

## **‘Unsatisfactory Saturation’: A critical exploration of the notion of saturated sample sizes in qualitative research**

### **Abstract**

Measuring quality in qualitative research is a contentious issue with diverse opinions and various frameworks available within the evidence base. One important and somewhat neglected argument within this field relates to the increasingly ubiquitous discourse of data saturation. While originally developed within grounded theory, theoretical saturation, and later termed data/thematic saturation for other qualitative methods, the meaning has evolved and become transformed. Problematically this temporal drift has been treated as unproblematic and saturation as a marker for sampling adequacy has become increasingly accepted and expected. In this paper we challenge the unquestioned acceptance of the concept of saturation and consider its plausibility and transferability across all qualitative approaches. By considering issues of transparency and epistemology we argue that adopting saturation as a generic quality marker is inappropriate. The aim of this paper is to highlight the pertinent issues and encourage the research community to engage with and contribute to this important area.

**Key words:** data saturation, thematic saturation, quality, qualitative research, transparency, epistemology

## Introduction

A paper of ours was recently peer reviewed for publication and while much of the feedback was constructive and fair, one of the criticisms of it was that we had failed to mention saturation of the sample. On closer inspection the request for such detail was embedded in the journal's own quality criteria. The paper under review had utilised the conversation analytic method and we were thus perturbed by this generic requirement for all qualitative inquiry. Our reasoning for this is that different qualitative perspectives have different indices for quality assurance and in this respect some qualitative approaches do not rely on saturation as a marker for sample size adequacy. Furthermore the notion of saturation was originally tied to grounded theory with a specific and theory driven meaning, and while this has been helpfully translated for other qualitative approaches it is not appropriate to impose it in all instances. Because of the increasing value being placed on theoretical/thematic/data saturation we believe this issue requires closer examination. In this paper, therefore, we aim to disentangle the debates surrounding this notion and provide a critical evaluation which will hopefully lead to opening up the debate further and to help demystify some of the complexities.

### *Quality criteria in qualitative research*

*“If we succumb to the lure of “one size fits all” solutions we risk being in a situation where the tail (the checklist) is wagging the dog (the qualitative research)”*

(Barbour, 2001:1115).

Within qualitative methodological discussions, the literature is littered with debates about whether there should be generic quality criteria for all qualitative research (Caelli et al, 2003; Mays and Pope, 2000; Tracy, 2010). While it is accepted that quality criteria checklists have contributed to increased confidence in the validity of qualitative inquiry and the wider acceptance of qualitative methods generally, these can be counterproductive if followed prescriptively (Barbour 2001). Barbour argues that an uncritical adoption of a range of criteria does not in itself equate with rigour. Furthermore, there is no singular way to measure the quality of qualitative research because it is so diverse (Guba and Lincoln, 2005: Mays and Pope, 2000). It is argued, therefore, that it is problematic to attempt to develop quality criteria applicable to all qualitative approaches as this would not respectfully value the wide range of methodologies which fall under the rubric of qualitative work and because of this each approach should be idiosyncratically evaluated against quality markers that are congruent with their epistemological origins (Caelli et al, 2003).

While there are areas of quality which potentially relate to all qualitative approaches and Tracy (2010) argues for eight such universal quality markers, other markers are less suitable for a blanket approach. One such issue that seems to have become confused in the literature relates to sampling, particularly the notion of saturation. Saturation seems to have become the gold standard against which the diversity of samples is determined (Guest et al, 2006) and yet saturation has multiple meanings and limited transparency. Defensibility of the quality of qualitative research, to a considerable extent, relates to sampling adequacy which should provide depth and maximum opportunity for transferability of findings (Spencer et al, 2003).

