Can sample size in qualitative research be determined *a priori*?

Julius Sim,[1]* Benjamin Saunders,[1] Jackie Waterfield,[2] Tom Kingstone[1]

[1] Institute for Primary Care and Health Sciences, Keele University, Staffordshire ST5 5BG, UK

[2] School of Health Sciences, Queen Margaret University, Edinburgh, UK

* corresponding author

Keywords: sampling, sample size, qualitative research

Abstract: There has been considerable recent interest in methods of determining sample size for qualitative research *a priori*, rather than through an adaptive approach such as saturation. Extending previous literature in this area, we identify four distinct approaches to determining sample size in this way: rules of thumb, conceptual models, numerical guidelines derived from empirical studies, and statistical formulae. Through critical discussion of these approaches, we argue that each embodies one or more questionable philosophical or methodological assumptions, namely: a naïve realist ontology; a focus on themes as enumerable 'instances', rather than in more conceptual terms; an incompatibility with an inductive approach to analysis; inappropriate statistical assumptions in the use of formulae; and an unwarranted assumption of generality across qualitative methods. We conclude that, whilst meeting certain practical demands, determining qualitative sample size *a priori* is an inherently problematic approach, especially in more interpretive models of qualitative research.

Introduction

Curtis et al (2000, p. 1002) note that sampling in qualitative research 'needs to be addressed rigorously and is fundamental to our understanding of the validity of qualitative research', but also suggest that it is a topic that has received insufficient attention in comparison to methods of data collection and analysis. Recently, however, the question of sample size in qualitative research has become a topical issue, with a number of papers debating whether the number of participants sampled in a given study should be decided upon *a priori*, and if so, the number of participants that is indicated. Clearly, there is often a perceived need to indicate sample size in advance of, or at the outset of, a qualitative research project. This may be in order to meet the demands of research funders or ethics committees, or simply to plan the resources for the study. Accordingly, a number of writers have addressed means of determining sample size for qualitative research *a priori* – particularly in relation to interview or focus group studies.

These attempts have at times, however, been challenged by those arguing that such *a priori* sample size decisions are incompatible with conceptual and methodological notions underpinning qualitative research. An edited collection by Baker and Edwards (2012) sought views from a range of researchers, primarily within the social sciences, as to the optimum sample size in qualitative research. Perhaps unsurprisingly, this was met with an overwhelming response of 'it depends', with a range contributing factors cited, including methodological considerations such as the nature and purpose of the individual study and the epistemological stance underpinning it, but also practical considerations around time and resources. This perhaps further highlights the competing priorities that are at play when making sampling decisions in qualitative research; i.e. the need to satisfy practical requirements of indicating sample size in advance, whilst at the same time seeking to adopt a

sampling approach that is in keeping with the methodological considerations pertinent to the particular study.

A more recent, and particularly spirited, discussion of these issues centred on a paper by Fugard and Potts (2015a), in which a statistical calculation of sample size for qualitative research is proposed. A back-and-forth debate ensued within the pages of this journal, with responses from Emmel (2015), Hammersley (2015), Byrne (2015), Braun and Clarke (2016) and Fugard and Potts (2015b, 2016). Prominent among these critiques was the argument that determining sample size *a priori* is inherently problematic in qualitative research, given that sample size is often adaptive and emergent, and – particularly if based on a grounded theory approach – adopts the principle of saturation. Saturation is operationalized in different ways (Saunders et al, 2017), but broadly speaking it guides data collection and/or analysis either in terms of 'informational redundancy' (Sandelowski, 2008, p.875) or in relation to the theoretical insights that develop as data accrue (O'Reilly & Parker 2013). In this way, sample size is decided *a posteriori*. For some (e.g. Fusch & Ness, 2015), saturation is an essential element within qualitative research, implying that sample size should always be determined by this means.

The present paper engages with the debate on these issues, but we adopt a broader critical overview of the literature, identifying four general approaches that have been taken towards determining qualitative sample size: 1. rules of thumb, based on a combination of methodological considerations and past experience; 2. conceptual models, based upon specific characteristics of the proposed study; 3. numerical guidelines derived from empirical investigation; and 4. statistical formulae, based on the probability of obtaining a sufficient sample size. We discuss these approaches in relation to the philosophical and methodological issues to which they give rise, ultimately arguing that each is underpinned by assumptions that are inherently problematic when applied to qualitative research, and in particular

inductive approaches in which themes[1] are iteratively developed through the researcher's ongoing engagement with the data. Through identifying, discussing and problematizing these varying approaches, this paper seeks to contribute novel insights to the ongoing discussion within the research literature, as well as potentially providing a resource for researchers who may be wrestling with decisions about sample size in their own qualitative studies.

Approaches to determining sample size

In this section, we describe in more detail – with examples – the four approaches alluded to above, as a basis for a critical, comparative discussion in subsequent sections.

Rules of thumb

A number of authors have proposed rules of thumb for sample size in qualitative research, based on methodological considerations and past experience with similar studies; this approach is reflected in one journal's policy on sample size for grounded theory studies (Dworkin, 2012). Some such recommendations are collated in Table 1.[2] These rules of thumb commonly lack a clear and detailed rationale, and whilst there is a degree of similarity in what they propose, there is also some diversity; for example, recommendations for grounded theory studies range from five to 35 and those for single case studies from four to 30.

[insert Table 1 about here]

Conceptual models

Some authors have used a rather more formal conceptual model, based upon specific characteristics of the proposed study, such as its aim, its underlying theoretical framework, and the type of analysis intended. Morse (2000), for example, argues that sample size will

depend upon: the scope of the research question (the broader the scope, the larger the sample size needed); the nature of the topic (the more 'obvious', the smaller the sample size); the quality of the data (the richer the data, the smaller the sample size); the study design (a longitudinal design in which a group is the unit of analysis will require a smaller sample size than one in which there is one interview per participant); and shadowed data (if interviews reveal something about others' perspectives, in addition to the interviewee's own, this may require a smaller sample size).

More recently, Malterud, Siersma and Guassora (2016) reason that sample size can be determined in relation to what they refer to as the 'information power' that a given sample holds.[3] This information power is influenced by: the aim of the study (the broader the aim, the greater the required sample size); the specificity of the sample (the more specific the characteristics of the participants in relation to the study aims, the smaller the sample size); the theoretical background (the less developed the underlying theory, the greater the sample size); the quality of dialogue (the richer the dialogue in the interviews, the smaller the sample size); and the analysis strategy (a study aiming for an exploratory cross-case analysis will require a larger sample size than one aiming for in-depth analysis of a few informants).

