

Sample Size and Saturation in PhD Studies Using Qualitative Interviews

Mark Mason

Key words:
saturation; sample
size; interviews

Abstract: A number of issues can affect sample size in qualitative research; however, the guiding principle should be the concept of saturation. This has been explored in detail by a number of authors but is still hotly debated, and some say little understood. A sample of PhD studies using qualitative approaches, and qualitative interviews as the method of data collection was taken from theses.com and contents analysed for their sample sizes. Five hundred and sixty studies were identified that fitted the inclusion criteria. Results showed that the mean sample size was 31; however, the distribution was non-random, with a statistically significant proportion of studies, presenting sample sizes that were multiples of ten. These results are discussed in relation to saturation. They suggest a pre-meditated approach that is not wholly congruent with the principles of qualitative research.

Table of Contents

- [1. Introduction](#)
 - [1.1 Factors determining saturation](#)
 - [1.2 Guidelines for sample sizes in qualitative research](#)
 - [1.3 Operationalising the concept of saturation](#)
 - [1.4 The issue of saturation in PhDs](#)
- [2. Method](#)
- [3. Results](#)
- [4. Discussion](#)
- [5. Conclusion](#)
- [References](#)
- [Author](#)
- [Citation](#)

1. Introduction

Samples for qualitative studies are generally much smaller than those used in quantitative studies. RITCHIE, LEWIS and ELAM (2003) provide reasons for this. There is a point of diminishing return to a qualitative sample—as the study goes on more data does not necessarily lead to more information. This is because one occurrence of a piece of data, or a code, is all that is necessary to ensure that it becomes part of the analysis framework. Frequencies are rarely important in qualitative research, as one occurrence of the data is potentially as useful as many in understanding the process behind a topic. This is because qualitative research is concerned with meaning and not making generalised hypothesis statements (see also CROUCH & MCKENZIE, 2006). Finally, because qualitative research is very labour intensive, analysing a large sample can be time consuming and often simply impractical. [1]

Within any research area, different participants can have diverse opinions. Qualitative samples must be large enough to assure that most or all of the perceptions that might be important are uncovered, but at the same time if the sample is too large data becomes repetitive and, eventually, superfluous. If a researcher remains faithful to the principles of qualitative research, sample size in the majority of qualitative studies should generally follow the concept of saturation (e.g. GLASER & STRAUSS, 1967)—when the collection of new data does not shed any further light on the issue under investigation. [2]

While there are other factors that affect sample size in qualitative studies, researchers generally use saturation as a guiding principle during their data collection. This paper examines the size of the samples from PhD studies that have used interviews as their source of data collection. It does not look at the data found in those studies, just the numbers of the respondents in each case. [3]

1.1 Factors determining saturation

While saturation determines the majority of qualitative sample size, other factors that can dictate how quickly or slowly this is achieved in a qualitative study. CHARMAZ (2006) suggests that the aims of the study are the ultimate driver of the project design, and therefore the sample size. She suggests that a small study with "modest claims" (p.114) might achieve saturation quicker than a study that is aiming to describe a process that spans disciplines (for example describing drug addiction in a specific group rather than a description of general addiction). [4]

Other researchers have also elucidated further supplementary factors that can influence a qualitative sample size, and therefore saturation in qualitative studies. RITCHIE et al. (2003, p.84) outline seven factors that might affect the potential size of a sample:

"the heterogeneity of the population; the number of selection criteria; the extent to which 'nesting' of criteria is needed; groups of special interest that require intensive study; multiple samples within one study; types of data collection methods use; and the budget and resources available". [5]

To this, MORSE (2000, p.4) adds, "the scope of the study, the nature of the topic, the quality of the data, the study design and the use of what MORSE calls "shadowed data". [6]

JETTE, GROVER and KECK (2003) suggested that expertise in the chosen topic can reduce the number of participants needed in a study—while LEE, WOO and MACKENZIE (2002) suggest that studies that use more than one method require fewer participants, as do studies that use multiple (very in-depth) interviews with the same participant (e.g. longitudinal or panel studies). [7]

Some researchers have taken this a step further and tried to develop a debate on the concept of saturation. MORSE (1995) feels that researchers often claim to have achieved saturation but are not necessarily able to prove it. This is also

suggested by BOWEN (2008) who feels that saturation is claimed in any number of qualitative research reports without any overt description of what it means or how it was achieved. To this end, CHARMAZ (2006) gives the example of a researcher studying stigma in obese women. It is entirely possible that a researcher will claim that the category "experiencing stigma" is saturated very quickly. However, while an inexperienced researcher might claim saturation, a more experienced researcher would explore the context of stigma in more detail and what it means to each of these women (p.114). [8]

According to DEY (1999), the concept of saturation is inappropriate. He suggests that researchers often close categories early as the data are only partially coded, and cite others to support this practice, such as and STRAUSS and CORBIN (1998 [1990]) who suggest that saturation is a "matter of degree" (p.136). They suggest that the longer researchers examine, familiarise themselves and analyse their data there will always be the potential for "the new to emerge". Instead, they conclude that saturation should be more concerned with reaching the point where it becomes "counter-productive" and that "the new" is discovered does not necessarily add anything to the overall story, model, theory or framework (p.136). They admit that sometimes the problem of developing a conclusion to their work is not necessarily a lack of data but an excess of it. As the analysis begins to take shape it is important for the researcher to become more disciplined and cut data where necessary. [9]

