involved embedding the qualitative study within the larger quantitative study. In one study (Gioia and Dziadosz 2008), criterion sampling was used in a simultaneous design where quantitative and qualitative data were merged together in a complementary fashion, and in two studies (Aarons et al. 2012; Zazzali et al. 2008), quantitative and qualitative data were connected together, one in sequential design for the purpose of developing a conceptual model (Zazzali et al. 2008), and one in a simultaneous design for the purpose of complementing one another (Aarons et al. 2012). Three of the six studies that used maximum variation sampling used a simultaneous structure with quantitative methods taking priority over qualitative methods and a process of embedding the qualitative methods in a larger quantitative study

(Henke et al. 2008; Palinkas et al. 2011b; Slade et al. 2008). Two of the six studies used maximum variation sampling in a sequential design (Aarons et al. 2009; Zazzali et al. 2008) and one in a simultaneous design (Henke et al. 2008) for the purpose of development, and three used it in a simultaneous design for complementarity (Bachman et al. 2009; Henke et al. 2008; Palinkas et al. 2011b). The two studies relying upon intensity sampling used a simultaneous structure for the purpose of either convergence or expansion, and both studies involved a qualitative study embedded in a larger quantitative study (Aarons and Palinkas 2007; Kramer and Burns 2008). The single typical case study involved a simultaneous design where the qualitative study was embedded in a larger quantitative study for the purpose of complementarity (Hoagwood et al. 2007). The snowball/maximum variation study involved a sequential design where the qualitative study was merged into the quantitative data for the purpose of convergence and conceptual model development (Green and Aarons 2011). Although not used in any of the 28 implementation studies examined here, another common sequential sampling strategy is using criteria sampling of the larger quantitative sample to produce a second-stage qualitative sample in a manner similar to maximum variation sampling, except that the former

narrows the range of variation while the latter expands the range.

Criterion-i sampling, as a purposeful sampling strategy, shares many characteristics with random probability sampling, despite having different aims and different procedures for identifying and selecting potential participants. In both instances, study participants are drawn from agencies, organizations or systems involved in the implementation process. Individuals are selected based on the assumption that they possess knowledge and experience with the phenomenon of interest (i.e., the implementation of an EBP) and thus will be able to provide information that is both detailed (depth) and generalizable (breadth). Participants for a qualitative study, usually service providers, consumers, agency directors, or state policy-makers, are drawn from the larger sample of participants in the quantitative study. They are selected from the larger sample because they meet the same criteria, in this case, playing a specific role in the organization and/or implementation process. To some extent, they are assumed to be “representative” of that role, although implementation studies rarely explain the rationale for selecting only some and not all of the available role representatives (i.e., recruiting 15 providers from an agency for semi-structured interviews out of an available sample of 25 providers). From the perspective of qualitative methodology, participants who meet or exceed a specific criterion or criteria possess intimate (or, at the very least, greater) knowledge of the phenomenon of interest by virtue of their experience, making them information-rich cases.