Numerical guidelines

A third, and seemingly popular, approach to sample size employs numerical guidelines derived from empirical investigation. An early and influential study by Guest, Bunce and Johnson (2006) used 60 interviews from a phenomenological study of West African women; the authors determined the degree of saturation as it occurred during a process of thematic analysis. Saturation was considered to occur within 12 interviews, a figure they suggest could be used for future studies. More recently, Guest, Namey and McKenna (2017) have published

a similar paper on focus groups, concluding that 80% of all themes would emerge within two to three groups, and 90% within three to six groups.

Francis et al (2010) reached conclusions on sample sizes required for saturation based on analysis of two previous studies. They defined an initial analysis sample of 10 interviews and a 'stopping criterion' for saturation of three, this criterion being defined as the number of additional interviews in which no new themes emerge. In the first study, of 14 interviews, they concluded that all of these interviews were required to achieve saturation, as the stopping criterion had not been met. In the second study, using the stopping criterion saturation was achieved after 15 of the 17 interviews. The authors are clear that their findings relate to 'theory-based content analysis', with *a priori* themes, and acknowledge that they may not apply to other approaches to analysis.

Hennink, Kaiser and Marconi (2017) also consider sample size in relation to saturation, distinguishing between what they term 'code saturation' (where no additional *issues* are identified) and 'meaning saturation' (where no further *insights* are gained). They examined 25 transcripts of semi-structured interviews, to determine the point at which both code and meaning saturation occurred. In terms of code saturation, most codes (53%) were identified in the first interview analysed, and 91% by the sixth interview. The point of meaning saturation was defined in terms of the last interview in which a novel code 'dimension' was identified. Meaning saturation was achieved earlier for codes representing concrete notions (normally by the ninth interview) whereas those representing more conceptual notions took longer (up to the 24th interview).

Extending Guest et al's (2006) work, Hagaman and Wutich (2017) focus on metathemes – which they define as themes that are cross-cultural rather than culture- or site-specific. They analysed data from interviews with 132 respondents from 4 sites, yielding 240

site-specific themes and 9 cross-cultural metathemes. Sixteen or fewer interviews were sufficient to identify common themes from sites with homogeneous groups, but 20–40 were needed to achieve saturation for metathemes. They specifically focused on themes across, rather than within, participants.

Finally, Ando, Cousins and Young (2014) examine saturation in the context of thematic analysis. In a two-stage process, codes were identified by inductive analysis of six interview transcripts (creating a codebook), and the resulting 39 codes, in seven themes, were then applied to the remaining 33 interview transcripts (resulting in some modifications to the number, definitions and labels of codes). The authors concluded that 12 interviews (six from the first stage and six from the second stage) provided all the themes and over 90% of codes, and recommended this as a sufficient sample size for thematic analysis with 'higher level concepts'.

Statistical formulae

At least five studies have used statistical formulae to determine sample size *a priori*, based on the probability of obtaining a sufficient sample size. A study that has promoted much discussion is that by Fugard and Potts (2015a), who present tables based on a binomial distribution[4] to show the minimum number of participants needed in order to detect, with a stated level of confidence (e.g. 80%), a given number of instances of a theme with an assumed prevalence in the population of interest.

Adopting a similar approach, Galvin (2015) utilizes the binomial distribution to i) calculate the probability that a theme with a given prevalence in the population will emerge at least once in a given number of interviews, or alternatively ii) calculate the number of interviews required for 95% probability of such a theme emerging at least once.

Van Rijnsoever (2015) uses simulations to determine the minimum number of informants (or, in his words, 'information sources') needed to achieve saturation. The addition of each informant is described as a 'sampling step'. Three scenarios are considered: *random chance* (where no prior information is used in successive sampling steps), *minimal information* (at each step, the researcher actively seeks an informant that will reveal at least one new code), and *maximum information* (at each step, the researcher seeks an informant that will lead to the largest possible increase in observed codes). This process is based upon knowledge of both the number of codes in the population and the probability of observing a code. If, for example, there are 101 codes in the population and the average probability of observing such a code is less than 10%, the random chance scenario required more than 1000 informants (in 95% of the simulations) to achieve saturation, whereas the minimal information and maximum information scenarios required 46 and 20 informants, respectively.

Finally, a study by Tran, Porcher, Falissard and Ravaud (2016) uses Monte-Carlo simulations[5] on an existing dataset of open-ended survey responses to determine the probability of identifying at least one new theme for each additional participant. A total of 150 participants allowed 92% of themes in the original study to be identified. A subsequent study from this group (Tran, Porcher, Tran & Ravaud, 2017) uses similar methodology to derive mathematical models to identify a 'point of saturation' that represents a desirable balance between the number of themes identified and the number of participants.[6]

Philosophical and methodological issues

The practical imperative to predetermine sample size *a priori* is understandable. However, it is important to consider the ontological and/or epistemological assumptions that are made in

the process and the methodological validity and coherence of strategies such as those outlined above.

Ontological status of a theme

Studies that rely on statistical formulae (e.g. Fugard & Potts, 2015a) or that otherwise arrive at predicted numbers of interviews at which saturation will occur (e.g. Francis et al, 2010) make a naïve realist assumption – i.e. that themes 'pre-exist' in participants' accounts, independently of the analyst, and are there to be discovered.[7] This emerges strongly in some of the language used. For example, van Rijnsoever (2015) suggests that a population will 'contain' codes and that information is 'extracted' from informants. He also refers to the possibility of codes 'becoming observed' or, alternatively, of being 'missed', similarly suggesting that codes have an objective existence, prior to the process of data collection. Fugard and Potts (2015a) – as Braun and Clarke (2016) point out – use similar language when referring to 'capturing' a theme.

This ontological assumption sits uneasily with perspectives that are more commonly associated with qualitative research, such as subtle realism (Hammersley, 1992) and constructionism (Gergen, 2001; Best, 2008). Subtle realism accepts that social phenomena and processes are 'real' in terms of having an independent existence from those who observe them, and are thereby in principle 'knowable', but denies that they are directly accessible through observation. Instead, they emerge from the engagement of the researcher with the data, and are therefore mediated by the researcher's prior cultural and theoretical understanding. Constructionism, meanwhile, takes a more clearly anti-realist stance, arguing that 'all claims to the "real" are traced to processes of relationship, and there is no extra-cultural means of ultimately privileging one construction of reality over another' (Gergen, 2001, p. 8). Accordingly, on this view social phenomena have no objective existence in the

external world, but are co-produced by analyst and informant. In neither perspective is a theme something to be 'discovered'.

These statistical and numerical approaches similarly assume that occurrences of a theme are fungible – they are deemed equivalent and interchangeable, regardless of the particular informant's account from which they may be generated, rather than as deriving their meaning and importance at least partly from the individuality of a person's account and the context in which it is set. In this way, data are decontextualized from participants' specific experiences or biographies, contrary to the spirit and mission of much qualitative research.