1.2 Guidelines for sample sizes in qualitative research

As a result of the numerous factors that can determine sample sizes in qualitative studies, many researchers shy away from suggesting what constitutes a sufficient sample size (in contrast to quantitative studies for example). However, some clearly find this frustrating. GUEST, BUNCE and JOHNSON (2006, p.59) suggest, "although the idea of saturation is helpful at the conceptual level, it provides little practical guidance for estimating sample sizes for robust research prior to data collection". During the literature search for the background to their study they found "only seven sources that provided guidelines for actual sample sizes" (p.61):

- Ethnography and ethnoscience: MORSE (1994, p.225) 30-50 interviews for both; BERNARD (2000, p.178) states that most studies are based on samples between 30-60 interviews for ethnoscience;
- grounded theory methodology: CRESWELL (1998, p.64) 20-30; MORSE (1994, p.225) 30-50 interviews.
- phenomenology: CRESWELL (1998, p.64) five to 25; MORSE (1994, p.225) at least six;
- all qualitative research: BERTAUX (1981, p.35) fifteen is the smallest acceptable sample (adapted from GUEST et al., 2006). [10]

While these numbers are offered as guidance the authors do not tend to present empirical arguments as to why these numbers and not others for example. Also

the issue of why some authors feel that certain methodological approaches call for more participants compared to others, is also not explored in any detail. [11]

Further to this, other researchers have tried to suggest some kind of guidelines for qualitative sample sizes. CHARMAZ (2006, p.114) for example suggests that "25 (participants are) adequate for smaller projects"; according to RITCHIE et al. (2003, p.84) qualitative samples often "lie under 50"; while GREEN and THOROGOOD (2009 [2004], p.120) state that "the experience of *most qualitative researchers* (emphasis added) is that in interview studies little that is 'new' comes out of transcripts after you have interviewed 20 or so people". [12]

While some researchers offer guidelines for qualitative samples, there is evidence that suggests others do not strictly adhere to them. THOMSON (2004) for example carried out a review of fifty research articles accessed using Proquest ABI Inform¹, with the search parameter "grounded theory" in citation and abstract, and found sample sizes ranging from five to 350. Just over a third (34%) used samples between CRESWELL's suggested range of 20 and 30 (1998, p.128)—while only 11 studies (or 22%) used samples in MORSE's range of over 30 (1994, p.225). [13]

1.3 Operationalising the concept of saturation

There is an obvious tension between those who adhere to qualitative research principles, by not quantifying their samples—and those who feel that providing guidance on sample sizes is useful. Some researchers have gone further than providing guidelines and have tried to operationalise the concept of saturation, based on their own empirical analysis. [14]

Possibly the first to attempt this were ROMNEY, BATCHELDER and WELLER (1986) who developed an analysis tool called the "Cultural Consensus Model" (CCM) for their ethnographic work. This sought to identify common characteristics between communities and cultural groups. The model suggests that each culture has a shared view of the world, which results in a "cultural consensus"—the level of consensus of different topics does vary but there are considered to be a finite set of characteristics or views. ROMNEY et al. suggest these views can then be factor analysed to produce a rigorous model of the cultures views on that topic. The subsequent analysis tool has also been used by some to estimate a minimum sample size—recently for example by ATRAN, MEDIN and ROSS (2005, p.753) who suggested that in some of their studies "as few as 10 informants were needed to reliably establish a consensus". [15]

GRIFFIN and HAUSER (1993) reanalysed data from their own study into customers of portable food containers. Using a model developed by VORBERG and ULRICH (1987) they examined the number of customer needs uncovered by various numbers of in-depth interviews and focus groups. Their work was

1 http://il.proquest.com/en-US/catalogs/databases/detail/abi_inform_complete.shtml [Accessed: May 24, 2010]. Proquest AB Inform is an electronic/online database which provides access to publications principally in the areas of: Industry, Business, Finance and Management.

undertaken from a market research perspective to assist in the development of robust bids and campaigns. Because of their analysis, they hypothesized that twenty to thirty in-depth interviews would be needed to uncover ninety to nine-five per cent of all customer needs. [16]

Most recently, GUEST et al. (2006) carried out a systematic analysis of their own data from a study of sixty women, involving reproductive health care in Africa. They examined the codes developed from their sixty interviews, in an attempt to assess at which point their data were returning no new codes, and were therefore saturated. Their findings suggested that data saturation had occurred at a very early stage. Of the thirty six codes developed for their study, thirty four were developed from their first six interviews, and thirty five were developed after twelve. Their conclusion was that for studies with a high level of homogeneity among the population "a sample of six interviews may [be] sufficient to enable development of meaningful themes and useful interpretations" (p.78). [17]

1.4 The issue of saturation in PhDs

GREEN and THOROGOOD (2009 [2004]) agree with GUEST et al., and feel that while saturation is a convincing concept, it has a number of practical weaknesses. This is particularly apparent in what they call "funded work" (or that limited by time). They suggest that researchers do not have the luxury of continuing the sort of open-ended research that saturation requires. This is also true when the point of saturation (particularly in relation to an approach like grounded theory methodology, which requires that all of the properties and the dimensions are saturated) they consider to be "potentially limitless" (p.120). They go on to add that sponsors of research often require a thorough proposal that includes a description of who, and how many people, will be interviewed at the outset of the research (see also SIBLEY, 2003). They further suggest that this also applies to ethics committees, who will want to know who will be interviewed, where, and when, with a clearly detailed rationale and strategy. This is no less relevant to PhD researchers. [18]

A reading of the application requirements of some of the World's top 50 Universities² suggests it is not uncommon for universities to require applicants to explicitly document their intended sample size, prior to registration. The University of Toronto for example (ranked 29th in 2009), requires prospective students of PhD research programmes to "[j]ustify the anticipated sample size and its representativeness. For example, how many documents, interviews, focus groups will be consulted/undertaken and why?"³ Further to this, University College Dublin (ranked 43rd in 2009) requires prospective students to give "an indication of the feasibility of the proposed project including some indication of sample size and selection"⁴. [19]

2 <http://www.timeshighereducation.co.uk/hybrid.asp?typeCode=438> [Accessed: May 24, 2010].

