SOCIOLOGY | REVIEW ARTICLE

Saturation controversy in qualitative research: Complexities and underlying assumptions. A literature review

Favourate Y. Sebele-Mpofu

Abstract: Judgement of quality in qualitative has been a contested and controversial issue amongst researchers. Contention has always emanated from the subjective nature of qualitative studies, absence of clear guidelines in sampling as well as the lack of generalisability of findings. Numerous avenues have been suggested to improve qualitative research quality and key amongst the suggestions is the concept of saturation. It is viewed as a contemporary measure to alleviate the subjectivity in qualitative research, a yardstick for estimating sample sizes in qualitative research as well as an assurance for rigour and quality. Despite its recognition as a vital tool, it has its own fair share of controversies and contradictions. This research, through a comprehensive and evaluative literature review sought to unpack the saturation puzzle, controversies in definitions and underlying assumptions. The objective was to make a contribution to the contemporary but growing body of knowledge on the saturation conundrum. The study found out that there are various forms of saturation and with varying underlying propositions, therefore in order to meaningfully apply the concept, researchers have to appreciate the forms of saturation, link the appropriate form to their qualitative research design. It

ABOUT THE AUTHOR

Favourate Y. Sebele-Mpofu is a Lecturer in the Accounting Department at the National University of Science and Technology (NUST), Zimbabwe. She currently lectures Taxation, Auditing and Strategic Performance Management. She holds a Masters in Finance and Investment and a Bachelor of Commerce in Accounting from the same University. She also holds a Master of Commerce in Accounting from the Midlands State University and a CIMA Advanced Diploma in Management Accounting. She is interested in tax policy research in developing and emerging economies. She has also researched on issues to do with the challenges faced by qualitative researchers in justifying their methodological choices. The current research on saturation was motivated by the researcher’s interest in taxation, which often covers both qualitative and quantitative research approaches.

PUBLIC INTEREST STATEMENT

Qualitative research has often been criticised for weaknesses in rigour and quality. The concept of saturation has been tabled by various researchers as a tool to enhance quality in qualitative research, especially in providing transparency and guidance in sample size selection. The concept itself is controversial, complicated and contested amongst researchers yet it holds a great deal of potential in improving the quality of qualitative research findings. Disagreements range from the definition to the underlying assumptions, the lack of adequate guidelines on how and when to apply the concept, the different types of saturation as well as how to assess whether the saturation point has been attained. This study therefore, sought to extensively review literature on saturation and to address the problematic areas highlighted earlier to guide qualitative researchers when applying the concept. Qualitative researchers need to define the concept, explain the type chosen and justify its appropriateness and explain the steps taken to ensure the saturation point was reached.
is undoubtedly important for research to define fully the form adopted, explicate the steps followed to achieve it and how it was ultimately achieved. In short, narrow the scope of saturation and contextualise it to your research.

**Subjects:** Education - Social Sciences; Sociology & Social Policy; Cognitive Psychology

**Keywords:** saturation; qualitative research; sampling; sample size; quality

1. **Introduction**

Saturation has become one of the novel and topical issues amongst researchers focusing on how to enhance rigor and validity in qualitative research as well as how to improve the quality and credibility in this approach (Fusch & Ness, 2015; Hennink et al., 2017, 2019; O'reilly & Parker, 2013; Saunders et al., 2018; Sim et al., 2018), despite being an old concept (Glaser & Strauss, 1967; Hennink et al., 2019). It is considered a fundamental: (1) “frequently touted guarantee of qualitative rigor” (Morse, 2015, p. 587) (2) guideline or “gold standard” to inform sample size determination in qualitative research designs (Guest et al., 2006, p. 60) (3) point of “information redundancy” (Sandelowski, 2008, p. 875) or “diminishing returns” (Rowlands et al., 2016, p. 40) (4) juncture at which “information power” is attained (Malterud et al., 2016, p. 2) (5) phase where no additional codes (code saturation) and themes and or further insights (meaning saturation) are emerging from the data (Hennink et al., 2017, p. 14). Interestingly Low (2019, p. 131), considers defining saturation as the point where “new information emerges as “problematic” and a “logical fallacy” that gives little or no advice as to how to achieve that point.