Linked to this, as Byrne (2015) points out in relation to Fugard and Potts' (2015a) paper, is a view of a theme as being in some sense an attribute of an individual. A one-to-one relationship appears to be posited between the participant and the theme, much as a respondent might be represented by a value on a particular variable in a more quantitative approach to research. Similarly, when Tran et al (2016, p. 91) suggest that 'enriching an initial sample with patients representing different characteristics would improve the number of themes identified,' the implication is that particular themes are directly generated by specific, identifiable characteristics of members of the sample, or possibly by definable subgroups of the sample. This understanding seems closer to a survey approach to research than a form of qualitative research in which participants exemplify, but are not reducible to, more abstract theoretical constructs, and in which such constructs are not generated by, or otherwise 'tied' to, individual attributes. In a similar way, any approach that attempts to determine sampling *a priori* in terms of the number of participants is liable to overemphasize the individual as the relevant unit. As Sandelowski (1995) points out, sample size can be thought of in terms of the number of *events*, *incidents* and *experiences*, not solely in terms of the number of *participants*. Accordingly, her recommendations regarding sample size make

reference to the number of descriptions of an experience rather than simply the number of informants. Adaptive sampling decisions based on saturation can accommodate these differing types of data, but strategies that predetermine a sample size purely in terms of the number of individual informants would struggle to do so.

Themes as 'instances'

When it is specified that a number of occurrences of a theme can be identified with a given level of confidence (e.g. Fugard & Potts, 2015a) or that a given number of interviews would be required in order to achieve saturation (e.g. Guest et al, 2006), an assumption is made that how a theme is defined is constant throughout the study, and that on this basis occurrences of a theme can be enumerated in a straightforward manner. However, what is considered to be an 'instance' of a theme might change over the process of data collection and analysis in line with changes in the researcher's understanding of the underlying theoretical concept that occur as insights develop. Furthermore, the extent to which a particular theme is considered to be important or relevant may change as data accumulate during a study (Kerr, Nixon & Wild, 2010; Emmel, 2013; Hammersley, 2015).

Accumulating instances of a theme do not, therefore, contribute information or insights to the researcher in a consistent or linear way. Accordingly, an attempt to predict the point of saturation cannot be tied down to the number of interviews in which a theme occurs, as it will also depend on the way in which, and the degree to which, each occurrence of a theme in successive interviews informs and modifies a deeper conceptual understanding of the theme concerned. As Becker (2012, p.15) notes, 'the number of interviews you need will change from day to day as you learn more and revise your ideas'. Hammersley (2015, p. 688) also points out that the qualitative researcher's focus is often not so much on *how many* informants are sampled, but on *which* informants are sampled, in terms of the 'fruitful

development of the emerging theory.' Hence, the development of a nuanced analysis depends upon how accumulating occurrences (and accompanying variations) of a theme assist in providing theoretical insights. Some accounts will contribute more to this process than others, and aiming only for a *number* of such accounts is therefore inadequate as a basis for determining sample size.

Statistical approaches to sample size calculation, and those such as Francis et al's (2010) that are centred on the empirical determination of saturation, focus explicitly on the *number* of instances of a theme. In the first case, the focus is on the number of participants required to identify *at least one* occurrence of a theme, whilst in the second it is on the number of interviews required in order to identify *no additional* such occurrences. Themes are thereby conceptualized only at the level of instances, rather than in the context of a broader analytical framework in which the concern would be not only with enumerable instances of a theme, but also with the way in which meaning is developed within a theme, and through the relationship between themes. For example, taking a grounded theory approach in a study of chronic pain, Howell (1994) describes the progressive development of analytical categories through the process of analysis, from more descriptive lower-order categories to more abstract higher-order categories. In contrast, statistical approaches appear to operate solely at the level of lower-order, descriptive themes and do not take account of the further development or elaboration of higher-order themes during analysis. Hence, the more that themes are regarded as holding interpretive, rather than descriptive, meaning, the less applicable will be an approach to sample size conceived in terms of instances.[8]

A further implication of statistical and empirical approaches to sample size is that the possibility of identifying instances of a theme is taken to be the same across members of the sample. If a specific number of participants is stated in advance of data collection in order to identify a theme, or a particular number of themes, with a stated degree of probability, it is

assumed that an occurrence of each theme can potentially be derived from each participant's account. However, this assumption will not necessarily apply in qualitative research. It is quite possible that some lower-order themes will be potentially identifiable in the accounts of some participants but may be extremely unlikely to occur in others. For example, in a study of physical activity and chronic pain, Richardson, Moore, Bernard, Jordan and Sim (2015) interviewed individuals who had either no pain, pain that did not interfere with their lives, or pain that did interfere. Themes relating to physical activity, social activity and a notion of 'involvement' were more commonly exemplified in the accounts of informants reporting pain without interference than in those informants reporting pain with interference. Only if one assumes a certain commonality of experience or perspective can one regard all participants as potentially contributing to a pre-identified theme.[9]

The analytical context

One of the most striking methodological aspects of many attempts to determine sample size *a priori* is their reliance on a deductive approach to analysis; i.e. one that relies wholly or predominantly on applying pre-identified themes to the data, rather than allowing these to emerge inductively.[10] Indeed, Francis et al (2010) are clear that their method is intended to apply to theory-driven themes.[11] This approach may lend itself, at least partially, to a method such as framework analysis (Ritchie & Spencer 1994), in which some or all themes are determined ahead of data analysis, or more broadly any approach in which a coding scheme is generated in the first few interviews and then applied to the analysis of subsequent interviews, as in Guest et al's (2006) and Ando et al's (2014) studies. However, attempts to pre-specify sample size – in other than a purely pragmatic way – in advance for a study in which analysis is intended to be inductive is essentially contradictory. If an understanding of the number and the nature of themes proceeds *pari passu* with the process of analysis, one

cannot meaningfully predict at the outset the point at which this understanding will be adequate.