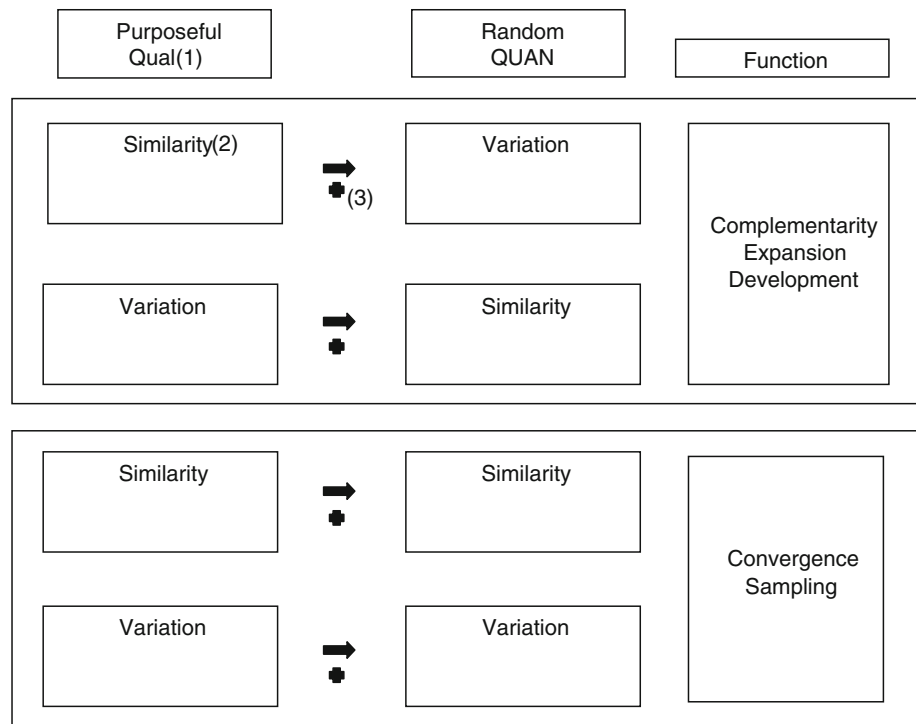
However, criterion sampling may not be the most appropriate strategy for implementation research because by attempting to capture both breadth and depth of understanding, it may actually be inadequate to the task of accomplishing either. Although qualitative methods are often contrasted with quantitative methods on the basis of depth versus breadth, they actually require elements of both in order to provide a comprehensive understanding of the phenomenon of interest. Ideally, the goal of achieving theoretical saturation by providing as much detail as possible involves selection of individuals or cases that can ensure all aspects of that phenomenon are included in the examination and that any one aspect is thoroughly examined. This goal, therefore, requires an approach that sequentially or simultaneously expands and narrows the field of view, respectively. By selecting only individuals who meet a specific criterion defined on the basis of their role in the implementation process or who have a specific experience (e.g., engaged only in an implementation defined as successful or only in one defined as unsuccessful), one may fail to capture the experiences or activities of other groups playing other roles in the process. For instance, a focus only on practitioners may fail to capture the insights,

experiences, and activities of consumers, family members, agency directors, administrative staff, or state policy leaders in the implementation process, thus limiting the breadth of understanding of that process. On the other hand, selecting participants on the basis of whether they were a practitioner, consumer, director, staff, or any of the above, may fail to identify those with the greatest experience or most knowledgeable or most able to communicate what they know and/or have experienced, thus limiting the depth of understanding of the implementation process.

To address the potential limitations of criterion sampling, other purposeful sampling strategies should be considered and possibly adopted in implementation research (Fig. 1). For instance, strategies placing greater emphasis on breadth and variation such as maximum variation, extreme case, confirming and disconfirming case sampling are better suited for an examination of differences, while strategies placing greater emphasis on depth and similarity such as homogeneous, snowball, and typical case sampling are better suited for an examination of commonalities or similarities, even though both types of sampling strategies include a focus on both differences and similarities. Alternatives to criterion sampling may be more appropriate to the specific functions of mixed methods, however. For instance, using qualitative methods for the purpose of complementarity may require that a sampling strategy emphasize similarity if it is to achieve depth of understanding or explore and develop hypotheses that complement a quantitative probability sampling strategy, achieving breadth of understanding and testing hypotheses (Kemper et al. 2003). Similarly, mixed methods that address related questions for the purpose of expanding or explaining results or developing new measures or conceptual models may require a purposeful sampling strategy aiming for similarity that complements probability sampling aiming for variation or dispersion. A narrowly focused purposeful sampling strategy for qualitative analysis that “complements” a broader focused probability sample for quantitative analysis may help to achieve a balance between increasing inference quality/trustworthiness (internal validity) and generalizability/transferability (external validity). A single method that focuses only on a broad view may decrease internal validity at the expense of external validity (Kemper et al. 2003). On the other hand, the aim of convergence (answering the same question with either method) may suggest use of a purposeful sampling strategy that aims for breadth that parallels the quantitative probability sampling strategy.