These descriptions are quite intriguing and depict two important aspects on saturation. Firstly, its criticalness and secondly the controversy surrounding perhaps its definition or its conceptualisation. Is it a phase, a rule, a measure or a standard? What is saturation? These questions continue to beg for answers. Explicating the intricacy of the term Morse et al. (2014) allude to the contradiction in meanings that are often attached to the term and the incompatibility in how to gauge it, describe it and even communicate effectual how it was attained in any study. Reiterating the dilemma on the “conceptualisation and operationalisation” of saturation, Saunders et al. (2018, p. 1893) assert that “There appears to be uncertainty on how saturation should be conceptualised and its use”. Putting more emphasis on the paradox, Fusch and Ness (2015) portend that qualitative researchers often find themselves in conundrum on how to address questions such as, what is saturation. How and when does one accomplish it? How does reaching it or not reaching it affects the research? Is the impact the same across qualitative designs, considering they are multiple?

This paper is motivated, firstly by the fact that being a tax researcher using the mixed method exploratory research design for my PhD studies, two different reviewers raised two different questions: Purposive sampling yes, but how did you address saturation when sampling? And the other question was, how did you ensure saturation was achieved in your thematic analysis? These got me thinking on the complicatedness of saturation and the dilemmas researchers go through in dealing with such questions. Secondly, by recommendation by Saunders et al. (2018, p. 1904) who foreground “the need not only for more transparent reporting, but also for a more thorough re-evaluation of how saturation is conceptualised and operationalised, including the recognition of potential inconsistencies and contractions in the use of the concept”. Thirdly, by the fact that “the concept is nebulous and lacks systematization” (Bowen, 2008, p. 139). Fourthly, despite the concept appearing to be crucial in qualitative research, contemporary literature expounding on it is comparatively paltry (Majid et al., 2018), it has been a “neglected” concept (Fusch & Ness, 2015, p. 1408). Lastly, saturation in sample sizes is a crucial aspect used by research reviewers, researchers, supervisors, ethical review committees and funders to assess the productive and acceptable sample sizes (Hennink et al., 2019), yet very scarce methodological research exists on delimitations that mould saturation, sample sizes needed to achieve it and ways to do so (Hennink et al., 2017; Walker, 2012). The various researchers described earlier, point to another controversy. When do we consider saturation in sampling, a priori (proposals and planning), when
conducting data collection or during the analysis stage or even though all stages and how? This study sought to contribute to the emerging theoretical body of literature on saturation, the ongoing argumentation on the subject to widen the discourse on the intricacies and underlying assumptions as well as to evaluate the areas of convergence and divergence among researchers.

