Ten steps for specifying saturation in qualitative research

Author

Buckley, Ralf

Published

2022

Journal Title

Social Science & Medicine

Version

Accepted Manuscript (AM)

DOI

10.1016/j.socscimed.2022.115217

COMMENTARY on
Hennink & Kaiser, 2022, *Social Science & Medicine*, 114523.
doi: 10.1016/j.socscimed.2021.114523

Ten steps for specifying saturation in qualitative research.

Abstract

Saturation is critical to reliability of qualitative social sciences, but methods descriptions differ greatly in the degree of detail provided. Rigour could be improved by routinely following a ten-step methodological protocol. 1. Define the underlying disciplinary framework. 2. Specify the target class precisely. 3. Show how participants or cases were selected or excluded. 4. Describe techniques to minimise inadvertent or indirect selection bias. 5. Report homogeneity or heterogeneity of cases, relative to the focus of analysis. 6. Report processes for elicitation or extraction of information content. 7. Select code, meaning or model saturation. 8. Specify code and concept fineness or granularity. 9. Report order and randomisation of cases. 10. Define the level of precision in *post facto* tests of saturation.

Keywords: reliability; thematic analysis; coding; granularity; randomisation; saturation

Introduction

Qualitative social science research, in medicine as in many other disciplines, focusses on identification and interrelationships of repeated themes in individual perceptions, revealed via language and communications. Key components are: whose perceptions are considered; how they are elicited, expressed, and examined; and the degree to which the researchers impose their own interpretations, or enable participants to co-construct findings. The focus of most qualitative research is on conceptual thematic hierarchies, often represented as coding trees. One major methodological consideration is saturation: eliciting or extracting a sufficiently comprehensive range of lowest-tier concepts, that the hierarchy of higher-tier constructs is efficient and stable, and any proposed explanatory model is correspondingly reliable.

Empirical tests of saturation were reviewed recently by Hennink and Kaiser (2022). Their focus was on *post facto* tests, carried out retrospectively after a study is complete, to check saturation. Here I aim to add to their findings, by identifying additional methodological concerns that merit consideration earlier on, at study design and data collection stages. These concerns apply to qualitative methods that aim to generate structured hierarchies (coding trees) of qualitative themes.

The concerns are: underlying disciplinary frame or epistemology; precise definition of target class; selection or exclusion of interviewees or cases; minimisation of inadvertent or indirect selection bias; degree of homogeneity or heterogeneity between cases; methods used to elicit interview information or extract case content; choice of code, meaning, or model saturation; fineness or granularity of coding; ordering and randomisation of cases; and precision in *post facto* statistical tests of saturation. Currently, each of these may be mentioned by some authors, but ignored or declared only implicitly by others. I propose that they could be adopted as a routine ten-point protocol to improve rigour in qualitative research. That is, I propose that each should be addressed during the design and conduct of research, and then described explicitly during research reporting and publication.

Step 1: Define the Underlying Disciplinary Framework

At the broadest scale, different disciplines may have very different views on the meanings and relative importance of different concepts. Medical research, for example, includes economics and geography, behaviour and communications, as well as anatomy, psychology, physiology and pharmacology. The academic discipline used as the underlying basis for any particular qualitative study defines the framework or epistemology for that study: the concepts that it recognises and considers relevant and important. Those concepts then define the codes used in thematic analysis, and choice of codes influences progress towards saturation. Disciplinary epistemology also influences the knowledge, training, and interests of researchers, which affect the topics addressed, cases and content selected, codes and concepts extracted, and constructs and coding trees created. The same interviews, for example, yield different coding trees within frameworks of mental health or place image respectively (Cooper and Buckley, 2021). Disciplinary frameworks are therefore a fundamental factor in saturation. It would be preferable to specify them explicitly rather than implicitly.

Step 2: Specify the Target Class Precisely

Qualitative research analyses human thoughts, feelings, preferences and opinions, and these different between individuals. During analysis, individual responses are dissected and sorted into codes, which are reassembled into coding trees or hierarchies. Since saturation refers to the completeness of the code set and coding tree structure, it is influenced strongly by the target group, the class of people whose opinions are sought. Therefore, one critical aspect of saturation in any qualitative study is to define and delimit that target class prospectively at the design stage, as precisely as possible. The match can then be tested retrospectively, by analysing the characteristics of actual respondents. For example, if we ask people's opinions about cancer treatments, we are likely to receive a very different set of responses from a narrow target class, such as those who have received successful treatments for a specific cancer type, than from a broader class, such as a representative sample of an entire national population; and progress towards saturation will differ correspondingly.