Furthermore, the specific nature of implementation research suggests that a multistage purposeful sampling strategy be used in many circumstances. Three different multistage sampling strategies are illustrated in Fig. 2 below. Several qualitative methodologists recommend sampling for variation (breadth) before sampling for

Fig. 1 Purposeful and random sampling strategies for mixed method implementation studies



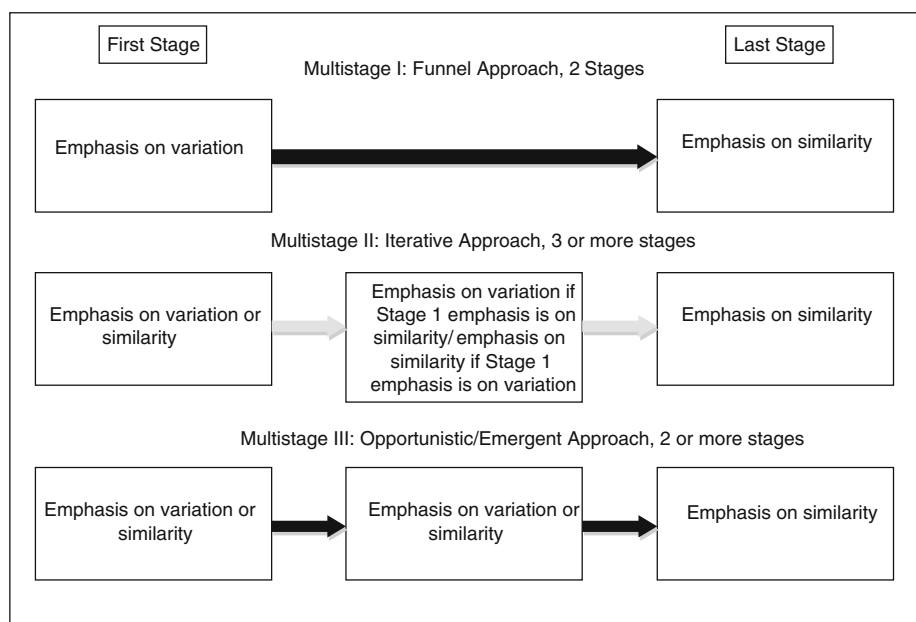
Legend:

(1) Priority and sequencing of Qualitative (QUAL) and Quantitative (QUAN) can be reversed.

(2) Refers to emphasis of sampling strategy.

(3) **➡** Refers to sequential structure; **➡** refers to simultaneous structure.

Fig. 2 Multistage purposeful sampling strategies



commonalities (depth) (Glaser 1978; Bernard 2002) (Multistage I). Also known as a “funnel approach”, this strategy is often recommended when conducting semi-structured interviews (Spradley 1979) or focus groups (Morgan 1997). This approach begins with a broad view of the topic and then proceeds to narrow down the conversation to very specific

components of the topic. However, as noted earlier, the lack of a clear understanding of the nature of the range may require an iterative approach where each stage of data analysis helps to determine subsequent means of data collection and analysis (Denzin 1978; Patton 2002) (Multistage II). Similarly, multistage purposeful sampling designs like

opportunistic or emergent sampling, allow the option of adding to a sample to take advantage of unforeseen opportunities after data collection has been initiated (Patton 2002, p. 240) (Multistage III). Multistage I models generally involve two stages, while a Multistage II model requires a minimum of 3 stages, alternating from sampling for variation to sampling for similarity. A Multistage III model begins with sampling for variation and ends with sampling for similarity, but may involve one or more intervening stages of sampling for variation or similarity as the need or opportunity arises.