2. Saturation definition controversy. What is it? How is it defined?
Saturation has its roots in the grounded theory when it was propounded by Glaser and Strauss (1967), as a means of designing theoretical and interpretative frameworks from qualitative information as cited by various researchers on this subject (Guest et al., 2017; Hennink et al., 2017; O’reilly & Parker, 2013; Sim et al., 2018). The conception has gained momentum in recognition over the years, as a contemporary route to enhance qualitative research potency, bearing in mind that this approach is often criticised for subjectivity, lack of clarity in arriving at samples sizes and problems of generalisability of findings. Despite its growing acceptance, it is marred with controversy. Its definition, nature, purpose and variations in use are subjects of intense debate among scholars (Saunders et al., 2018). According to Low (2019, p. 131), most of the current studies on saturation concentrate largely on how many interviews, how big the sample size or how many focus groups are required to attain saturation point “rather than developing a conceptual and didactic definition of what it is”. Very minimal methodological research is available on the specifications or guidelines that shape saturation, what it entails, how to evaluate it as well as on the specific and transparent parameters on how to accomplish it. Glaser and Strauss (1967, p. 61) described saturation as a parameter for judging when to cease sampling, this being the point where “no additional data are being found where the sociologist can develop properties of the category. And he sees similar instances over and over again, thereby the researcher becomes empirically confident that data is saturated”. The resolution is that the saturation point is defined in relation to the cessation of sampling, shaped by designing of conceptual categories when analysing data, implying that sampling and data analysis occur as combined or concurrent process as opposed to being sequential or stage by stage process. Hennink et al. (2019) posit that this describes theoretical saturation, which leans largely on the sufficiency of the sample, to enable the researcher to generate adequate, logical, relevant and copious data to philosophically buttress emerging models. On another angle, citing Starks and Trinidad (2007, p. 1375), Saunders et al. (2018) advance that theoretical saturation takes place when all the concepts that characterise a theory are fully reflected in the data. This addresses, perhaps elements of “meaning saturation” or the “information power” suggested by Malterud et al. (2016). Data saturation is explained as when evaluative and philosophical adequacy is attained in relation to the guiding theoretical framework. The question to be answered is “do we have sufficient data to illustrate” the theoretical framework underpinning the study? (Saunders et al., 2018, p. 1895).

Hennink et al. (2017, p. 15) define saturation in two forms, code and meaning saturation, these being the stage where “no additional codes are emerging” and where no “further insights” are originating from the data. Re-affirming the former, Urquhart (2012, p. 194) details it as the point where “There are mounting instances of the same codes, but no new ones”. Reiterating the latter, O’reilly and Parker (2013) and Walker (2012) assert that saturation ideal occurs where enough information has been collected to reproduce the study. Fusch and Ness (2015) delineates saturation from the thematic, meaning and coding angles expressing it as the juncture where further coding becomes unfeasible as there is no emerging information, codes or themes from further interviewing. Three forms or definitions of saturation become evident here: code saturation, thematic saturation and meaning saturation (data or information saturation). Are these three the same? Are they different? The controversy in the definition and how to explicate saturation is evident. An increase in the areas of concentration in adjudging saturation becomes apparent. The unfolding of new codes, themes and information becomes the measure to assess analysis. This is a slight deviation from the breadth of the development and refinement of those already determined (Hennink et al., 2019; Saunders et al., 2018), perhaps the information power suggested by Malterud et al. (2016).
From the discussion above, defining saturation appears, problematic and the quandary in which researchers often find themselves visible. How do they define it, from the cessation of analysis angle (Urqhart, 2012), theoretical perspective (theory development) (Glaser & Strauss, 1967) or when all theoretical constructs are fully captured in the data (Starks and Trinidad (2007) as articulated by Saunders et al. (2018), data adequacy (Fusch & Ness, 2015), “informational redundancy” (Guest et al., 2006) or even the narrower angle of saturation at individual interview level where the informant has no new information to provide and their stand point has been fully comprehended (Hennink et al., 2019).

Let’s suppose, for interest sake the contradictory saturation definitions above are well understood, its attainment remains a formidable challenge. Contention surrounds, how to attain it and when as well as which methods are more likely to ensure saturation is reached? (Guest et al., 2020).

3. Saturation applicability controversy: when and how?
Researchers table diverse definitions and accounts of saturation, but they converge on some commonalities in their conceptualisations, such as the point where no new themes, codes and information other than the one already attained from the data and the point where the study can be recreated (Fusch & Ness, 2015; Guest et al., 2006). These common principles display some interconnectedness as no new information normally signifies the achievement of the other concepts on themes, codes and replication (O’reilly & Parker, 2013). Saunders et al. (2018, p. 1900) questions even these commonalities, querying the definition of a theme, stating “However, interpretations at this stage regarding what might constitute a theme, before even beginning to consider whether identified themes are saturated, will be superficial at best”. The way saturation is defined influences the time and context of when and how it can be achieved (Guest et al., 2017; Saunders et al., 2018; Sim et al., 2018). The answers to the when and how questions are influenced by the research design and will accordingly vary (Morse, 2015; Morse et al., 2014). Morse (2015) leans more on the saturation posited in the grounded model by Glaser and Strauss (1967). The variation in qualitative research designs compound the intricacy of the saturation puzzle (content analysis, ethnographic, phenomenological and meta-analysis) together with the multiple methods and instruments of data collection (literature review, focus groups and interviews among others).