The conceptual models of Morse (2000) and Malterud et al (2016), which stress the particular characteristics of individual studies as crucial to adequate sample size, avoid many of the shortcomings or questionable assumptions associated with the more empirical approaches. Thus, Malterud et al (2016) reject the use of formulae to calculate sample size in qualitative research, and acknowledge that initial assessments of sample size should be revisited during a study. However, there are other respects in which this conceptual approach is questionable. A basic premise of Malterud et al's (2016, p. 1754) model is that 'tools to guide sample size should not rely on procedures from a specific analysis method, but rest on shared methodological principles'. First, this claim is contestable. As a method of analysis will be at least partly determined by the nature of the research question, which will in turn be based on a particular philosophical perspective, it is likely that sample size *will* be related to particular analytical procedures. For example, adopting an approach based on interpretive phenomenological analysis, with its focus on cases (Smith, Flowers & Larkin, 2009), may have very different implications for sample size from one based on qualitative content analysis (Granaheim and Lundman, 2004; Schreier, 2014), or one based on conversation analysis (Grundy, 2008; Clift, 2016). This point can be illustrated further by focusing on Morse's (2000) discussion of shadowed data as a determinant of sample size. Shadowed data occur when:

> '[i]n addition to talking about their own experience, participants may discuss the experience of others, how their experience resembles or differs from others, and why… The use of shadowed data… provides the investigator with some idea of the range of experiences and the domain of the phenomena beyond the single person's individual experience, and it provides some explanation of the rationale for these differences'. (Morse, 2000, p. 4)

The way in which, and the extent to which, such shadowed data will influence sample size is not straightforward, and will depend upon the specific approach to analysis that is undertaken. If, in the context of an ethnographic study, analysis is directed at understanding how individuals contribute to and partake of a particular culture or set of social practices, there is a sense (albeit limited, perhaps) in which one person's account may tell us something about the experiences of others and the way in which these differ, and thereby influence the sample size. In contrast, an analysis founded on more phenomenological principles may simply regard an informant's discussion of others' experiences as elaborating that individual's own particular perspective, rather than providing insight into what others experience or perceive. The effect on sample size is very different in each case. In the ethnographic approach, shadowed data may affect sample size in terms of the number of different perspectives that are represented in the data, whereas in the phenomenological approach the effect on sample size is mediated through the richness of the data obtained from an individual informant.

Second, the insistence on 'shared methodological principles' suggests a form of methodological unity that sits uneasily within qualitative research. Malterud et al (2016) seem to assume that the various dimensions of 'information power', and hence of sample size, operate in a uniform and predictable way. No doubt there is a *tendency* for a study with broad aims to require a larger sample than one with narrower aims, but is this a matter of necessity? Might it not have to do with the *nature* of these aims, rather than just their breadth? It is argued that 'a researcher who never challenges his or her participant runs the risk of developing empirical data holding low information power, which, during analysis, only reproduces what is known from before' (Malterud et al, 2016, p. 1756). This seems to be a very broad claim, and assumes a false orthodoxy in interviewing. In a narrative approach to interviewing, for example, a technique of challenging might be inimical, rather than

conducive, to obtaining insightful data. For example, it would sit uneasily within the 'receptive' approach to narrative interviewing described by Wengraf (2009), which is characterized by minimal intervention on the part of the researcher, so as to permit the informant to unfold his or her story freely and spontaneously. Here, following a single initial question, any interventions by the researcher are 'effectively limited to facilitative noises and non-verbal support' (Wengraf, 2009, p. 113). This non-directive style of narrative interview is also reflected in Chase's (2005) characterization of the research relationship as one of narrator and listener, rather than informant and interviewer.

The way in which adequacy of sample size is conceptualized can also suggest an approach to analysis that has little resonance with qualitative research. The focus on instances, discussed earlier, encourages an undue emphasis on a process of enumeration in qualitative data analysis. Admittedly, authors such as Silverman (1985) and Dey (1993) have argued convincingly that deriving theoretical insights from qualitative data may at times depend upon some form of basic quantification. Dey (1993, p. 179), for example, maintains that 'it is difficult to see how, in practice, it is possible to identify associations between categories or to assess the strength of relationships without recourse to a numerical evaluation'. However, it is hard to see this type of enumeration as playing more than an ancillary role in qualitative data analysis. As argued earlier, a meaningful analysis will ultimately depend upon the nature and meaning of concepts expressed in the data, not their prevalence, frequency or typicality. Moreover, it is dangerous to equate the number of instances straightforwardly with some measure of analytical importance, as van Rijnsoever (2015, p. 12) appears to when suggesting that 'to enhance the credibility of the research, it is possible to aim deliberately for a minimum number of observations of each code.'

Diversity of the participants

Some of the rules of thumb in Table 1 refer to the notion of homogeneity, suggesting that this may influence the required sample size. It is indeed reasonable to think that sampling in qualitative research needs to take account of the differing characteristics, or the diversity of experience, in those who are the focus of the study, and that the greater their supposed heterogeneity, the larger the number of participants that should be included. On occasions, this way of thinking may inform *a priori* judgments. A study intending to explore how the experience of an illness is gendered may, for example, foresee analytical comparisons between men's and women's perspectives and determine the number of informants accordingly. Equally, and perhaps more likely, a sense of heterogeneity may emerge from participants' accounts during the process of data analysis, and may appropriately inform the number of participants sampled in line with the principles of theoretical sampling (Glaser & Strauss, 1967).

In both cases, however, what underlies this consideration is not some notion of empirical representativeness – as might be the case in survey research – but rather a concern 'to generate a full range of variation in the set of descriptions to be used in analysing a phenomenon' (Polkinghorne, 1989, p. 48). The homogeneity of the sample does not, therefore, drive sample size in the same way as in quantitative research, where the principal concern is with minimizing random sampling error (Barnett, 2002).

Determinants of sample size as interdependent

Malterud et al's (2016) and Morse's (2000) models present the various determinants of sample size as if they operate independently and thus summatively. Hence, Malterud et al (2016, p. 1756) argue:

> A study will *need the least amount of participants* when the study aim is narrow, if the combination of participants is highly specific for the study aim, if it is supported by established theory, if the interview dialogue is strong, and if the analysis includes longitudinal in-depth exploration of narratives or discourse details.[12]

It is, however, conceivable that in some instances these determinants will interact. Hence, the factors influencing sample size may well not be independent, and the extent to which one particular dimension is present in a study will influence the extent to which some or all of the others influence the required sample size. Morse (2000, pp. 3–4) considers separately the effect of the scope of the study ('the broader the scope of the research question, the longer it will take to reach saturation') and the effect of the topic of the study ('If the topic being studied is obvious and clear, and the information is easily obtained in the interviews, then fewer participants are needed than if the topic is below the surface and intriguing and difficult to grab'). However, a study whose scope is broad may conceivably address either an obvious topic or one that is below the surface, as may one whose scope is narrower. The effect of the scope of the study cannot therefore be considered separately from that of the topic – and how, and to what extent, this occurs cannot be readily judged in advance of undertaking the study. Consequently, it may not be realistic to rely on generic determinants of sample size – except as very provisional guidance or as 'springboards' for one's thinking (Bryman, 2012, p. 19) – but instead take a more particularistic focus on the specific characteristics in the study at hand, and the way in which these inter-relate.