Step 3: Show How Participants or Cases Were Selected or Excluded

Few studies can obtain data from every case or participant within the target class. Therefore, the mechanisms used to select cases or participants will influence the themes or concepts identified, and the speed of progress towards saturation. The case or participant selection process is therefore an important factor for readers to judge reliability of findings. In many qualitative studies, descriptions of case selection are imprecise, mentioning only convenience or snowball approaches. Such specifications are inadequate, since individuals have widely differing interests, perceptions, and opinions. Generally, to achieve saturation of concepts, we need techniques to maximise the range of views recorded from within the target class. Methods should therefore describe what steps were taken to maximize diversity within the target class, e.g. by ensuring that participants included a range of ages, genders, and ethnicities, and by seeking responses at different times of day and week, and perhaps in different places.

Step 4: Describe Measures to Minimise Inadvertent or Indirect Selection Bias

Qualitative methods may run the risk of indirect and inadvertent bias towards particular subgroups within the target class, because of the particular ways in which respondents or cases are identified and approached, and information elicited. Since different subgroups may hold different views incorporating different themes or concepts, any such bias can affect progress towards saturation. If the target class is nationwide, for example, but the study is carried out within a single hospital catchment or public-health area, then interviewees may not be socioeconomically, demographically, or psychologically representative of broader geographical regions. Similarly, if respondent panels are assembled using commercial survey organisations rather than census-representative permanent panels, then they may be biased towards individuals in poor mental health, because individuals in distressed circumstances sign up preferentially to self-selected commercial surveys (Chauvenet et al., 2020). For on-site interviews in person, bias can be introduced through non-random sources of rejection: e.g., by people in a hurry, or with children, or with a different first language, or fear of authority. Methods should therefore state explicitly, what potential sources of bias were considered, and how they were minimised.

Inadvertent bias can also be introduced by: any incentives offered, including form as well as amount; exact format and wording of privacy, anonymity, and confidentiality guarantees; and information provided to participants about the research and researchers. These are routine considerations in ethics declarations, but they are also possible sources of inadvertent bias in cases or content, and hence in progress towards saturation. If data were compiled covertly or neutrally, for example, without explicit identification of the researcher's role and identity, then readers need to know exactly how this was achieved. They may also want to know how similar the researchers were to the respondents, eg in medical history, age, gender, ethnicity, language, socioeconomic status, and personal interests, since all of these may affect exchange of information.

Step 5: Report Homogeneity or Heterogeneity of Cases, *cf.* Focus of Analysis

Progress towards saturation is strongly affected by the homogeneity or heterogeneity of cases and content. This is recognised widely in discussions of saturation (Guest et al., 2020; Hagaman and Wutich, 2017), but there are additional considerations, not commonly reported. If the focus of analysis is narrower than that of data collection, for example, so that only some cases and content are relevant, then more cases in total are needed, to achieve a given degree of saturation. Examples include: (a) life narrative approaches (Zins and Ponocny, 2022), where health is only one of many topics discussed; (b) directed content analysis

(Assarroudi et al., 2018; Ebrahimi Belil et al., 2018), where the themes for qualitative data collection and analysis are predefined, and only some of the cases and content collected are relevant; or (c) analyses that extract health components from data that were initially compiled for unrelated reasons (Cooper and Buckley, 2021).

Similarly, additional cases are needed to achieve saturation if a differential stakeholder narrative analysis is adopted (Reed et al., 2009), since this indicates consistent heterogeneity of opinion between subgroups within the overall class of participants. This can occur either at broad scale, e.g. between different interest groups; or at finer scale, e.g. between different groups of individuals receiving similar treatments (Akhoundogli and Buckley, 2021; Buckley etal., 2017). The same themes (codes) may be raised by each stakeholder group, but with different views (meanings) on each issue. To achieve overall saturation, either code or meaning, requires saturation for each subgroup separately. In practice, homogeneity or heterogeneity between cases is only discussed explicitly where differential stakeholder narratives are constructed; but it is equally relevant for all qualitative social science research.

Step 6. Report Processes for Elicitation or Extraction of Information Content

Mode and means used to extract information from text, or elicit opinions or perceptions using spoken language, are influential in determining the range of opinions expressed, and hence the accumulation of themes and concepts, and consequent progress towards saturation. If documents or social media materials are analysed, then methods should therefore describe how they were selected and screened (eg by language, topic, content or author), and what stopping rules were used to determine how many were compiled. If interviews or focus groups were used, methods sections should describe where and when were they carried out, how long they took, and whether they were conducted individually or jointly, eg as formal or informal focus groups.