*Sampling and the concept of saturation*

Sampling is a core concern for researchers to determine the success of a project and continual examination is required (Tucket 2004). In qualitative research the selection of respondents cannot follow the procedures of quantitative sampling because the purpose is not to count opinions or people but explore the range of opinions and different representations of an issue (Gaskell, 2000). Thus, sampling in qualitative research is concerned with the richness of information (Kuzel, 1992) and the number of participants required, therefore, depends on the nature of the topic and the resources available (Gaskell, 2000). There are two key considerations that guide the sampling methods in qualitative research, appropriateness and adequacy (Morse and Field, 1995). It is argued, therefore, that the researcher should be pragmatic and flexible in their approach to sampling and that an adequate sample size is one that sufficiently answers the research question (Marshall, 1996).

In this sense then generalisability is not sought by the researcher and the focus is less on sample size and more on sample adequacy (Bowen, 2008). Bowen argues that adequacy of sampling relates to the demonstration that saturation has been reached which means that depth as well as breadth of information is achieved. Qualitative researchers often make decisions related to the adequacy of their sample based on the notion of saturation. There has been, however, a development of the ways in which saturation is understood and utilised by researchers. The consequence of this has been that there is now some confusion in terms of what saturation means, how it should be used and when it is applicable.

### *History of saturation*

There are various forms of saturation, with the original being theoretical saturation developed in the approach of grounded theory (Guest et al, 2006). Other variations of the concept for

other qualitative methods include, data saturation (Francis et al, 2010; Guest et al 2006), thematic saturation (Guest et al, 2006) and in some cases simply saturation (Starks and Trinidad, 2007). While there is some diffusion and vagueness surrounding these terms (Guest et al, 2006) they do have distinct meanings and are typically applied to all qualitative methods. Generally, however, thematic/data saturation are normatively taken to mean that data should continue to be collected until nothing new is generated (Green and Thorogood 2004); the point at which there are fewer surprises and there are no more emergent patterns in the data (Gaskell, 2000). This is quite different to theoretical saturation, the form of saturation used by grounded theorists.

The original meaning of saturation pioneered within grounded theory, of theoretical saturation, is still used within this approach in current work and has retained its central importance. In grounded theory the notion of saturation does not refer to the point at which no new ideas emerge, but rather mean that categories are fully accounted for, the variability between them are explained and the relationships between them are tested and validated and thus a theory can emerge (Green and Thorogood, 2004). This is congruent with the underpinning epistemological position and the goals of grounded theory, which are to develop an explanatory theory of the social processes that are studied in the environments in which they have taken place (Glaser and Strauss, 1967; Starks and Trinidad, 2007).

Despite the different meanings that have been applied to saturation, the changes of those meanings over time and the general acceptance of the new meaning, it has significant influence and has attracted some debate in terms of its practical application and transparency in dissemination. While we recognise that uniform acceptance of using one form of saturation as a quality marker has considerable drawbacks, we note that for those approaches that

warrant its use, there is an onus on transparent practice and thus prior to discussions regarding problematic expectations of saturation we explore these important debates first.

### *Quality and transparency*

In qualitative research transparency is a recognised marker of quality (Spencer et al, 2003) which means that sufficient detail should be included about how the data were collected (Meyrick, 2006). Within qualitative research sufficiency of sample size is measured by depth of data rather than frequencies and therefore samples should consist of participants who best represent the research topic (Morse et al, 2002). The aim of some qualitative work is to have generalisability or transferability and thus sample size is important (Onwuegbuzie, 2003). The corpus needs to be large enough to capture a range of experiences but not so large as to be repetitious, and the common guiding principle is saturation (Mason, 2010). Notably, within the literature the notion of saturated samples tends to be used as an indication of quality (Guest et al, 2006), and yet it has been argued that transparency, with regards to adequacy of sample sizes, is generally inadequate in dissemination (Bowen, 2008). Bowen argues that researchers have tended to gloss over the details about how saturation was determined.