3 http://www.phs.utoronto.ca/ssh_phd_thesis_guidelines.asp [Accessed: September 15, 2009].

4 [http://www.ucd.ie/t4cms/phd%20proposal%20guidelinesjan%2025%20\(2\).doc](http://www.ucd.ie/t4cms/phd%20proposal%20guidelinesjan%2025%20(2).doc) [Accessed: September 15, 2009].

This also appears to trouble current postgraduate students. When STRAUSS and CORBIN redrafted their "Basics of qualitative research" (1998 [1990]), they included what they considered twenty of the most frequently asked questions in their classes and seminars—question sixteen is, "How many interviews or observations are enough? When do I stop gathering data?" The answer they give once again outlines the concept of saturation but finishes with only a reiteration of the concept, suggesting that there are constraints (including time, energy, availability of participants etc.): "Sometimes the researcher has no choice and must settle for a theoretical scheme that is less developed than desired" (p.292). [20]

As an example of how this issue affects current postgraduate students, a brief search of the prominent Internet forum PostgraduateForum.com⁵ found at least three live discussion threads, specifically set up to debate and discuss the number of research participants required for their studies: "How many qual research interviews?"⁶; "No. of participants"⁷; and "How many qualitative (sic.) interviews"⁸. [21]

The PhD is probably the one time that a researcher (often mature and in the middle of his/her career) should get to examine a subject in a great deal of detail, over the course of a number of years. Throughout the supervisory process the study is scrutinised by national, and often international, experts, and once completed, the methodology and findings scrutinised further. If a high level of rigour is to be found in the types of methods used in research studies then it should be in PhDs. [22]

With this in mind it was decided to examine the issue of sample size in the context of PhDs studies. The following research questions were developed to explore this issue:

- How many participants are used in PhD studies utilising qualitative interviews? And do these numbers vary depending on the methodological approach? [23]

2. Method

A content analysis of a PhD database was undertaken on the website: "Index To Theses: A comprehensive listing of theses with abstracts accepted for higher degrees by universities in Great Britain and Ireland since 1716"⁹ ("the only comprehensive published listing of British theses accepted annually for higher

5 <http://www.postgraduateforum.com/> [Accessed: September 15, 2009].

6 <http://www.postgraduateforum.com/threadViewer.aspx?TID=4763> [Created: August 15, 2006, 16:19:25, accessed: September 15, 2009].

7 <http://www.postgraduateforum.com/threadViewer.aspx?TID=6664> [Created: May 2, 2007, 21:28:30, accessed: September 15, 2009].

8 <http://www.postgraduateforum.com/threadViewer.aspx?TID=8099> [Created: November 28, 2007, 10:07:41, accessed: September 15, 2009].

9 <http://www.theses.com/> [Accessed between August 3, 2009 and August 24, 2009].

degrees by some of the most prestigious educational institutions in the world; the Universities of Great Britain and Ireland"¹⁰). [24]

Searching was undertaken between 3rd August 2009 and the 24th August 2009 (on 532,646 abstracts in the collection; last updated 2 July 2009, volume 58, 3rd update of 8) to identify PhD studies which stated they had used qualitative (i.e. structured; semi-structured or unstructured) interviews as a method of data collection. [25]

To explore any differences between diverse research approaches a categorisation of 26 different qualitative research approaches from diverse disciplines was used (TESCH, 1990). While studying qualitative research software TESCH found 26 different types of qualitative methodological tradition and categorised them into four groups: the characteristics of language, the discovery of regularities, the comprehension of the meaning of text or action, and reflection. [26]

A "standard search" was used, with the following parameters applied: "Any field" contains "INSERT METHODOLOGY (e.g. "Grounded Theory"¹¹); "Any field" contains "interviews"; and "Degree" contains "PhD". The following criteria were used to exclude cases:

- Abstracts that did not state the exact number of interviews (i.e. studies where the author stated that "over fifty interviews were undertaken", for example, were excluded.
- Abstracts that stated that the author had been part of a fieldwork team were excluded.
- Abstracts that specified more than one interview for one participant were excluded (i.e. repeat interviews, longitudinal studies or panel studies).
- Abstracts from other professional qualifications such as PhDs in clinical psychology (DClinPsy¹²), for example, where single client case studies are prevalent, were excluded. [27]

This was intended to provide consistent criteria, that meant only studies that explicitly detailed the actual number of people interviewed once as part of the work, are included. Also, this study looks only at the use of one to one personal interviewing, and as such, the use of focus groups is not included in this analysis. [28]

The remaining studies were collected into the sample. The abstracts were searched and the following details recorded from each:

10 http://www.nationalschool.gov.uk/policyhub/evaluating_policy/magenta_book/key-data.asp
[Accessed: May 24, 2010].

11 All areas of qualitative research were entered separately.

12 Doctorate in Clinical Psychology

- number of participants interviewed;
- methodological approach used; and
- category of qualitative research. [29]

3. Results

Table 1 shows the results from the analysis. It provides the number of studies identified from each research approach (as identified by TESCH, 1990), and the number of studies which made up the study sample, when the inclusion criteria were applied. It further provides the highest number of participants/interviews used for that approach along with the measures of central dispersion. Finally, it identifies how many of these studies utilised interviews as the only method, and provides this figure as a percentage of the overall sample.