There is no one size fit all saturation and “What is data saturation for one is nearly not enough for another” (Fusch & Ness, 2015, p. 1408). This suggests a diversity in parameters. For example, viewing saturation in relation to meta-analysis and phenomenology can mean entirely different aspects of consideration and degrees of saturation. The former depends on reviewing literature from published studies, saturation is constructed upon the previous researchers’ own explication, definition and achievement of saturation, yet the latter requires an in-depth understanding of the phenomenon under investigation from the views and experiences of participants (McKerchar, 2008; Wilson, 2015). Therefore, the latter could require greater saturation points or richer data to do so and the former lower degrees of saturation (Fusch & Ness, 2015). The other complication relates to what yardstick is used to measure saturation, is it the codes, themes or meanings? Hennink et al. (2017, p. 15) submit that the reliance on codes only is a narrow focus or analysis of saturation and “misses the point of saturation”. The researchers suggest code saturation as a preliminary point to build on so as to achieve “meaning saturation”, this being the point where viewpoints, variations, accurate and deep understanding of information are all reflected in the data (Hennink et al., 2017, 2019; Saunders et al., 2018). This implies that focusing on codes alone is a deficient measure of saturation, the codes can be saturated but vital information remain unconsidered. Hennink et al. (2019) interrogates the validity of using themes, suggesting that appraising saturation on the non-emergence of themes is rather a premature assessment, because occurrence alone without comprehension of the themes across data is superficial. Thematic saturation just like code saturation must be considered an initial analysis that lays a foundation for more thoughtful and comprehensive data analysis that pays attention to significance and denotation of the issue at hand as well as to comprehend “the depth, breadth and nuance of the issue” (Hennink et al., 2019).
The question is, when will all the necessary information be captured and how can you tell that all critical information is represented with the data? Researchers provide different answers ranging from when new information becomes redundant, nothing coming up, when the topics are well understood and multiple examples can be used to explain phenomenon and where no new codes or themes emerge (Hancock et al., 2016; Hennink et al., 2017; Malterud et al., 2016). According to Morse (2015, p. 587) when the understanding of the “phenomenon becomes stronger, more evident, more consistent, more comprehensive and more mature”, saturation has been attained. The researcher advances that it is not about hearing it all from the participants, but the fact that saturation is more evident when the research report or publication is comprehensively presented in a competent and confident manner. The resultant outcome is abstract and connected to literature, findings are capable of generalisability and “findings surprise and delight the reader” (Morse, 2015, p. 588). Boddy (2016, p. 428) proposes that saturation attainment provides the study findings with “some degree of generalisability. A notion disputed by Saunders et al. (2018, p. 1899) as they suggest that such a proposition deviates from the idea of “theoretical adequacy” and the “explanatory scope of theory” in research suggested by Glaser and Strauss (1967, p. 61). The researchers argue that such a deviation points to a mix up on the meanings, aim and measure of achieving saturation. The saturation conundrum is visible.