Achieving transparency is, however, a complex requirement for researchers given that there are limited guidelines for the research community to utilise (Francis et al, 2010; Ziebland and McPherson, 2006) and differing meanings of the term. There are two areas of importance in relation to the need for guidance about saturation. Firstly, it is becoming an increasingly common requirement at the design stage, for planning, funding and ethical review, to state in advance the proposed size of the sample. This is potentially problematic given that researchers shy away from making suggestions about sample size sufficiency (Mason, 2010)

and in some approaches a priori estimations are inappropriate (Morse, 1995). Secondly, there is limited practical guidance or help to show researchers when saturation has been reached (Bowen, 2008; Guest et al, 2006). This prompts questions about how the research community might agree on principles that researchers and reviewers can use to determine when saturation has been reached and how to best defend judgements in a way that is transparent to readers (Francis et al, 2010). This is particularly important given that in reality, researchers often stop recruitment when resources become limited and are driven by time and money, rather than sample adequacy (Green and Thorogood, 2004).

Interestingly in qualitative work it is often left to the reader to unpack the data collection and analysis to generate clues as to the methodology with little transparency or evidence to how or why saturation was achieved (Caelli et al, 2003). This is evidenced by a review of a leading journal which found that during a sixteen month period, eighteen articles mentioned saturation and yet none of them were transparent about how it was achieved (Francis et al, 2010). This is problematic given that it is an expectation that researchers make explicit the process of saturation during dissemination (Bowen, 2008). In reality there are practical constraints on the researcher in terms of unforeseen participant attrition (Tuckett, 2004) and in terms of time and resources (Green and Thorogood, 2004). This is particularly important as there are arguments relating to saturation and quality within each interview and therefore, researchers need to pay attention to both the length of interviews as well as the number of interviews (Onwuegbuzie and Leech 2005). Transparency about these limitations on reaching saturation does not necessarily invalidate the findings. If saturation is not reached this simply means that the phenomenon has not yet been fully explored rather than that the findings are invalid (Morse, 1995). It is acceptable, therefore, that any limitations of sampling adequacy

are transparently reported. Researchers thus need to be clear in dissemination if they reached saturation, how they reached it and what issues they faced during recruitment

Saturation is a convincing concept but has a number of practical weaknesses especially as in some cases the number of emergent themes are potentially limitless (Green and Thorogood, 2004). This is because each life is unique and in this sense data are never truly saturated as there will always be new things to discover (Wray et al, 2007). In qualitative inquiry researchers can take an inductive approach or deductive approach. Those who use inductive reasoning use the data to generate ideas and those who use deductive reasoning begins with an idea and then use the data to confirm it (Thorne, 2000). In theory therefore, those who have a focussed idea have a related research agenda, and thus this guides the direction of the data collection. In this way parameters are created and particular areas of interest pursued within which saturation can be achieved. On the other hand a deductive approach is much broader and the researcher is unaware of the types of categories that may emerge from data collection. In this sense, the potential for achieving saturation becomes an unrealistic target.

What this section has highlighted is that transparency of process is essential during dissemination to ensure quality in qualitative research. Achieving quality, however, transcends transparency of saturation as authors have a responsibility for evidencing the merits of the overall approach (Caelli et al, 2003). Caelli et al further argue that researchers must address their theoretical position, evidence the congruence between methodology and methods and highlight the strategies they used to establish rigour. What this means is that the concept of saturation is not always an appropriate criterion for establishing quality across all qualitative approaches. Theoretical position and establishing congruence between methods means that differing data collection methods are favoured and different quality markers are

utilised. For example although interviews are a common and popular method of data collection and deemed suitable for some approaches (Gaskell, 2000), they are not universally favoured by all. This is particularly noteworthy given that the majority of debate and discussion around saturation almost exclusively focuses on interview and focus group studies. Thus there seems to be an omission of critical thinking about its versatility when applied to collection methods such as naturally occurring data, diary entries or observations.