They should also include: choice of languages; degree of interview structure; questions and issues addressed; whether the order of questions was fixed, varied, or randomized (Buckley, 2020); and what prompts were used, if any. If any additional elicitation measures were used, such as audio, photo or video prompts, these should also be described. It is also important to know whether analyses were performed on written text only, or on audio or video recordings where pauses and intonations are also available for interpretation. All of these factors influence the depth and richness of responses, and hence, progress to saturation.

Step 7: Select Code, Meaning or Model Saturation

In thematic analysis, and in methods that adopt thematic analysis to hypothesize functional models, saturation is a functional parameter of the analytic process (Guest et al., 2020; Hennink and Kaiser, 2022). Data are compiled as cases, such as interviews, documents, or social media posts. Each case contains content, expressed as language, text, or images, that includes one or more themes, extracted as codes. The codes signify concepts, and thematic analysis arranges these into a hierarchical coding tree, with successively higher tiers of constructs. Iterative analysis of additional cases generates a cumulative set of codes, that grows asymptotically.

Full code saturation is reached once further cases contribute no further codes. Effective code saturation is reached once further cases do not change the coding tree. Distinctions may also be drawn between saturation of code names, saturation of multiple meanings within codes, and saturation of latent theoretical models (Hennink and Kaiser, 2022). Data collection is commonly expected to continue until the number of cases exceeds that required to reach saturation, since this is a necessary condition for confidence in findings. Some studies employ explicit stopping rules to halt data collection. Others use tacit judgement, and test saturation retrospectively via *post facto* tests (Hennink and Kaiser, 2022). Others again, perhaps the majority across many disciplines, do not make any formal tests, but rely simply on collecting a large volume of data. In medicine, where research results can be highly consequential, explicit formal methodologies are especially important, and this applies for saturation as for any other aspect of method (Hennink and Kaiser, 2022).

Step 8: Specify Code & Concept Fineness or Granularity

Saturation is strongly influenced by the granularity or fineness of the concept codes extracted and presented. In practice this is represented by the number of tiers in a coding tree, with saturation for broad themes reached more rapidly than for narrower themes. There is no ultimate philosophical limit to subdivision of concepts, so granularity is determined by culture and language, research topic, and researcher interests. Similar concepts are more or less finely divided in different languages, and the granularity of case content reflects that linguistic structure. This merits particular attention for cross-cultural, multi-lingual analyses, where translation of linguistic terms may not provide one-to-one correspondence in concepts (Buckley et al., 2021). Granularity also reflects methods used to elicit information, such as the level of fine detail in interview questions and prompts. It reflects the coding process adopted by the researchers and/or co-constructing participants, who can decide whether to code concepts at finer or coarser levels, and how much to compress raw qualitative data from multiple sources (Buckley, 2018). In addition, researchers may only present the upper tiers of a coding tree, if more finely-divided lower tiers are not critical to the question investigated (Cooper and Buckley, 2021).

Step 9: Report Order & Randomisation of Cases in *Post Facto* Tests

Post facto empirical approaches to testing for saturation are widespread (Guest et al., 2020; Hagaman and Wutich, 2017; Hennink and Kaiser, 2022). Progress towards saturation, and hence the outcomes of these *post facto* tests, are affected by the order of cases in the analysis, since cases rarely provide equal detail. If we start with the most detailed cases, we will reach saturation with fewer cases than if we start with the sparsest cases. For a *post facto* test of code saturation, we can repeatedly randomize the order of cases, and re-plot cumulative number of codes against cumulative number of cases. With enough cases, these will converge to the same asymptote. If they do not, then more cases are needed to achieve saturation, and to complete reliable *post facto* tests. Some database and qualitative analysis software includes options to randomise case order. Alternatively, manual randomization is straightforward, using random number tables or generators. Methods descriptions should specify whether case order was randomized in testing for saturation; and if not, what factors determined the order. To date, this seems to have been reported only rarely (Hennink and Kaiser, 2022).