On the other hand Sandelowski (2008, p. 875) points that saturation is reached, when the researcher fully agrees “that the properties and dimensions of the concepts and conceptual relationships selected to render the target event are fully described and that they have captured its complexity and variation”. These elucidations by the researchers to on how to determine saturation attainment, point to subjectivity in the judgement. Even when using thematic saturation to measure saturation achievement (which is explicated as the point where no new themes emerge), Sim et al. (2018) argues that thematic conceptualisation basing on the number of theme occurrences or number of times is flawed because what is important is not the numerical instance but the analytical frame that focuses on meanings and relationships. The frequency of the theme might not comparatively correspond to its impact or contribution to overall research (Roy et al., 2015). The key question to answer when gauging whether saturation has been attained is, Have we exhausted all the “unique dimensions that flesh out, clarify, transform or dimensionalise data that leads to a fully saturated concept?” (Roy et al., 2015, p. 254).

4. Forms of saturations: complexities and underlying assumptions

As highlighted in literature above, researchers define the term differently or don’t define it at all but just proclaim to have reached saturation or where they make an effort to define it, definitions vary (Guest et al., 2020; Low, 2019; Mason, 2010). The variability in interpretations and meanings given to the term has led to some researchers drawing negative conclusions on the concept. For example O’reilly and Parker (2013, p. 190) consider the multi-disciplinary application of the concept of saturation in qualitative research as rather inappropriate yet others emphasise its importance (Hancock et al., 2016; O’reilly & Parker, 2013). Saunders et al. (2018) pronounce the challenges in the “operationalisation and conceptualisation” of saturation and further point out the hazy and often overlapping espousal of the term. They allude to the fact that researchers often combine two or more forms of saturation making its denotation complex and opaque. The researchers identify four types of saturation (theoretical, inductive thematic, a priori thematic and data saturation) and explain what these fundamentally entail and their major focal areas in the research process. The description given seems to overlook the breakdown of theoretical saturation in the two forms tabled in literature (Glaser & Strauss, 1967). “Meaning saturation” or information adequacy saturation is not evident from the models. An adapted table of saturation forms showing those outlined by Saunders et al. (2018) and in other studies are presented in Table 1.

The explanations of the above forms of saturation are themselves overlapping and their presuppositions multiplex. Constantinou et al. (2017, p. 6) advocate that perhaps for a clear delimitation of the different forms of saturation, the question to be addressed is, “what exactly is being saturated?” For example, querying the delineation of **theoretical saturation** as the juncture where
Table 1. Different Forms of Saturation and their Explanations

| Type               | Explanation                                                                 | Focal point                      | References                                                                 |
|--------------------|-----------------------------------------------------------------------------|-----------------------------------|-----------------------------------------------------------------------------|
| Theoretical saturation | Describes the generation of theoretical or conceptual categories as guided by the grounded theory. Relates to when are the concepts or dimensions of a theory are fully reflected in the data. | Sampling Analysis                 | (Glaser & Strauss, 1967; Morse et al., 2014) Starks and Trinidad (2007, p. 1375) as cited in Saunders et al. (2018), (Morse, 2015). |
| Thematic saturation | Inductive-Linked the point where no new codes and/or themes are emerging from the data. A priori hinges on the extent to which the determined codes or themes epitomise or illustrate the data. | Analysis Sampling                 | (Hancock et al., 2016; Hennink et al., 2017; Urquhart, 2012) |
| Data Saturation    | Explicates the level to which new data repeats what was expressed in previous data (data replication). | Data collection and analysis      | (Fusch & Ness, 2015)                                                      |
| Meaning Saturation | Relates to the quality of data, “richness and thickness” when no additional information from the data emerges. Quality, deep, detailed and relevant data has been gathered. | Throughout the research process (planning, data collection and analysis) | (Hennink et al., 2017; Hennink et al., 2019) |

Adapted from Saunders et al. (2018) and enhanced from various researchers.

“no new information” is cropping up from the analysis of data, Low (2019), adduces that the definition is controversial and lacking in some important dimensions. Focusing on just “no new information” overlooks the initial pronouncements by Glaser and Strauss (1967) which focused on theory building and testing, suggesting that the stabilisation of the theory or when data reflect fully the constructs in the theory, saturation point has been achieved. The researcher declares that “the definition provides no didactic guidance on how researchers can determine such a point and is a logical fallacy, as there are always new theoretic insights to be made as long as data continue to be collected and analysed” (Low, 2019, p. 131). In a more compatible opinion theoretical saturation is “specifically intended for the practice of building and testing theoretical models using qualitative data and refers to the point at which the theoretical model being developed stabilises” (Guest et al., 2020).