Multistage purposeful sampling is also consistent with the use of hybrid designs to simultaneously examine intervention effectiveness and implementation. An extension of the concept of “practical clinical trials” (Tunis et al. 2003), effectiveness-implementation hybrid designs provide benefits such as more rapid translational gains in clinical intervention uptake, more effective implementation strategies, and more useful information for researchers and decision makers (Curran et al. 2012). Such designs may give equal priority to the testing of clinical treatments and implementation strategies (Hybrid Type 2) or give priority to the testing of treatment effectiveness (Hybrid Type 1) or implementation strategy (Hybrid Type 3). Curran et al. (2012) suggest that evaluation of the intervention’s effectiveness will require or involve use of quantitative measures while evaluation of the implementation process will require or involve use of mixed methods. When conducting a Hybrid Type 1 design (conducting a process evaluation of implementation in the context of a clinical effectiveness trial), the qualitative data could be used to inform the findings of the effectiveness trial. Thus, an effectiveness trial that finds substantial variation might purposefully select participants using a broader strategy like sampling for disconfirming cases to account for the variation. For instance, group randomized trials require knowledge of the contexts and circumstances similar and different across sites to account for inevitable site differences in interventions and assist local implementations of an intervention (Bloom and Michalopoulos 2013; Raudenbush and Liu 2000). Alternatively, a narrow strategy may be used to account for the lack of variation. In either instance, the choice of a purposeful sampling strategy is determined by the outcomes of the quantitative analysis that is based on a probability sampling strategy. In Hybrid Type 2 and Type 3 designs where the implementation process is given equal or greater priority than the effectiveness trial, the purposeful sampling strategy must be first and foremost consistent with the aims of the implementation study, which may be to understand variation, central tendencies, or both. In all three instances, the sampling strategy employed for the implementation study may vary based on the priority assigned to that study relative to the effectiveness trial. For

instance, purposeful sampling for a Hybrid Type 1 design may give higher priority to variation and comparison to understand the parameters of implementation processes or context as a contribution to an understanding of effectiveness outcomes (i.e., using qualitative data to expand upon or explain the results of the effectiveness trial). In effect, these process measures could be seen as modifiers of innovation/EBP outcome. In contrast, purposeful sampling for a Hybrid Type 3 design may give higher priority to similarity and depth to understand the core features of successful outcomes only.

Finally, multistage sampling strategies may be more consistent with innovations in experimental designs representing alternatives to the classic randomized controlled trial in community-based settings that have greater feasibility, acceptability, and external validity. While RCT designs provide the highest level of evidence, “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). Randomized trials are also “relatively poor in assessing the benefit from complex public health or medical interventions that account for individual preferences for or against certain interventions, differential adherence or attrition, or varying dosage or tailoring of an intervention to individual needs” (Brown et al. 2009, p. 2). Several alternatives to the randomized design have been proposed, such as “interrupted time series,” “multiple baseline across settings” or “regression-discontinuity” designs. Optimal designs represent one such alternative to the classic RCT and are addressed in detail by Duan et al. (this issue). Like purposeful sampling, optimal designs are intended to capture information-rich cases, usually identified as individuals most likely to benefit from the experimental intervention. The goal here is not to identify the typical or average patient, but patients who represent one end of the variation in an extreme case, intensity sampling, or criterion sampling strategy. Hence, a sampling strategy that begins by sampling for variation at the first stage and then sampling for homogeneity within a specific parameter of that variation (i.e., one end or the other of the distribution) at the second stage would seem the best approach for identifying an “optimal” sample for the clinical trial.

Another alternative to the classic RCT are the adaptive designs proposed by Brown et al. (2006, 2008, 2009). Adaptive designs are a sequence of trials that draw on the results of existing studies to determine the next stage of evaluation research. They use cumulative knowledge of current treatment successes or failures to change qualities of the ongoing trial. An adaptive intervention modifies what an individual subject (or community for a group-based trial) receives in response to his or her preferences or initial responses to an intervention. Consistent with

multistage sampling in qualitative research, the design is somewhat iterative in nature in the sense that information gained from analysis of data collected at the first stage influences the nature of the data collected, and the way they are collected, at subsequent stages (Denzin 1978). Furthermore, many of these adaptive designs may benefit from a multistage purposeful sampling strategy at early phases of the clinical trial to identify the range of variation and core characteristics of study participants. This information can then be used for the purposes of identifying optimal dose of treatment, limiting sample size, randomizing participants into different enrollment procedures, determining who should be eligible for random assignment (as in the optimal design) to maximize treatment adherence and minimize drop-out, or identifying incentives and motives that may be used to encourage participation in the trial itself.

Alternatives to the classic RCT design may also be desirable in studies that adopt a community-based participatory research framework (Minkler and Wallerstein 2003), considered to be an important tool on conducting implementation research (Palinkas and Soydan 2012). Such frameworks suggest that identification and recruitment of potential study participants will place greater emphasis on the priorities and “local knowledge” of community partners than on the need to sample for variation or uniformity. In this instance, the first stage of sampling may approximate the strategy of sampling politically important cases (Patton 2002) at the first stage, followed by other sampling strategies intended to maximize variations in stakeholder opinions or experience.