	No. of studies found	No. of studies after inclusion criteria applied	Range		Measures of central dispersion			
			High	Low	Mode	Mean	Median	St. Dev.
Action research	140	28	67	3	6	23	17	18.4
Case study	1401	179	95	1	40	36	33	21.1
Collaborative research	8	2	25	5	-	15	15	14.1
Content analysis	213	42	70	2	30	28	25	14.7
Critical / emancipatory research	6	3	42	21	-	35	41	11.8
Discourse analysis	157	44	65	5	20	25	22	15.3
Ecological Psychology	0	0	-	-	-	-	-	-
Educational ethnography	0	0	-	-	-	-	-	-
Education connoisseurship	0	0	-	-	-	-	-	-
Ethnographic contents analysis	2	2	52	22	-	37	37	21.2

Ethnography of communication	1	1	34	34	-	34	34	-
Ethno-methodology	7	2	55	11	-	31	31	27.6
Ethnoscience	0	0	-	-	-	-	-	-
Event structure	0	0	-	-	-	-	-	-
Grounded theory	429	174	87	4	25	32	30	16.6
Holistic ethnography	1	0	-	-	-	-	-	-
Hermeneutics	19	9	42	7	-	24	26	10.2
Heuristic research	0	0	-	-	-	-	-	-
Life history	61	35	62	1	21	23	20	16.1
Naturalistic enquiry	2	1	26	26	-	26	26	-
Phenomenology	57	25	89	7	20	25	20	19.9
Qualitative evaluation	7	1	42	42	-	42	42	-
Reflective phenomenology	0	0	-	-	-	-	-	-
Structural ethnography	0	0	-	-	-	-	-	-
Symbolic interactionism	22	12	4	87	-	33	28	26.5
Transcendental realism	0	0	-	-	-	-	-	-
TOTAL	2533	560	95	1	30	31	28	18.7

Table 1: Descriptive statistics for each methodological group [30]

Table 1 shows that the overall range of the numbers of participants used was from 95 (using a case study approach) to 1 (also using a case study and a life history approach). Of the 560 studies analysed the median and mean were 28, and 31 respectively, suggesting perhaps that the measures of central dispersion were generally consistent. However, the distribution is bi-modal (20 and 30) and the standard deviation was 18.7, which suggests that distribution of studies was somewhat positively skewed and the range of studies was comparatively widely dispersed from the mean. Below this Figure 1 provides an illustration of the

distribution of the sample, i.e. how many studies used 1 participant in their study, how many used 2, how many used 3 etc.

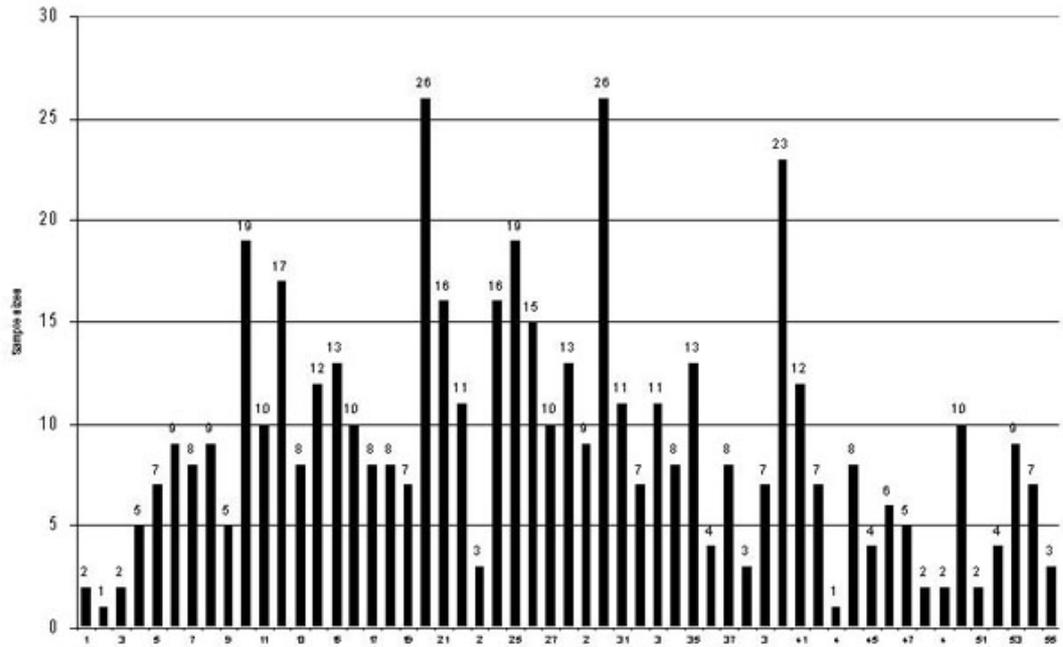


Figure 1: Number of studies by each individual sample size [31]

Figure 1 shows a bi-modal distribution with a skewness¹³ of 0.936 and a Kurtosis¹⁴ of 0.705, potentially suggesting a positively skewed distribution. The most important result from the chart however is the distribution of the sample. [32]

Figure 1 shows the prevalence of the studies that included 10, 20, 30 and 40 participants as their sample size. These were the four highest sample sizes, and provided 17% of the total number of studies in this analysis. This pattern continues with the prevalence of studies using 50 and 60 as their sample size comparative to the numbers around them (i.e. 50 is the most prevalent sample size with samples using any number from 50-59, and the same for 60). In total, the sample sizes ending in a zero account (i.e. studies with 10, 20, 30 participants etc.) for 114 of the studies in the sample. These nine sample sizes accounted for 20% of the total number of studies used in this analysis. [33]

13 Skewness is a measure of distribution of any variable. Statistics of any distribution can be calculated which indicates how randomly distributed the variable is. The closer to zero the final skewness statistic is the more randomly distributed it is felt to be. However the further away from zero (either positively or negatively) the more likely there is to be some underlying uniformity to the distribution.