On code or categories saturation, expressing dissatisfaction on the angle, Morse (2015, p. 588) suggests that what is being saturated is not the categories per se, but instead the features of data within those categories, emphasising that coding in terms of categories robs the research of the recognition of individual experiences of participants. Categorisation should be considered as an initial “step in the processes of conceptualisation, synthesis and abstraction” towards saturation. The researcher asseverates that “saturation is the building of rich data within the process of inquiry, by attending to scope and replication, hence in turn, building the theoretical aspects of enquiry”.

Data saturation depicts a broader use of the conception. In this broadness, saturation is explained as the “point in data collection and analysis when new incoming data produces little or no new information to address the research question” (Guest et al., 2020, p. 2). Critiquing data saturation, Constantinou et al. (2017) table that, “what is saturated is not the data but the categories or themes”. According to the researchers, data is raw views or information collected from study participants and hence can never be saturated because perspectives and words tend to
vary across participants as these are shaped by various factors such as experience, beliefs, occupation, education and understanding of the subject under study. The words or views are grouped according to homogenous characteristics or “commonalities”. In this case what is being considered to have been saturated or as a measure of saturation is not the raw data itself but the categorisation of that data into themes. Saturation is therefore described as the point where “no themes emerge” from the data (Bowen, 2008; Guest et al., 2006) as opposed to where no new data emerges. It is on this line of thought that Constantinou et al. (2017) decide to adopt “themes saturation instead of data saturation”.

Thematic saturation is not without disputation either. Blaikie (2018) poses the question, “what constitutes a theme? Saunders et al. (2018) observe that it is problematic to talk about thematic saturation, without giving a comprehensive definition of a theme, yet they are quick to point out that arriving at that definition is a complicated task, “superficial at its best”. Recapitulating the challenge, Morse (2015) state that there is little evidence on how to accomplish thematic saturation. Further emphasizing the controversy, Braun and Clarke (2016) submits that contrary to the conceptualisation of themes by Fugard and Potts (2015) as ontologically, clear and discrete things that are in the littered in the data, just like “diamonds” waiting to be picked, themes are determined and conceptualised in various ways. Have they been “identified or developed?” (Braun & Clarke, 2016) For example, these can be “imposed on data; discovered in data or constructed from and for data” (Blaikie, 2018). This implies various ontological views and subjectivity in generating themes, adding to complexity of saturation. What are the themes being saturated and how have they been derived and conceptualised? Braun and Clarke (2016, p. 740) maintain that thematic saturation tends to turn a blind eye to the “problematic conceptualisation of a ‘theme’: the reporting of not themes, but of topics or domains of discussion, albeit claiming them as themes”. Low (2019) suggests that themes alone are not an adequate gauge of saturation as they ordinarily become fused into the narratives that answer the research questions. It is therefore not a matter of how recurrent a particular theme is in data but whether the data enable the researcher to fruitfully develop and test evaluative arguments that allow for research objectives to be fully addressed. Highlighting the puzzle even further, Braun and Clarke (2019), continue to interrogate thematic data saturation as a yardstick to gauge the rigor and cogency of qualitative research, tabling the question, “To saturate or not to saturate?” Therefore, questions still continue as to which type of saturation should be considered vital in any study and how should it be achieved, or perhaps the resolution will be influenced by the nature of the study, its design, its objectives and the data collection methods adopted.