Summary

On the basis of this review, the following recommendations are offered for the use of purposeful sampling in mixed method implementation research. First, many mixed methods studies in health services research and implementation science do not clearly identify or provide a rationale for the sampling procedure for either quantitative or qualitative components of the study (Wisdom et al. 2011), so a primary recommendation is for researchers to clearly describe their sampling strategies and provide the rationale for the strategy.

Second, use of a single stage strategy for purposeful sampling for qualitative portions of a mixed methods implementation study should adhere to the same general principles that govern all forms of sampling, qualitative or quantitative. Kemper et al. (2003) identify seven such principles: (1) the sampling strategy should stem logically from the conceptual framework as well as the research questions being addressed by the study; (2) the sample

should be able to generate a thorough database on the type of phenomenon under study; (3) the sample should at least allow the possibility of drawing clear inferences and credible explanations from the data; (4) the sampling strategy must be ethical; (5) the sampling plan should be feasible; (6) the sampling plan should allow the researcher to transfer/generalize the conclusions of the study to other settings or populations; and (7) the sampling scheme should be as efficient as practical.

Third, the field of implementation research is at a stage itself where qualitative methods are intended primarily to explore the barriers and facilitators of EBP implementation and to develop new conceptual models of implementation process and outcomes. This is especially important in state implementation research, where fiscal necessities are driving policy reforms for which knowledge about EBP implementation barriers and facilitators are urgently needed. Thus a multistage strategy for purposeful sampling should begin first with a broader view with an emphasis on variation or dispersion and move to a narrow view with an emphasis on similarity or central tendencies. Such a strategy is necessary for the task of finding the optimal balance between internal and external validity.

Fourth, if we assume that probability sampling will be the preferred strategy for the quantitative components of most implementation research, the selection of a single or multistage purposeful sampling strategy should be based, in part, on how it relates to the probability sample, either for the purpose of answering the same question (in which case a strategy emphasizing variation and dispersion is preferred) or the for answering related questions (in which case, a strategy emphasizing similarity and central tendencies is preferred).

Fifth, it should be kept in mind that all sampling procedures, whether purposeful or probability, are designed to capture elements of both similarity and differences, of both centrality and dispersion, because both elements are essential to the task of generating new knowledge through the processes of comparison and contrast. Selecting a strategy that gives emphasis to one does not mean that it cannot be used for the other. Having said that, our analysis has assumed at least some degree of concordance between breadth of understanding associated with quantitative probability sampling and purposeful sampling strategies that emphasize variation on the one hand, and between the depth of understanding and purposeful sampling strategies that emphasize similarity on the other hand. While there may be some merit to that assumption, depth of understanding requires both an understanding of variation and common elements.

Finally, it should also be kept in mind that quantitative data can be generated from a purposeful sampling strategy and qualitative data can be generated from a probability

sampling strategy. Each set of data is suited to a specific objective and each must adhere to a specific set of assumptions and requirements. Nevertheless, the promise of mixed methods, like the promise of implementation science, lies in its ability to move beyond the confines of existing methodological approaches and develop innovative solutions to important and complex problems. For states engaged in EBP implementation, the need for these solutions is urgent.

Acknowledgments This study was funded through a Grant from the National Institute of Mental Health (P30-MH090322; K. Hoagwood, PI).