14 Kurtosis is a measure of the level of the peak in any variable, i.e. how much there are extreme deviations. Statistics of any distribution can be calculated which indicates how "peaked" distributed the variable is. The higher the kurtosis statistic is the more likely there will be infrequent, and extreme deviations in the data.

A test for the randomness of fluctuations¹⁵ indicated that there was very strong evidence against the randomness of fluctuations: test statistic 5.17; $p=0.00025$. The pattern of non-fluctuation is more clearly illustrated below in Figure 2.

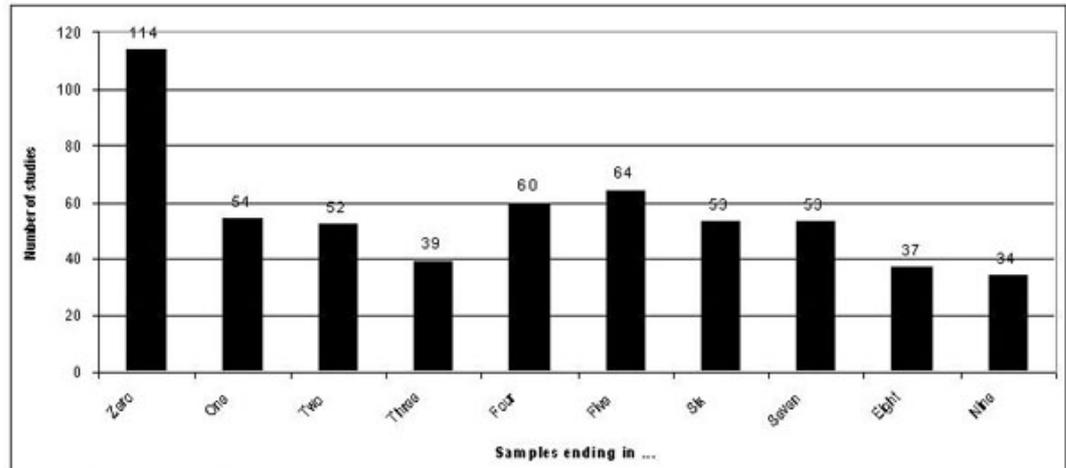


Figure 2: Number of studies with a sample ending in each integer [34]

A Chi-squared "goodness-of-fit" test¹⁶ was then used to test the null hypothesis that samples used in qualitative studies are equally likely to end on any integer. Results from the test indicate that Chi-square = 108.475; $p=0.000$, as a result the null hypothesis is rejected¹⁷. [35]

Table 1 also shows the descriptive results of the analysis of the 26 approaches identified by TESCH, in an attempt to discover whether the methodological approach affects the number of interviews undertaken. The analysis returned an uneven distribution of approaches among the studies used in the sample. Of the 26 approaches identified by TESCH, seven did not return any studies that fitted the search criteria, and a further one did not return any studies into the sample once the inclusion criteria were applied. As a result, detailed statistical analysis was not possible. [36]

However, it was clear that there were approaches that utilised interviews in their method more frequently than others did. Of the 26 qualitative approaches, nine returned more than 10 studies (eight after the inclusion criteria were applied). The most popular approaches used in PhD studies for this analysis were: case study,

15 The test for the Randomness of Fluctuations examines how randomly a set of data is organised. It can indicate whether there is a normal level of randomness which might be expected under normal conditions or whether there is some underlying pattern at work.

16 The Goodness-of-fit of a statistical test which aims to show how well a set of data that has been gathered "fits" with what might be expected under normal circumstances.

17 Before any statistical analysis is carried out on any variable or variables, the tester develops a hypothesis to test against—i.e. something they expect to happen as a result of an intervention or observation. For each hypothesis there is a null hypothesis which assumes that any kind of difference or significance seen in a set of data is due to purely to chance. It is a statistical convention to report any results in their relationship to the null hypothesis, e.g. if the null hypothesis is accepted there is an understanding that there no significant outcome as a result of the test. However if the null hypothesis is rejected it is understood that the change is thought to be as a result of intervention.

grounded theory methodology, content analysis, discourse analysis, action research, life history, phenomenology, symbolic interactionism, and hermeneutics. [37]

The approach utilising interviews most frequently were case study projects (1,401). However, only 13% of these fitted the inclusion criteria. This is followed by grounded theory studies (429). A greater proportion of these fitted the inclusion criteria (41%). Case study and grounded theory designs accounted for nearly two thirds (63%) of the entire sample. [38]

Qualitative evaluation had the highest mean number of participants (42); followed by ethnographic contents analysis (37), critical/emancipatory research (35), ethnography of communication (34). However, these means are achieved from comparatively few studies. The more studies returned into the sample for this analysis, the lower the mean tended to become. [39]

Perhaps more worthy of note is the fact that of the major approaches (i.e. those that returned the largest numbers of studies into the sample), case study approaches had the highest mean number of participants in their studies (36), while action research and life history approaches each showed mean numbers of participants of 23 in their studies. [40]

Finally, the data were compared to the guidelines given by various authors for achieving saturation in qualitative interviews (see page 3). The number of studies used in this analysis is shown below as a proportion of the whole for that approach:

- Sixty per cent of the ethnographic studies found fell within the range of 30-50 suggested by MORSE (1994) and BERNARD (2000). No ethnoscience studies were found that fitted the inclusion criteria.
- Just under half (49%) of the studies in this analysis fell within CRESWELL's (1998) suggested range of 20-30 for grounded theory studies: while just over a third (37%) fell within the range of 30-50 suggested by MORSE.
- All of the phenomenological studies identified had at least six participants, as suggested by MORSE: while just over two thirds identified (68%) fell within CRESWELL's suggested range of five to 25.
- Eighty per cent of the total proportion of qualitative studies met BERTAUX's (1981) guideline: while just under half (45%) met CHARMAZ's (2006) guidelines for qualitative samples, with up to 25 participants being "adequate" (p.114). A third of the studies (33% or 186) used sample sizes of 20 or under (GREEN & THOROGOOD, 2009 [2004]). Finally, 85% met RITCHIE et al.'s (2003) assertion that qualitative samples "often lie under 50" (p.84). [41]

4. Discussion

A wide range of sample sizes was observed in the PhD studies used for this analysis. The smallest sample used was a single participant used in a life history study, which might be expected due to the in-depth, detailed nature of the approach, while the largest sample used was 95 which was a study utilising a case study approach. The median, and mean were 28 and 31 respectively, which suggests a generally clustered distribution. However, the standard deviation (at 18.7) is comparatively high and the distribution is bi-modal and positively skewed. [42]

The most common sample sizes were 20 and 30 (followed by 40, 10 and 25). The significantly high proportion of studies utilising multiples of ten as their sample is the most important finding from this analysis. There is no logical (or theory driven) reason why samples ending in any one integer would be any more prevalent than any other in qualitative PhD studies using interviews. If saturation is the guiding principle of qualitative studies it is likely to be achieved at any point, and is certainly no more likely to be achieved with a sample ending in a zero, as any other number. However, the analysis carried out here suggests that this is the case. [43]

Of the samples achieved in this study there does not seem to be any real pattern as to how far PhD researchers are adhering to the guidelines for saturation, established by previous researchers. A large proportion of the samples (80%) adhered to BERTAUX's guidelines of 15 being the smallest number of participants for a qualitative study irrespective of the methodology. At the lower end of the spectrum, a higher proportion of researchers seem more ready to adhere to RITCHIE et al.'s guidelines that samples should "lie under 50". However, there were a proportion of studies that used more than 50 as their sample—these larger qualitative studies are perhaps the hardest to explain. [44]

While none of the guidelines presented here are intended to be faultless reference tools for selecting qualitative samples sizes, all authors agree that saturation is achieved at a comparatively low level (e.g. GUEST et al., 2006; GRIFFIN & HAUSER, 1993; and ROMNEY et al., 1986), and generally don't need to be greater than 60 participants (CHARMAZ, 2006; MORSE, 1994; CRESWELL, 1998). [45]

Without more detail of the studies it is not possible to conclude whether these larger samples were truly inappropriate. GREEN and THOROGOOD (2009 [2004]) give an example of a study where they explored how bilingual children work as interpreters for their parents. They constructed a sample of 60 participants: "but within that were various sub-samples, such as 30 young women, 15 Vietnamese speakers- 40 young people born outside the UK, and 20 people who were the only people to speak their 'mother tongue' in their school class" (p.120). [46]

MOHRMAN, TENKASI and MOHRMAN (2003) also used a comparatively large sample size for a qualitative study. Their study was longitudinal and utilised over

350 participants in eight different organisations. The study required MOHRMAN et al. to assess differences between multiple groups so each unit of analysis required its own sub-set (more in the nature of a quantitative quota sample). There is no way of knowing that the samples analysed in this study were similarly arranged. [47]

LEECH (2005) suggests that it is a mistake to presume that all qualitative research must inevitably use small samples. She feels that this ignores what she calls a growing body of research studies that utilise text-mining¹⁸ (e.g. POWIS & CAIRNS, 2003; DEL-RIO, KOSTOFF, GARCIA, RAMIREZ & HUMENIK, 2002; and LIDDY, 2000) as their method. Text-mining was not identified by TESCH (1990) as a separate methodological approach and as a result was not used in this analysis. Further analysis might examine samples from these studies in more detail. This highlights a potential weakness of this study—the interpretation of methodological approach. While it is believed that PhD researchers own descriptions of their work are likely to be accurate, it may place studies into certain categories when they might be better suited to others. [48]

Further research might also seek to quantify the other issues that affect sample size and undertake regression analysis to see what percentage of variance in the sample size can be explained by these factors. This would require a larger sample than that achieved in this paper as the unit of analysis would be the methodological approach or the existence of supplementary methods for example. Finally, this paper has sought to examine the use of personal interviewing in PhD studies for the reasons already given. Further research could feasibly examine whether these patterns exist in published research. [49]

Ultimately, qualitative samples are drawn to reflect the purpose and aims of the study. A study schedule is then designed, the study is carried out and analysed by researchers with varying levels of skill and experience. The skill of the interviewer clearly has an effect on the quality of data collected (MORSE, 2008) and this will have a subsequent effect in achieving saturation (GUEST et al., 2006)—the sample size becomes irrelevant as the quality of data is the measurement of its value. This is as a result of an interaction between the interviewer and the participant. There could be an argument, for example, which suggests that ten interviews, conducted by an experienced interviewer will elicit richer data than 50 interviews by an inexperienced or novice interviewer. Any of these factors along the qualitative journey can affect how and when saturation is reached and when researchers feel they have enough data. [50]

However, while it is clear that these issues can affect saturation, they should not dictate it. Results from this analysis suggest that researchers are not working with saturation in mind, but instead a quota that will allow them to call their research

18 Text mining is the process of collecting data from text. It involves clustering data in blocks and then searching for issues such as relevance, novelty and interest. By assessing blocks of text and categorising it, patterns can be identified. For more information see the UK National Centre for Text Mining, <http://www.nactem.ac.uk/> [Accessed: May 24, 2010].