Van Rijnsoever (2015, p. 12) emphasises that it is important for qualitative researchers not to focus solely on the occurrence of themes but more on the characteristics and meaning of concepts reflected in the data to make meaningful assessments. This explains “meaning saturation” (Hennink et al., 2017). Reitering meaning saturation Sim et al. (2018) propose that it is essential to consider not only how many times the theme emerges but its analytic conceptualisation, thus move from descriptive meaning of the theme to its interpretive cogency.

5. Qualitative research designs and the saturation paradox
The importance of saturation in qualitative research is explicated from two seemingly related angles: sample size determination (Guest et al., 2006, 2017) and enhancement of research quality and validity (Hancock et al., 2016). Qualitative researchers often find themselves in a predicament when striving to address these two important areas in qualitative research. There is imprecision and lack of clarity in the methodological conceptualisation of the saturation notion, “especially providing no description of how saturation might be determined and no practical guidelines for estimating the sample sizes for purposively sampled interview” (Guest et al., 2006, p. 60). The various forms of saturation explicated by researchers: theoretical, data, thematic, meaning and code saturation compound the quandary that the researchers find themselves in (Guest et al., 2020; Malterud et al., 2016; Rowlands et al., 2016; Saunders et al., 2018). What is saturation, which of the forms of saturation do they seek to address in the their research, how are they going to achieve it and how does accomplishing the chosen form impact on the other forms as well as on
research validity, are some of the difficult questions that researchers have to contend with (Fusch & Ness, 2015; Morse et al., 2014; O’reilly & Parker, 2013).

On sampling and sample size determination, the controversy lies on the fact that contrary to the quantitative approach where sample size decisions are guided by some cardinal principles such as the N rule and confidence levels, for a qualitative researcher the process is fraught with “subjectivity and arbitrariness” (Rowlands et al., 2016, p. 40). It is a matter of judgement, yet a relevant and representative sample must allow the research to address the fundamental measures of validity in qualitative research such as rigor, credibility of findings, conformability, trustworthiness and acceptability (Fetters et al., 2013). Saturation point consideration is argued to be vital in the resolution of this conundrum of sample size assessment, although most qualitative researchers fail to define their samples, sample sizes, explain how they addressed saturation in choosing their sample and others just allude to the fact that they reached saturation but without adequate elaboration on how and when aspects (Guest et al., 2017; Marshall et al., 2013). Some researchers would, for example state that saturation was achieved at between 12 and 30 interviews. This explains very little regarding the sample size and offers no justification for it or when the sample was chosen, was it a priori (Rowlands et al., 2016) or during the data collection stage study, perhaps the “interviewing until saturation” (Guest et al., 2020, p. 2) or during the analysis stage. Morse (2000) points out that saturation is largely declared and not explained by researchers, notwithstanding that complexities in measuring saturation in real life contexts are immense.

The other confusion in the saturation in qualitative research puzzle has to do with differences in the research designs as well as the data collection methods used such as literature review, observation, interviews and focus groups. These are discussed briefly in 5.1 and 5.2, with 5.3 covering the intricacy in sample size estimation.