References

- Aarons, G. A., Green, A. E., Palinkas, L. A., Self-Brown, S., Whitaker, D. J., Lutzker, J. R., et al. (2012). Dynamic adaptation process to implement an evidence-based maltreatment intervention. *Implementation Science*, 7, 32.
- Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in child welfare. *Administration and Policy in Mental Health and Mental Health Services Research*, 38, 4–23.
- Aarons, G. A., & Palinkas, L. A. (2007). Implementation of evidence-based practice in child welfare: Service provider perspectives. *Administration and Policy in Mental Health and Mental Health Services Research*, 34, 411–419.
- Aarons, G. A., Wells, R., Zagursky, K., Fettes, D. L., & Palinkas, L. A. (2009). Implementing evidence-based practice in community mental health agencies: Multiple stakeholder perspectives. *American Journal of Public Health*, 99(11), 2087–2095.
- Bachman, M. O., O'Brien, M., Husbands, C., Shreeve, A., Jones, N., Watson, J., et al. (2009). Integrating children's services in England: National evaluation of children's trusts. *Child: Care, Health and Development*, 35, 257–265.
- Bernard, H. R. (2002). *Research methods in anthropology: Qualitative and quantitative approaches* (3rd ed.). Walnut Creek, CA: Alta Mira Press.
- Bloom, H. S., & Michalopoulos, C. (2013). When is the story in the subgroups? Strategies for interpreting and reporting intervention effects for subgroups. *Prevention Science*, 14, 179–188.
- Brown, C., Ten Have, T., Jo, B., Dagne, G., Wyman, P., Muthén, B., et al. (2009). Adaptive designs for randomized trials in public health. *Annual Review of Public Health*, 30, 1–25.
- Brown, C. H., Wang, W., Kellam, S. G., Muthén, B. O., & Prevention Science and Methodology Group. (2008). 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*, S95, S74–S104.
- Brown, C. H., Wyman, P. A., Guo, J., & Peña, J. (2006). Dynamic wait-listed designs for randomized trials: New designs for prevention of youth suicide. *Clinical Trials*, 3, 259–271.
- Brunette, M. F., Asher, D., Whitley, R., Lutz, W. J., Weider, B. L., Jones, A. M., et al. (2008). Implementation of integrated dual disorders treatment: A qualitative analysis of facilitators and barriers. *Psychiatric Services*, 59, 989–995.
- Cresswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed method research* (2nd ed.). Thousand Oaks, CA: Sage.
- Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care*, 50, 217–226.
- Denzen, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York: McGraw Hill.
- Duan, N., Bhaumik, D. K., Palinkas, L. A., & Hoagwood, K. (this issue). Purposeful sampling and optimal design. *Administration and Policy in Mental Health and Mental Health Services Research*.
- Gioia, D., & Dziadosz, G. (2008). Adoption of evidence-based practices in community mental health: A mixed method study of practitioner experience. *Community Mental Health Journal*, 44, 347–357.
- Glaser, B. G. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine de Gruyter.
- Glasgow, R., Magid, D., Beck, A., Ritzwoller, D., & Estabrooks, P. (2005). Practical clinical trials for translating research to practice: Design and measurement recommendations. *Medical Care*, 43(6), 551.
- Green, A. E., & Aarons, G. A. (2011). A comparison of policy and direct practice stakeholder perceptions of factors affecting evidence-based practice implementation using concept mapping. *Implementation Science*, 6, 104.
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82.
- Henke, R. M., Chou, A. F., Chanin, J. C., Zides, A. B., & Scholle, S. H. (2008). Physician attitude toward depression care interventions: Implications for implementation of quality improvement initiatives. *Implementation Science*, 3, 40.
- Hoagwood, K. E., Vogel, J. M., Levitt, J. M., D'Amico, P. J., Paisner, W. I., & Kaplan, S. J. (2007). Implementing an evidence-based trauma treatment in a state system after September 11: The CATS Project. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46(6), 773–779.
- Kemper, E. A., Stringfield, S., & Teddlie, C. (2003). Mixed methods sampling strategies in social science research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in the social and behavioral sciences* (pp. 273–296). Thousand Oaks, CA: Sage.
- Kramer, T. F., & Burns, B. J. (2008). Implementing cognitive behavioral therapy in the real world: A case study of two mental health centers. *Implementation Science*, 3, 14.
- Landsverk, J., Brown, H., Chamberlain, P., Palinkas, L. A., & Horwitz, S. M. (2012). Design and analysis in dissemination and implementation research. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Translating science to practice* (pp. 225–260). New York: Oxford University Press.
- Marshall, T., Rapp, C. A., Becker, D. R., & Bond, G. R. (2008). Key factors for implementing supported employment. *Psychiatric Services*, 59, 886–892.
- Marty, D., Rapp, C., McHugo, G., & Whitley, R. (2008). Factors influencing consumer outcome monitoring in implementation of evidence-based practices: Results from the National EBP Implementation Project. *Administration and Policy In Mental Health*, 35, 204–211.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Minkler, M., & Wallerstein, N. (Eds.). (2003). *Community-based participatory research for health*. San Francisco, CA: Jossey-Bass.
- Morgan, D. L. (1997). *Focus groups as qualitative research*. Newbury Park, CA: Sage.
- Morse, J. M., & Niehaus, L. (2009). *Mixed method design: Principles and procedures*. Walnut Creek, CA: Left Coast Press.
- Padgett, D. K. (2008). *Qualitative methods in social work research* (2nd ed.). Los Angeles: Sage.