Statistical assumptions

If we turn to the approaches that employ statistical methods to calculate sample size, these make a number of specific assumptions. First, as part of the probabilistic model they assume that occurrences of a theme are statistically independent – i.e. that the identification of a

theme in the account of one informant does not influence, and is not influenced by, its occurrence in any other informant. This is unlikely to be the case. As suggested earlier, within a study a theme may be more likely to arise from the accounts of certain informants than of certain other informants. For example, in a study conducted within a particular community or organization, particular relationships or shared experiences involving a number of participants may suggest that the accounts of these particular individuals will evoke similar concepts. The statistical basis of the sample size calculation will be affected accordingly. In a similar way, if a process of snowball sampling is adopted – as commonly occurs in qualitative research (Noy, 2008) – the use of social networks or prior relationships to identify additional informants almost guarantees that themes will not be independent. Additionally, interest may centre on how a theme is repeatedly exemplified by an individual informant's account, not merely with the presence of themes across accounts – these occurrences *within* an informant will similarly not be independent.

Second, these statistical calculations rely upon a posited underlying probability distribution, such as the binomial distribution. So, if a theme is considered to have a prevalence of 15% in the population, and the researcher wishes to be 95% confident of finding at least one occurrence of this theme, the following formula (where 'ln' denotes natural logarithm) will provide the minimum number of informants required (Viechtbauer et al, 2015):

$$n = \frac{\ln(1-.95)}{\ln(1-.15)} = 19$$

A consequence of this is that the sample of informants in any study based on such a calculation should reflect this hypothesized probability distribution. If, however, the actual prevalence differs from that assumed, the calculation of the minimum number of informants will not be valid – e.g. if the prevalence of the theme is actually 10% rather than 15%, the

appropriate number of informants would be 29. Furthermore, using a hypothesized distribution requires a random sampling procedure. Not only is random sampling impracticable in most qualitative research, but a notable drawback is that such a prescribed method also rules out any form of sampling – such as purposive (Mason, 2002) or theoretical (Glaser & Strauss, 1967) sampling – that adjusts the selection of cases adaptively during analysis in relation to developing insights or other emergent features of the study.

Directly linked to this is a third assumption: that the prevalence of a theme (e.g. Fugard & Potts, 2015a; Galvin, 2015) and/or the number of themes (e.g. van Rijnsoever, 2015) can be estimated at the outset. This is probably only possible with clearly defined *a priori* themes within a deductive framework, and certainly not in an inductive approach to analysis. In the latter, not only are themes liable to be unknown at the outset, but any that are known will be subject to refinement or re-definition as the study proceeds, largely precluding any prior sense of their prevalence.

An assumption of generality

Statistical approaches to sample size assume that the basis on which the calculation is made – e.g. the prevalence of a theme – should be transferable to the context of a new study. Moreover, when Fugard and Potts (2015a, p. 671) note that 'for any given study, the saturation point may vary, making planning difficult', they see this as a problem to be overcome, rather than as a reflection of the individuality of specific studies.

This notion of generality also underlies some of the other approaches that have been considered. Establishing guidelines on the basis of past experience with 'similar' studies runs the risk of assuming a false homogeneity among studies, even with the same methodological or analytical tradition. Studies carried out from the perspective of interpretative phenomenological analysis, for example, will clearly have something in common, but there is

a danger of overstating their similarity and overlooking the unique features of individual studies, in response to which an apparently equivalent methodology may take quite a different form or emphasis. Similarly, empirical studies, such as those by Ando et al (2014), Francis et al (2010) and Guest et al (2006, 2016), are conducted in particular contexts, with specific samples: Ando et al (2014) interviewed individuals with multiple sclerosis, whilst Guest et al (2006) used data from sex workers in Africa. The generality of the figures they derive for sample size may therefore be limited (as the authors often admit), and this undercuts their usefulness in relation to future studies in which the topic, context or population of the study may differ. Moreover, this is not just a matter of generalizability. Ignoring the unique features of a particular study goes against the contextual orientation of most qualitative research.

Discussion

It is not difficult to understand the desire of many researchers undertaking qualitative research to pursue methods for determining sample size *a priori*, given requirements of funding bodies, the practicalities of managing time and resources, and perhaps, amongst some, a desire to reduce the perceived 'messiness' of subjective judgements about sampling by introducing a degree of standardization. We have critiqued four main approaches to determining sample size identified in the literature. We have argued that what appears common to all of these different approaches are problematic assumptions of uniformity across different qualitative methods.

Both statistical approaches to determining sample size and empirically-derived guidelines appear to suggest a degree of uniformity in the identification of a theme, in terms of the assumption of a direct relationship between the number of participants and the number

of identified instances of a theme. In such approaches, the number of instances of a theme appears to be most significant, rather than the theoretical insights these instances offer in relation to the theme concerned. We have argued that such models appear more suited to deductive approaches to analysis in which themes are predefined, and appear somewhat incompatible with more exploratory, inductive approaches in which sampling decisions are guided by a principle such as saturation.

Conceptual model approaches such as that proposed by Malterud et al (2016) are perhaps more suited to an inductive methodology – they argue that judgments on sample size 'should be stepwise revisited along the research process and not definitely decided in advance' (2016, p. 1757). However, the model they propose does not elaborate on this idea, and an assessment of the 'information power' of a sample from the initial point of designing the study does appear to place more emphasis on *a priori* decision-making, even if such decisions are not set in stone. Malterud et al (2016) also make the questionable assumption of a degree of standardization across qualitative methods – that there is a set of shared principles spanning different study designs and analytic frameworks. Additionally, like statistical formulae and empirically-derived guidelines, the notion of 'information power' appears, albeit implicitly, to adopt a realist assumption that data are somehow extracted from participants, suggesting the incremental gaining of objective information – an epistemological stance we have proposed is at odds with approaches that consider themes as being *developed* as part of an ongoing, interpretive analysis. Many of these approaches also assume a questionable homogeneity across studies – even within an identified methodological approach, as in the case of approaches based on rules of thumb – leading to an excessively generalist approach to sample size recommendations or calculations.

Researchers – particularly those working at the interpretive end of the spectrum of qualitative research – face a genuine challenge when seeking to justify sample sizes to

funding bodies or ethical review committees. However, resorting to predictions or calculations that rest upon questionable assumptions, or that import inappropriate methodological or statistical principles from quantitative research, is an ill-advised response to this challenge. A preferable approach would be, perhaps, to address the pragmatic necessity of providing *some* indication of sample size by presenting an approximate, and very provisional, anticipated upper limit, without presenting an unwarranted empirical basis for such a figure or attempting precise predictions of a point of saturation, but with a clear caveat that a firm judgment of the number of participants ultimately required can only be reached once the study is underway. This can be further justified by including a clear explanation of how an adaptive approach to sample size fits with the inductive methodology adopted in the study, as well as an outline of how sample size decisions will be made during the course of the study – for example, the way in which saturation will be assessed and demonstrated in the research – thereby providing a rationale underpinned by clear methodological reasoning.