"finished". Further research to shed light on this might explore whether the number of methods used might affect saturation. [51]

This is connected to issues raised by RITCHIE et al. (2003) and GUEST et al. (2006), who suggest that a number of factors, irrespective of sample size affect when a study is saturated. MORSE (2000) also supports this by saying that "the number of participants required in a study is one area in which it is clear that too many factors are involved and conditions of each study vary too greatly to produce tight recommendation's" (p.5). [52]

However, to have too detailed a discussion of the flexible nature of saturation (and whether the sample sizes in this analysis are appropriate or not) is to lose sight of the most important findings of this study. The over-representation of certain sample sizes in qualitative PhD studies suggests a potential deficiency in the teaching and supervision of qualitative methods. [53]

As external criteria often impinge on sample sizes (such as ethics committee requirements to state numbers at the outset, where guestimates are likely to be round numbers), it is obviously easier for a student to defend the approach they cited, rather than stop data collection apparently early, when compared to their proposal. [54]

On closer examination however, perhaps this should not be surprising. When the guidelines for saturation by various researchers are examined the integers zero and five are equally prevalent, even in those presented in more detail by GREEN and THOROGOOD. Nearly all of the examples of sample guidelines presented here by previous researcher are in multiples of five. This is all the more curious when empirical examples presenting guidelines for saturation (e.g. ROMNEY et al., GRIFFIN & HAUSER; and GUEST et al.) shy away from such simplistic estimates. [55]

5. Conclusion

So to conclude then, there is a range of potential conclusions that might be drawn as a result of this analysis:

- On the one hand, PhD researchers (and/or their supervisors) don't really understand the concept of saturation and are doing a comparatively large number of interviews. This ensures that their sample sizes, and therefore their data, are defensible.
- Alternatively PhD researchers do understand the concept of saturation but they find it easier to submit theses based on larger samples than are needed "just to be on the safe side" (and therefore feel more confident when it comes to their examination).
- Irrespective of their understanding of saturation, PhD researchers are using samples in line with their proposal to suit an independent quality assurance process (i.e. doing what they said they were going to do). [56]

Whatever the appropriate reason there are clear implications for research students using qualitative interviews. SIBLEY (2003) suggests that in her experience the teaching of qualitative research is often not as rigorous as other forms of social research, and many assumptions about its teaching are taken for granted. This is more important now than ever before, as some authors seem to have identified an increased interest in qualitative research (FLICK, 2002; CROUCH & MCKENZIE, 2006) and an increase in the number of qualitative studies printed in medical journals (BORREANI, MICCINESI, BRUNELLI & LINA 2004). It is therefore more important than ever to make qualitative methods as robust and defensible as possible. [57]

So what does this all mean? The common sample sizes and the preference for a certain "type" of approach suggest something preconceived about the nature of the PhD studies analysed here. Are students completing their samples based on what they feel they can defend, and what their supervisors and institutions require, rather than when they feel their work is actually complete? [58]

Interviewing 38 or 57 people potentially risks triggering awkward questions from readers less familiar with the concept of saturation. However, we might expect PhD students to be predominantly orientated towards their discipline, and less concerned by challenge. Sample sizes of round numbers suggest, perhaps, an insufficient grounding in the concept of saturation. [59]

The point of saturation is, as noted here, a rather difficult point to identify and of course a rather elastic notion. New data (especially if theoretically sampled) will always add something new, but there are diminishing returns, and the cut off between adding to emerging findings and not adding, might be considered inevitably arbitrary. [60]

Whether this reflects an arbitrary approach to qualitative research by students, or an overly prescriptive approach by institutions, supervisors, or a combination of both, is beyond the scope of this study. What is also beyond the scope of this study is whether all qualitative researchers see saturation as the aim of their study—as noted in the introduction saturation might not always be the only way to design/do qualitative studies. However this author firmly believes that it should be the guiding principle for qualitative data collection, at the very least. [61]

However, what is apparent is that there appears to be something pre-meditated about the samples in these studies. This perhaps further supports the assertion by BORREANI et al. (2004) that while the interest and prevalence of qualitative research is growing, the approaches used sometimes show a limited understanding of qualitative research principles (e.g. see BOWEN, 2008, who presents a full and clear explication of how his study reached saturation). With the development of sophisticated qualitative software packages, it is now much easier to present screenshots of coding frameworks and analytical models, which can help to illustrate this more fully. [62]

What is clear from this analysis is that there are issues prevalent in the studies that are not wholly congruent with the principles of qualitative research. This has clear implications for students who should ensure that they, at the very least, understand the concept of saturation, and the issues that affect it, in relation to their study (even if it is not the aim of the study). Once they are fully aware of this, they, and their supervisors, can make properly informed decisions about guiding their fieldwork and eventually closing their analysis. Alternatively, if this has to be done before saturation is achieved, they are better able to understand the limitations and scope of their work. Either way, this will contribute to a fuller and more rigorous defence of the appropriateness in their sample, during the examination process. [63]