Step 10: Define the Level of Precision in *Post Facto* Tests of Saturation

Post facto tests of saturation also need to specify a precision or error factor: how close to the estimated asymptote do we need to get? Some empirical tests of saturation (Guest et al., 2020; Hennink and Kaiser, 2022) refer to saturation percentages; but those are themselves dependent on case ordering and coding granularity. Percentage saturation can be defined only if we can estimate the asymptote in the graph of cumulative number of codes against cumulative number of cases. This is only meaningful if we can fit a negative exponential function to that graph: #codes = A{1 – 1/exp(#cases)}, with A defining full saturation. If the curve is better represented by, eg, a logarithmic function, then there is no definable measure of full saturation, only of effective saturation in the sense of a stabilised coding tree, at a particular granularity. If we use coarse constructs, we may reach full saturation; if we use finer concepts, we may not. Whether or not that matters, depends on whether the analysis focusses on the broad structure of the case-content material, or on finer components. Rather than percentage saturation, therefore, results could preferably report the numbers of cases needed to reach saturation at each tier of the coding tree. This has been done occasionally in past published research, but not routinely.

Conclusions

I argue here that saturation is a critical criterion in assessing the reliability of findings from qualitative social science research, in medicine as well as other disciplines. I agree with the suggestion of Hennink and Kaiser (2022) that it should be considered in explicit detail by authors, editors and reviewers. Reader access to entire datasets is rare in qualitative research, because of privacy and confidentiality obligations in research ethics protocols. Readers, including medical practitioners, healthcare policymakers, and other researchers, rely on depth and detail of methods descriptions and presentation of results, to judge reliability of findings. Saturation is a key component, and steps taken to achieve and test for saturation are therefore correspondingly important. To provide sufficiently detailed information may require some expansion of word limits for qualitative articles, or separate supplementary materials. My suggestion is that authors of qualitative social science research should routinely include responses, at least briefly, to a ten-step methodological protocol (Table 1), analogous to protocols such as PRISMA for reviews and meta-analyses (Moher et al., 2015).

Table 1. Ten-Step Methodological Protocol for Specifying Saturation

1. Define the underlying disciplinary framework
2. Specify the target class precisely
3. Show how participants or cases were selected or excluded
4. Describe techniques to minimise inadvertent or indirect selection bias
5. Report homogeneity or heterogeneity of cases, *cf.* focus of analysis

6. Report processes for elicitation or extraction of information content
7. Select code, meaning or model saturation
8. Specify code and concept fineness or granularity
9. Report order and randomisation of cases in *post facto* tests of saturation
10. Define the level of precision in *post facto* tests of saturation

References

Akhoundogli, M., Buckley, R.C. 2021. Outdoor tourism to escape social surveillance: health gains but sustainability costs. J. Ecotour., doi: 10.1080/14724049.2021.1934688

Buckley, R.C. 2018. Ageing adventure athletes assess achievements and alter aspirations to maintain self-esteem. Front. Psych. 9, 225.

Buckley, R.C., 2020. Nature tourism and mental health: parks, happiness, and causation. J. Sust. Tour. 28, 1409-1424.

Buckley, R.C., Guitart, D., Shakeela, A. 2017. Contested surf tourism resources in the Maldives. Ann. Tour. Res. 64, 185-199.

Buckley, R.C., Zhong, L.S., Martin, S. 2021. Mental health key to tourism infrastructure in China's new megapark. Tour. Manage. 82, 104169.

Assarroudi, A., Heshmati Nabavi, F., Armat, M.R., Ebadi, A., Vaismoradi, M., 2018. Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. J. Res. Nurs. 23, 42-55.

Chauvenet, A.M., Buckley, R.C., Hague, L., Fleming, C., Brough, P., 2020. Panel sampling in health research. Lancet Psych. 7, 840-841.

Cooper, M.A., Buckley, R., 2021. Tourist mental health drives destination choice, marketing, and matching. J. Trav. Res. 61, 786-799.

Ebrahimi Belil, F., Alhani, F., Ebadi, A., Kazemnejad, A., 2018. Self-efficacy of people with chronic conditions: A qualitative directed content analysis. J. Clin. Med. 7, 411.

Guest, G., Namey, E., Chen, M., 2020. A simple method to assess and report thematic saturation in qualitative research. PloS One 15, e0232076.

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.

Hennink, M., Kaiser, B.N., 2022. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. Soc. Sci. Med., 114523.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P. & Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst. Rev. 4, 1-9.

Reed, M.S., Graves, A., Dandy, N., Posthumus, H., Hubacek, K., Morris, J., Prell, C., Quinn, C.H. & Stringer, L.C., 2009. Who's in and why? A typology of stakeholder analysis methods for natural resource management. J. Enviro. Manage. 90, 1933-1949.

Zins, A.H., Ponocny, I., 2022. On the importance of leisure travel for psychosocial wellbeing. Ann. Tour. Res. 93, 103378.