Whilst in this paper we have focused principally on sample size in relation to interview studies – reflecting the majority of literature in this field – we might propose that similar arguments be put to funding bodies and ethics committees regarding other methods of qualitative data collection; for instance, observation, diaries and visual methods. Although the nature of sampling decisions may be very different across these methods, much the same logic must hold that sampling decisions should be principally driven by methodological considerations over and above practical imperatives. Closer attention to sampling decisions in relation to these other types of qualitative data could be a fruitful area for further research.

Conclusion

On this evidence, we believe that defining sample size *a priori* is inherently problematic in the case of inductive, exploratory research, which, by definition, looks to explore phenomena in relation to which we cannot identify the key themes in advance. In such an approach, specifying *a priori* how many participants will be needed to create sufficient understanding of what is as yet unknown is, in its essence, illogical (Saunders et al, 2017). Thus, whilst we acknowledge the practical imperative to give a rough estimation of sample size at the beginning of a study, we argue that the decision over what constitutes an adequate sample size to meet a study's aims is one that is necessarily a process of ongoing interpretation by the researcher. It is an iterative, context-dependent decision made during the analytical process as the researcher begins to develop an increasingly comprehensive picture of the developed themes, the relationship between these themes, and where the conceptual boundaries of these themes lie.

Moreover, although there is clearly a need to make a decision on the number of participants in a study by one means or another – whether through *a priori* determination or a more adaptive approach such as saturation – there is also a need to ensure that the whole issue of sample size does not assume a disproportionate prominence and overshadow other essential elements with the process of qualitative data collection and analysis. As Emmel (2013, p. 154) reminds us, 'it is not the number of cases that matters, it is what you do with them that counts.'

[1] We use this term broadly, to embrace 'codes', 'categories', and similar terms.

[2] Some such recommendations relate not to the number of informants but to the number of interviews with an individual informant. For example, Spradley (1979:51) recommends at least six to seven one-hour interviews for an ethnographic study.

[3] Although these authors indicate that their model applies to the planning of a study, it is not solely focussed on the prior determination of sample size; they note that the adequacy of the sample size should be continuously reassessed during a study.

[4] The binomial distribution is a probability distribution used for binary variables, i.e. those in which an observation can take one of two possible values, such as 'present' versus 'absent'.

[5] Monte Carlo simulations estimate the sampling distribution of a particular statistic by drawing numerous random samples from a stimulated population of values (Mooney, 2004).

[6] More specifically, a theme accumulation curve was constructed, such that a value of 0.05 for this curve (i.e. one new theme for each 20 additional participants) allowed 97.5% of themes to be identified whilst limiting the inclusion of further participants who would not yield further themes; this was proposed as a possible stopping criterion for sampling.

[7] Interestingly, an early example of this approach (Romney, Weller and Batchelder 1986) focused on calculating the number of participants in terms of investigating their *knowledge* – the truth or falsity of their responses to specific questions. Moving from this to matters of belief or experience, on which qualitative research characteristically focuses, is questionable.

[8] Hagaman and Wutich (2017) are, however, explicit that their themes are descriptive.

[9] Higher-order themes, owing to their greater theoretical abstraction and the fact that they are likely to subsume a number of lower-order themes, might be more readily assumed to encompass most or all accounts. However, these numerical and statistical approaches to sample size tend to focus on lower-order themes.

[10] The identification of themes by an inductive process is most commonly associated with grounded theory (Glaser and Strauss, 1967). Initially, the process of analysis in grounded theory is indeed inductive, but as theoretical categories evolve, a more abductive logic is employed whereby instances of data are related to the theoretical category with which they best fit (Charmaz, 2009), thereby allowing these categories to be further developed and refined. Accordingly, we are using 'inductive' in a broad rather than a narrow sense.

[11] Curiously, despite proposing that codes can be foreknown in terms of their number and probability of occurrence, van Rijnsoever (2015) sets his approach within an inductive approach, aligned with the principles of grounded theory.

[12] Malterud et al (2016:1754) acknowledge that the determinants may have a 'mutual impact on each other', but this is not explicated within the model that they propose.

Table 1. Rules of thumb for sample size

Author(s)	Recommendation
Adler and Adler (2012)	Recommend a 'broad range of between a dozen and 60, with 30 being the mean' (p. 10).
Bernard (2000, 2013)	Bernard (2000) recommends 30–60 for ethnographic studies. Elsewhere (Bernard, 2013, p. 175), he suggests that '10–20 knowledgeable people are enough to uncover and understand the core categories in any well-defined cultural domain or study of lived experience'.
Boddy (2005)	Suggests an upper limit of 12 focus groups or 30 in-depth interviews, if researching a relatively homogeneous population.
Creswell (2013)	Recommends 20 to 30 informants for a grounded theory study, 4 to 5 cases per study for case study research, and 2 or 3 for narrative research.
Dukes (1984)	Recommends 3 to 10 participants in a phenomenological study.
Kuzel (1999)	Recommends 5 to 8 participants in a homogeneous sample (or homogeneous subgroups within a sample) or 12 to 20 if looking for disconfirming evidence or to achieve maximum variation sampling.
Lincoln and Guba (1985 p235)	Recommend between 12 and 20 participants in interview studies.
Marshall, Cardon, Poddar and Fontenot (2013)	Recommend 20 to 30 interviews for grounded theory studies and 15 to 30 for single case studies.
Morse (1994, 2000)	Morse (1994) recommends at least 6 participants for phenomenological studies and approximately 35 for studies based in grounded theory, ethnography and ethnoscience. Morse (2000) gives similar figures, pointing out that the number of interviews per informant is important, as well as the number of informants.
Onwuegbuzie and Leech (2007)	Do not recommend specific numbers, but advocate examining sample sizes in previous studies of similar design in which saturation was reached and using a figure within the range of such sample sizes.
Parse (2009)	Recommends 2 to 10 participants in order to achieve 'redundancy or saturation' (p. 10).
Ray (1994)	Suggests that phenomenological studies usually focus on a group of between 8 and 12 people, but may also focus on just a single person (p. 127).
Ritchie, Lewis, Elam, Tennant and Rahim (2014)	Advise an upper limit of 50 for interview studies, on the basis that the quality of data collection and analysis may suffer with larger studies (p. 118).
Smith, Flowers and Larkin (2009)	Advise between 3 and 10 for studies based on interpretative phenomenological analysis, but indicate that the appropriate sample size depends on a number of factors specific to the study concerned, including level of study for student work (undergraduate, postgraduate).