References

- Atran, Scott; Medin, Douglas L. & Ross, Norbert O. (2005). The cultural mind: Environmental decision making and cultural modeling within and across populations. *Psychological Review*, 112(4), 744-776.
- Bernard, Harvey R. (2000). *Social research methods*. Thousand Oaks, CA: Sage.
- Bertaux, Daniel (1981). From the life-history approach to the transformation of sociological practice. In Daniel Bertaux (Ed.), *Biography and society: The life history approach in the social sciences* (pp.29-45). London: Sage.
- Borreani, Claudia; Miccinesi, Guido; Brunelli, Cinzia & Lina, Micaela (2004). An increasing number of qualitative research papers in oncology and palliative care: Does it mean a thorough development of the methodology of research? *Health and Quality of Life Outcomes*, 2(1), 1-7.
- Bowen, Glenn A. (2008). Naturalistic inquiry and the saturation concept: A research note. *Qualitative Research*, 8(1), 137-152.
- Charmaz, Kathy (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks, CA: Sage.
- Creswell, John (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Crouch, Mira & McKenzie, Heather (2006). The logic of small samples in interview based qualitative research. *Social Science Information*, 45(4), 483-499.
- Del-Rio, Jesus A.; Kostoff, Ronald N.; Garcia, Esther O.; Ramirez, Ana M. & Humenik, James A. (2002). Phenomenological approach to profile impact of scientific research: Citation mining. *Advances in Complex Systems*, 5(1), 19-42.
- Dey, Ian (1999). *Grounding grounded theory*. San Diego, CA.: Academic Press.
- [Flick, Uwe](#) (2002). *An introduction to qualitative research*. London: Sage.
- [Glaser, Barney](#) & Strauss, Anselm (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine Publishing Company.
- Green, Judith & Thorogood, Nicki (2009 [2004]). *Qualitative methods for health research* (2nd ed.). Thousand Oaks, CA: Sage.
- Griffin, Abbie & Hauser, John R. (1993). The voice of the customer. *Marketing Science*, 12(1), 1-27.
- Guest, Greg; Bunce, Arwen & Johnson, Laura (2006). "How many interviews are enough? An experiment with data saturation and variability". *Field Methods*, 18(1), 59-82.
- Jette, Dianne J.; Grover, Lisa & Keck, Carol P. (2003). A qualitative study of clinical decision making in recommending discharge placement from the acute care setting. *Physical Therapy*, 83(3), 224-236.
- Lee, Diane T.F.; Woo, Jean & Mackenzie, Ann E. (2002). The cultural context of adjusting to nursing home life: Chinese elders' perspectives. *The Gerontologist*, 42(5), 667-675.

- Leech, Nancy L. (2005). The role of sampling in qualitative research. *Academic Exchange Quarterly*, [http://www.thefreelibrary.com/The role of sampling in qualitative research-a0138703704](http://www.thefreelibrary.com/The+role+of+sampling+in+qualitative+research-a0138703704). [Accessed: September 08, 2009].
- Liddy, Elizabeth D. (2000). Text mining. *Bulletin of the American Society for Information Science & Technology*, 27(1), 14-16.
- Mohrman, Susan A.; Tenkasi, Ramkrishnan V. & Mohrman, Allan M. Jr. (2003). The role of networks in fundamental organizational change: A grounded analysis. *The Journal of Applied Behavioural Science*, 39(3), 301-323.
- Morse, Janice M. (1994). Designing funded qualitative research. In Norman K. Denzin & Yvonna S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp.220-35). Thousand Oaks, CA: Sage.
- Morse, Janice M. (1995). The significance of saturation. *Qualitative Health Research*, 5(3), 147-149.
- Morse, Janice, M. (2000). Determining sample size. *Qualitative Health Research*, 10(1), 3-5.
- Morse, Janice, M. (2008). Styles of collaboration in qualitative inquiry. *Qualitative Health Research*, 18(1), 3-4.
- Powis, Tracey & Cairns, David R. (2003). Mining for meaning: Text mining the relationship between of reconciliation and beliefs about Aborigines. *Australian Journal of Psychology*, 55, 59-62.
- Ritchie, Jane; Lewis, Jane & Elam, Gillian (2003). Designing and selecting samples. In Jane Ritchie & Jane Lewis (Eds.), *Qualitative research practice. A guide for social science students and researchers* (pp.77-108) Thousand Oaks, CA: Sage.
- Romney, A. Kimball; Batchelder, William & Weller, Susan C. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, 88(3),13-38.
- Sibley, Susan (2003). Designing qualitative research projects. National Science Foundation Workshop on Qualitative Methods in Sociology, July 2003, http://web.mit.edu/anthropology/faculty_staff/silbey/pdf/49DesigningQuaRes.doc [Accessed: September 08, 2009].
- Strauss, Anselm & Corbin, Juliet (1998 [1990]). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.
- Tesch, Renata (1990). *Qualitative research: Analysis types and software tools*. New York, NY: Falmer.
- Thomson, Bruce S. (2004). *Qualitative research: Grounded theory—Sample size and validity*, <http://www.buseco.monash.edu.au/research/studentdocs/mgt.pdf> [Accessed: September 08, 2009].
- Vorberg, Dirk & Ulrich, Rolf (1987) Random search with unequal search rates: Serial and parallel generalisations of McGill's model. *Journal of Mathematical Psychology*, 31(1), 1-23.

Author

Mark MASON has spent more than fifteen years in public sector environments carrying out, commissioning, managing and disseminating research. He began his career working in addictions in psychiatric hospitals. After a successful career in local authorities Mark moved into a policy research analyst post within central government, working initially for the Drugs Prevention Advisory Service in the Home Office. Since then he has worked in a number of social research positions nationally for the last ten years. Over the course of his career Mark has presented and published work on a range of subjects including substance use and community safety. He is currently studying for a PhD at Oxford Brookes University.

Contact:

Mark Mason

Oxford Brookes University
School of Health and Social Care
Marston Road Campus
Jack Straw's Lane
Oxford
OX3 0FL
UK

E-mail: Mark.mason@brookes.ac.uk

Citation

Mason, Mark (2010). Sample Size and Saturation in PhD Studies Using Qualitative Interviews [63 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 11(3), Art. 8, <http://nbn-resolving.de/urn:nbn:de:0114-fqs100387>.

Revised: 9/2010