*The problem of inappropriate expectations about saturation*

It is clear from the argument thus far that considering saturation is more complex than the literature has suggested. Differing data collection methods frame the sufficiency of data quantity in different ways and because of this it is questionable whether saturation can be applied in all cases. The legacy of quantitative science appears to have left a cultural residue of larger numbers having greater impact. This is not applicable to qualitative work as more data does not necessarily lead to more information (Mason, 2010). This is a particularly important issue as it is not just unnecessary but also potentially unethical to recruit further participants to a study and not make full use of the data they provide (Francis et al, 2010). In qualitative inquiry, the aim is not to acquire a fixed number of participants rather it aims to gather sufficient depth of information as a way of fully describing the phenomenon being studied (Fossey et al, 2002). As such there are differences in how various approaches frame research questions, sample participants and collect data (Starks and Trinidad, 2007) In order to achieve richness and depth of analysis.

The central aim of research is to extend and advance knowledge (Caelli et al, 2003) and yet fundamental arguments about the nature of and mechanisms for its acquisition are diverse.

Different assumptions about knowledge thus inform the epistemological starting point of research and, therefore, also determine the aims and objectives of any given project. It is these aims and objectives which will guide the trajectory of the whole research process. We argue that this is particularly important when applying the notion of saturation to sampling adequacy. The adequacy of the sample is, therefore, not determined solely on the basis of the number of participants but the appropriateness of the data. For example conversation analysts have a preference for small data sets of naturally occurring data as more appropriate for their unique mode of inquiry (Potter, 2002). This is one approach where saturation as a marker for quality becomes redundant. There is a sophisticated literature about how to manage decisions in the research process using conversation analysis and there are robust mechanisms for ensuring quality within it (Hutchby and Wooffitt, 2008; Ten Have, 2007).

The irrelevance of saturation is not limited to conversation analysis, however, and the epistemological and methodological frameworks should guide researchers in their decision making and application of quality criteria. The problem arises when markers such as saturation are rigidly applied in all cases. The unquestioned acceptance of concepts like saturation consequently become part of an institutional discourse of quality which perpetuate unhelpful myths about optimum sampling adequacy and simultaneously undermine the value of research not conforming to these expectations. This is understandable as researchers are accustomed to using language and concepts that are relevant for their own research community (Caelli et al, 2003). Nonetheless, respect for other traditions is important and in academic activities such as writing, reviewing and teaching, a reflexive attitude is essential. It is possible to maintain methodological integrity within a particular tradition, while fairly assessing other qualitative methods against their own measures of quality. In terms of

saturation, therefore, theoretical congruence should be maintained so as to not dilute its usefulness. As Caelli et al (2003: 9) propose:

*“While saturation has a distinct theoretically embedded meaning in grounded theory, its ubiquitous and non-selective use risks rendering the term meaningless to the qualitative research community”*

### *Concluding remarks*

What is evident from this discussion is that the debates around the application of saturation beyond its origins in grounded theory, have received limited attention. While grounded theory has clear guidance about what constitutes theoretical saturation, how to apply it and when to use it, the new meanings in relation to other qualitative approaches are less developed. Our tentative arguments in this paper are designed to disentangle some of the key issues that have emerged since the transformation of the concept and to open up debates. As saturation becomes unquestioned and expected, it is necessary to take time to reflect on what this actually means for research practice. As journals are starting to incorporate questions about saturation in the process of review, it is imperative that questions are raised before the research community becomes complacent. We note that different academics from different traditions hold interesting and diverse opinions on this issue and we hope that this paper will stimulate debate so that academia and qualitative research can progress.