Corbin and Strauss (2015)	Propose at least five one-hour interviews for theoretical saturation in grounded theory studies.
Warren (2002)	Suggests that the norm in terms of the minimum number of interviews is in the range of 20–30 in order to 'have a nonethnographic interview study published' (p. 99).

References

Adler, P.A., & Adler, P. (2012). How many qualitative interviews is enough? In S.E. Baker & R. Edwards (Eds.) *How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research* (pp. 8–11). Southampton: ESRC National Centre for Research Methods, University of Southampton.

Ando, H., Cousins, R., & Young, C. (2014). Achieving saturation in thematic analysis: development and refinement of a codebook. *Comprehensive Psychology, 3*, 1–7. doi: 10.2466/03.CP.3.4

Baker, S.E., & Edwards, R. (2012). *How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research*. Southampton: ESRC National Centre for Research Methods, University of Southampton.

Barnett, V. (2002). *Sample survey: Principles and methods* (3rd ed). London: Arnold.

Becker, H.S. (2012). How many qualitative interviews is enough? In S.E. Baker & R. Edwards (Eds.) *How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research* (p. 15). Southampton: ESRC National Centre for Research Methods, University of Southampton.

Bernard, H.R. (2000). *Research methods in anthropology: Qualitative and quantitative approaches* (3rd ed). Lanham, MD: AltaMira Press.

Bernard, H.R. (2013). *Social research methods: Qualitative and quantitative approaches* (2nd ed). Thousand Oaks, CA: Sage.

Best, J.D. (2008). Historical development and defining issues of constructionist research. In J.F. Holstein & J.A. Gubrium (Eds.) *Handbook of constructionist research* (pp. 41–64). New York, NY: Guilford Press.

Boddy, C.R. (2005) Sample size for qualitative research. *Qualitative Market Research 19,* 426–432. doi: 10.1108/QMR-06-2016-0053

Braun, V., & Clarke, V. (2016). (Mis)conceptualising themes, thematic analyses, and other problems with Fugard and Potts' (2015) sample-size tool for thematic analysis. *International Journal of Social Research Methodology, 19*, 739–743. doi: 10.1080/13645579.2016.1195588

Bryman, A. (2012). How many qualitative interviews is enough? In S.E. Baker & R. Edwards (Eds.) *How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research* (pp. 18–19). Southampton: ESRC National Centre for Research Methods, University of Southampton.

Byrne, D. (2015). Response to Fugard and Potts: supporting thinking on sample sizes for thematic analyses: a quantitative tool. *International Journal of Social Research Methodology, 18*, 689–691. doi: 10.1080/13645579.2015.1005455

Charmaz, K. (2007). Shifting the grounds: constructivist grounded theory methods. In J.M. Morse, P.N. Stern, J. Corbin, B. Bowers, K. Charmaz & A.E. Clarke (Eds.) *Developing grounded theory: The second generation* (pp. 127–154)*.* Walnut Creek, CA: Left Coast Press.

Chase, S.E. (2005). Narrative inquiry: multiple lenses, approaches, voices. In N.K. Denzin & Y.S. Lincoln (Eds.) *The Sage handbook of qualitative research* (3rd ed.) (pp. 651–679). Thousand Oaks, CA: Sage.

Clift, R. (2016). *Conversation analysis*. Cambridge: Cambridge University Press.

Corbin J., & Strauss A. (2015) *Basics of qualitative research: Techniques and principles for developing grounded theory*. Thousand Oaks, CA: Sage.

Creswell, J.W. (2013). *Qualitative inquiry & research design: Choosing among five approaches* (3rd ed). Thousand Oaks, CA: Sage.

Curtis, S., Gesler, W., Smith, G., & Washburn, S. (2000). Approaches to sampling and case selection in qualitative research: examples in the geography of health. *Social Science & Medicine,* 50, 1001–1014. doi: 10.1016/S0277-9536(99)00350-0

Dey, I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists*. London: Routledge.

Dukes, S. (1984). Phenomenological methodology in the human sciences. *Journal of Religion and Health,* 23, 197–203. doi: 10.1007/BF00990785

Dworkin, S.L. (2012). Sample size policy for qualitative studies using in-depth interviews. *Archives of Sexual Behavior, 41*, 1319–1320. doi: 10.1007/s10508-012-0016-6

Emmel, N. (2013). *Sampling and choosing cases in qualitative research: A realist approach*. London: Sage.

Emmel, N. (2015). Themes, variables, and the limits to calculating sample size in qualitative research: a response to Fugard and Potts. *International Journal of Social Research Methodology, 18*, 685–686. doi: 10.1080/13645579.2015.1005457

Francis, J.J., Johnston, M., Robertson, C., Glidewell, L., Entwhistle, V., Eccles, M.P., & Grimshaw, J.M. (2010). What is an adequate sample size? Operationalising data saturation for theory-driven interview studies. *Psychology and Health, 25*, 1229–1245. doi: 10.1080/08870440903194015

Fugard, A.J.B., & Potts, H.W.W. (2015a). Supporting thinking on sample sizes for thematic analyses: a quantitative tool. *International Journal of Social Research Methodology, 18*, 669–684. doi: 10.1080/13645579.2015.1005453

Fugard, A.J.B., & Potts, H.W.W. (2015b). Response to comments. *International Journal of Social Research Methodology, 18*, 693–694. doi: 10.1080/13645579.2015.1005454

Fugard, A.J.B., & Potts, H.W.W. (2016). 'Shine bright like a diamond'? A reply to Braun and Clarke. *International Journal of Social Research Methodology, 19*, 745–746. doi: 10.1080/13645579.2016.1205794

Fusch, P.I., & Ness, L.R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report, 20*, 1408–1416. http://www.nova.edu/ssss/QR/QR20/9/fusch1.pdf

Galvin, R. (2015). How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge? *Journal of Building Engineering, 1*, 2–12. doi: 10.1016/j.jobe.2014.12.001

Gergen, K.J. (2001). *Social construction in context*. London: Sage.

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.

Granaheim, U.K., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24*, 105–122. doi: 10.1016/j.nedt.2003.10.001

Grundy, P. (2008). *Doing pragmatics* (3rd ed). London: Hodder-Arnold.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*, 59–82. doi: 10.1177/1525822X05279903

Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field Methods, 29*, 3–22. doi: 10.1177/1525822X16639015

Hagaman, A.K., & Wutich, A. (2017). How many interviews are enough to identify metathemes in multisited and cross-cultural research? Another perspective on Guest, Bunce and Johnson's (2006) landmark study. *Field Methods, 29*, 23–41. doi: 10.1177/1525822X16640447

Hammersley, M. (1992). *What's wrong with ethnography?* London: Routledge.