- Palinkas, L. A., Aarons, G. A., Horwitz, S. M., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011a). Mixed method designs in implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*, 44–53.
- Palinkas, L. A., Ell, K., Hansen, M., Cabassa, L. J., & Wells, A. A. (2011b). Sustainability of collaborative care interventions in primary care settings. *Journal of Social Work*, *11*, 99–117.
- Palinkas, L. A., Holloway, I. W., Rice, E., Fuentes, D., Wu, Q., & Chamberlain, P. (2011c). Social networks and implementation of evidence-based practices in public youth-serving systems: A mixed methods study. *Implementation Science*, *6*, 113.
- Palinkas, L. A., Fuentes, D., Garcia, A. R., Finno, M., Holloway, I. W., & Chamberlain, P. (2012). Inter-organizational collaboration in the implementation of evidence-based practices among agencies serving abused and neglected youth. *Administration and Policy in Mental Health and Mental Health Services Research*. doi:10.1007/s10488-012-0437-5.
- Palinkas, L. A., & Soydan, H. (2012). *Translation and implementation of evidence-based practice*. New York: Oxford University Press.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Proctor, E. K., Knudsen, K. J., Fedoracivius, N., Hovmand, P., Rosen, A., & Perron, B. (2007). Implementation of evidence-based practice in community behavioral health: Agency director perspectives. *Administration and Policy in Mental Health*, *34*, 479–488.
- Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, C. (2009). 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*, *36*, 24–34.
- Rapp, C. A., Etzel-Wise, D., Marty, D., Coffman, M., Carlson, L., Asher, D., et al. (2010). Barriers to evidence-based practice implementation: Results of a qualitative study. *Community Mental Health Journal*, *46*, 112–118.
- Raudenbush, S., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, *5*, 199–213.
- Slade, M., Gask, L., Leese, M., McCrone, P., Montana, C., Powell, R., et al. (2008). Failure to improve appropriateness of referrals to adult community mental health services—Lessons from a multi-site cluster randomized controlled trial. *Family Practice*, *25*, 181–190.
- Spradley, J. P. (1979). *The ethnographic interview*. New York: Holt, Rinehart & Winston.
- Swain, K., Whitley, R., McHugo, G. J., & Drake, R. E. (2010). The sustainability of evidence-based practices in routine mental health agencies. *Community Mental Health Journal*, *46*, 119–129.
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in the social and behavioral sciences* (pp. 3–50). Thousand Oaks, CA: Sage.
- Tunis, S. R., Stryer, D. B., & Clancey, C. M. (2003). Increasing the value of clinical research for decision making in clinical and health policy. *Journal of the American Medical Association*, *290*(1624–1632), 2003.
- Wisdom, J. P., Cavaleri, M. C., Onwuegbuzie, A. T., & Green, C. A. (2011). Methodological reporting in qualitative, quantitative, and mixed methods health services research articles. *Health Services Research*, *47*, 721–745.
- Woltmann, E. M., Whitley, R., McHugo, G. J., et al. (2008). The role of staff turnover in the implementation of evidence-based practices in health care. *Psychiatric Services*, *59*, 732–737.
- Zazzali, J. L., Sherbourne, C., Hoagwood, K. E., Greene, D., Bigley, M. F., & Sexton, T. L. (2008). The adoption and implementation of an evidence based practice in child and family mental health services organizations: A pilot study of functional family therapy in New York State. *Administration and Policy in Mental Health*, *35*, 38–49.