## References

- Barbour, R. (2001). Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *British Medical Journal*, 322, 1115-1117.
- Bowen, G. (2008). Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research*, 8 (1), 137-142.
- Caelli, K., Ray, L., & Mill, J. (2003). 'Clear as mud': Toward greater clarity in generic qualitative research. *International Journal of Qualitative Methods*, 2 (2), Open access.
- Fossey, E., Harvey, McDermott, F. & Davidson, L. (2002). Understanding and evaluating qualitative research. *Australian and New Zealand Journal of Psychiatry*, 36, 717-732.
- Francis, J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M., & Grimshaw, J. (2010). What is adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology and Health*, 25 (10), 1229-1245.
- Gaskell, G. (2000). Individual and group interviewing. In Bauer, M, & Gaskell, G. (Eds). *Qualitative Researching with Text, Image and Sound*. (pp: 38-56). London: SAGE
- Glaser, B. & Strauss, A. (1967). *The Discovery of Grounded Theory; Strategies for Qualitative Research*. Chicago: Aldine

Green, J. & Thorogood, N. (2004). *Qualitative Methods for Health Research*. London: SAGE Publications.

Guba, E. G. & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions and emerging influences. In N. Denzin & Y. S. Lincoln (Eds). *The Sage Handbook of Qualitative Research (Third Edition)*. (pp: 191-216). Thousand Oaks, CA: SAGE

Guest, G. Bruce, A. & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18 (1), 59-82.

Kuzel, A. (1992). Sampling in qualitative inquiry. In B. Crabtree, & W. Miller, (Eds). *Doing Qualitative Research*. (pp: 31-44). California: SAGE.

Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. *Forum: Qualitative Social Research*, 11 (3) Art.8, online.

Mays, N. & Pope, C. (2000). Quality in Qualitative Health Research. In C. Pope, & N. Mays, (Eds). *Qualitative Research in Health Care*. (pp: 89-102). London: BMJ Books.

Meyrick, J. (2006). What is good qualitative research? A first step towards a comprehensive approach to judging rigour/quality. *Journal of Health Psychology*. 11 (5): 799-808.

Morse, J. (1995). The significance of Saturation. *Qualitative Health Research*, 5 (2), 147-149.

Morse, J., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Institute for Qualitative Methodology, 1* (2), 13-22.

Morse, J. and Field, P. (1995). *Qualitative Methods for Health Professionals*. (Second Edition). California: SAGE

Onwuegbuzie, A. (2003). Effect sizes in qualitative research: a prolegomenon. *Quality & Quantity: An International Journal of Methodology, 37*, 393-409

Onwuegbuzie, A. & Leech, N. (2007). A call for qualitative power analyses. *Quality & Quantity: An International Journal of Methodology, 41*, 105-121.

Potter, J. (2002). Two kinds of natural. *Discourse Studies, 4* (4): 539-542.

Spencer, L., Ritchie, J & Lewis, J. et al (2003). *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*, Government Chief Social Researcher's Office, Prime Minister's strategy Unit, London. [www.strategy.gov.uk](http://www.strategy.gov.uk)

Starks, H. & Trinidad, S.B. (2007). Choose your method: a comparison of phenomenology, discourse analysis, and grounded theory. *Qualitative Health Research, 17* (10), 1372-1380

Ten Have, P. (2007). *Doing Conversation Analysis*. London; SAGE.

Thorne, S. (2000). Data analysis in qualitative research. *Evidence Based Nursing, 3*, 68-70.

Tracy, S. (2010). Qualitative quality: eight “Big-Tent” criteria for excellent qualitative research. *Qualitative Inquiry*, 16 (10), 837-851.

Tuckett, A. (2004). Part 1: Qualitative research sampling – the very real complexities. *Nurse Researcher*, 12 (1), 47-61.

Wray, N., Markovic, M. & Manderson, L. (2007). “Researcher saturation”: The impact of data triangulation and intensive-research practices on the researcher and qualitative research process. *Qualitative Health Research*, 17 (10), 1392-1402.

Ziebland, S. & McPherson, A (2006). Making sense of qualitative data analysis: an introduction with illustrations from DIPEX (personal experiences of health and illness). *Medical Education*, 40, 405-414.