Hammersley, M. (2015). Sampling and thematic analysis: a response to Fugard and Potts. *International Journal of Social Research Methodology, 18*, 687–688. doi: 10.1080/13645579.2015.1005456

Hennink, M.M., Kaiser, B.N., & Marconi, V.C. (2017). Code saturation versus meaning saturation: how many interviews are enough? *Qualitative Health Research,* 27, 591–608. doi: 10.1177/1049732316665344

Howell, S.L. (1994). A theoretical model for caring with women with chronic non-malignant pain. *Qualitative Health Research, 4*, 94–122. doi: 10.1177/104973239400400107

Kerr, C., Nixon, A., & Wild, D. (2010). Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *Expert Reviews in Pharmacoeconomics and Outcomes Research, 10*, 269–281. doi: 10.1586/erp.10.30

Kuzel, A.J. (1999). Sampling in qualitative research. In B.F Crabtree & W.L. Miller (Eds.) *Doing Qualitative Research*, (2nd ed). (pp. 33–45). Thousand Oaks, CA: Sage.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

Malterud, K., Siersma, V.D., & Guassora, A.D. (2016). Sample size in qualitative interview studies: guided by information power. *Qualitative Health Research, 26*, 1753–1760. doi: 10.1177/1049732315617444

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research? A review of qualitative interviews in IS research. *Journal of Computer Information Systems, 54*, 11–22. doi: 10.1080/08874417.2013.11645667

Mason, J. (2002). *Qualitative researching* (2nd ed). London: Sage.

Mooney, C.Z. (2004). Monte Carlo simulation. In M.S. Lewis-Beck, A. Bryman & T.F. Liao (Eds.) *The Sage encyclopedia of social science research methods, volume 2* (pp. 662–664). Thousand Oaks, CA: Sage.

Morse, J. (1994). Designing funded qualitative research. In N. Denzin & Y. Lincoln (Eds.) *Handbook of qualitative research* (pp. 220–235). Thousand Oaks, CA: Sage.

Morse, J.M. (2000). Determining sample size. *Qualitative Health Research, 10*, 3–5. doi: 10.1177/104973200129118183

Noy, C. (2008). Sampling knowledge: the hermeneutics of snowball sampling in qualitative research. *International Journal of Social Research Methodology,*11, 327–344. doi: 10.1080/13645570701401305

Onwuegbuzie, A.J., & Leech, N.L. (2007). A call for qualitative power analyses. *Quality and Quantity* 41, 105–121. doi: 10.1007/s11135-005-1098-1

O'Reilly, M., & Parker, N. (2013). 'Unsatisfactory saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research, 13*, 190–197. doi: 10.1177/1468794112446106

Parse, R.R. (1990). Parse's research methodology with an illustration of the lived experience of hope. *Nursing Science Quarterly* 3, 9–17. doi: 10.1177/089431849000300106

Polkinghorne, D.E. (1989). Phenomenological research methods. In R.S. Valle & S. Halling (Eds.) *Existential-phenomenological perspectives in psychology: Exploring the breadth of human experience* (pp. 41–60). New York, NY: Plenum Press.

Ray, M.A. (1994). The richness of phenomenology: philosophic, theoretic and methodologic concerns. In J.M. Morse (Ed.) *Critical issues in qualitative research methods* (pp. 117–133). Thousand Oaks, CA: Sage.

Richardson, J.C., Moore, A.J., Bernard, M., Jordan, K.L., & Sim, J. (2015). Living well with chronic pain in later life: the role and meaning of activity and involvement. *Activities, Adaptation and Aging 39*, 200–213. doi: 10.1080/01924788.2015.1063329

Ritchie, J., Lewis, J., Elam, G., Tennant, R., & Rahim, N. (2014). Designing and selecting samples. In J. Ritchie, J. Lewis, C. McNaughton Nicholls & R. Ormston (Eds.) *Qualitative research practice: A guide for social science students and researchers* (2nd ed). (pp. 111–146). London: Sage.

Ritchie, J., & Spencer, L. (1994). Qualitative data analysis for applied policy research. In A. Bryman & R.G. Burgess (Eds.) *Analyzing qualitative data* (pp. 173–194). London: Routledge.

Romney, A.K., Weller, S.C., & Batchelder, W.H. (1986). Culture as consensus: a theory of culture and informant accuracy. *American Anthropologist, 88*, 313–338. doi: 10.1525/aa.1986.88.2.02a00020

Sandelowski, M. (1995). Sample size in qualitative research. *Research in Nursing and Health, 18*, 179–183. doi: 10.1002/nur.4770180211

Sandelowski, M. (2008). Theoretical saturation. In L.M. Given (Ed.) *The SAGE encyclopedia of qualitative research methods, volume 2* (pp. 875–876). Thousand Oaks, CA: Sage.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2017). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality and Quantity,* doi: 10.1007/s11135-017-0574-8

Schreier, M. (2014). Qualitative content analysis. In U. Flick (Ed) *The Sage handbook of qualitative data analysis* (pp. 170–183). London: Sage.

Silverman, D. (1985). *Qualitative methodology and sociology: Describing the social world*. Aldershot: Gower.

Smith, J.A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*. London: Sage.

Spradley, J.P. (1979). *The ethnographic interview*. Belmont, CA: Wadsworth.

Tran, V.-T., Porcher, R., Falissard, B., & Ravaud, P. (2016). Point of data saturation was assessed using resampling methods in a survey with open-ended questions. *Journal of Clinical Epidemiology, 80*, 88–96. doi: 10.1016/j.jclinepi.2016.07.014

Tran, V.-T., Porcher, R., Tran, V.-C., & Ravaud, P. (2017). Predicting data saturation in qualitative surveys with mathematical models from ecological research. *Journal of Clinical Epidemiology, 82*, 71–78. doi: 10.1016/j.jclinepi.2016.10.001

van Rijnsoever, F. (2015). *(I can't get no) saturation: a simulation and guidelines for minimum sample sizes in qualitative research. Research Innovation Studies Utrecht Working Paper Series 15.05*. Utrecht: University of Utrecht.

Viechtbauer, W., Smits, L., Kotz, D., Budé, L., Spigt, M., Serroyen, J., & Crutzen, R. (2015). A simple formula for the calculation of sample sizes in pilot studies. *Journal of Clinical Epidemiology, 68,* 1375–1379. doi: 10.1016/j.jclinepi.2015.04.014

Warren, C.A. (2002). Qualitative interviewing. In J.F. Gubrium & J.A. Holstein (Eds.) *Handbook of interview research: Context and method* (pp 83–101). Thousand Oaks, CA: Sage.

Wengraf, T. (2009). *Qualitative research interviewing: Biographic narrative and semi-structured methods*. London: Sage.