5.1. Qualitative research designs and the saturation puzzle

Blaikie (2018) alludes to the fact that qualitative research is quite broad and that the term is often imprecisely used in a blanket manner ignoring the different logics of inquiry that characterise the research domain (induction, abduction, deduction, retroduction) and the varying epistemological assumptions that define each logic. (This was not delved into in detail in this research). As highlighted earlier the multiplicity of qualitative research designs is commonly problematic with regards to saturation as there is no blanket form of saturation. The type and breadth of saturation is often influenced by the chosen research design. The responses to the questions, when and how are shaped the research design (Fusch & Ness, 2015). What is considered saturation or even the appropriate level of saturation might vary contextually from one design to the other. For example, is it meta-analysis, ethnography or phenomenology. Meta-analysis study could possibly require lower levels of saturation because they are constructed on studies which in most cases would have addressed saturation point (Fusch & Ness, 2015). This argument is open to debate because researchers allude to the failure by most qualitative researchers to expound on saturation and how it was reached, others claim it without giving relevant facts to back up their claims (Marshall et al., 2013; Mason, 2010; Rowlands et al., 2016). Looking at the focus of ethnography and phenomenology research designs, these could demand for high degrees of saturation (Fusch & Ness, 2015, p. 1409). Saturation refers to different aspects to different researchers and suffers from inconsistency in evaluation and reporting (Morse et al., 2014; Tran et al., 2017). For example pointing to the ideal sample size to achieve saturation, Roy et al. (2015) citing Morse (1994) proposed 6 interviewees for phenomenological studies, 30 to 50 interviews or observations in the case of ethnographic and grounded theory studies and 100 to 200 sample participants where the study is ethological in nature. Marshall et al. (2013) suggests that for grounded theory qualitative studies, a sample of 20 to 30 interviews is more appropriate and that single case studies should ordinarily hold 15 to 30 interviews. For meta-analysis studies, multiples of 10 were found to be sufficient (Mason, 2010). Already these proposed sample sizes point to a diversity in qualitative studies and the point of saturation variation yet qualitative
researchers rarely give these details as bemoaned by Morse (2015), Guest et al. (2020), and Saunders et al. (2018).

The absence of compatibility in definitions and forms of saturation reflects a broadness in the term saturation and at same the controversy that compromises transparency and translates to poor reproducibility of studies and jeopardises rigour as well as the depth of the study (Constantinou et al., 2017). These are the very attributes that saturation seeks to enhance. In relation to the broadness, Saunders et al. (2018, p. 1893) argue that saturation operationalisation should be informed by the “research question (s), theoretical position and the analytical framework adopted”. This suggests the need to narrow the scope of saturation conceptualisation so as to preserve its purview or perhaps to contextualise. Reiterating the concern the researchers emphasise that for saturation to be “conceptually meaningful and practically useful”, its scope of application must be constricted and properly defined (Saunders et al., 2018, p. 1899).

With regards to the variation of qualitative research designs, for example, saturation point for phenomenology and that of meta-analysis will expectedly vary. Sim et al. (2018, p. 626) posit that “in the phenomenology approach, the effect on a sample size is mediated through the richness of the data obtained from individual informant”. Malterud et al. (2016) and Sim et al. (2018) posit that sample size determination is also dependent on the nature of evaluation strategy, for example a research working towards an in-depth and gaining a complete picture of a phenomenon from few informants will suffice with a smaller sample size.

Table 2 makes a summary of some of the qualitative research designs and their foci as well as the multiple data collection methods that can be used to collect data, thus heightening the saturation point complexity. Each collection method has its aim and its associated challenges when it comes to saturation.

5.2. Sample size determination and the saturation point predicament
In spite of the fact that saturation is increasingly gaining ground as a tool to estimate sample sizes in qualitative research, how part of it is still fraught with confusion (Guest et al., 2020, p. 1). Morse (2015, p. 587) proclaims that “Saturation as the most frequently touted guarantee of qualitative rigour offered by authors to reviewers and readers, yet it is the one that we know least about”. Various sample sizes (Guest et al., 2006, 2017; Hennink et al., 2017) have been tabed by various

| Qualitative Research Design | Purpose of Design | Data Collection methods |
|-----------------------------|-------------------|-------------------------|
| Ethnography                 | Cultural interpret data. Important features are naturalism, small samples, multiple data sources, ‘emic’ and ‘etic’ (Lambert et al., 2011) | Participant observation mostly used, In-depth interviews and focus groups suitable for use |
| Phenomenology               | Aims to understand phenomenon in-depth through studying human experience. Meaning is derived from the feelings, perceptions and cognitions (McKerchar, 2008; Wilson, 2015) | In-depth inter views suitable for use |
| Grounded                    | Seeks to build theory from a research situation. It is iterative and integrative research design (McGhee et al., 2007) | Participant observation rarely used, in-depth and focus groups interviews suitable for use |
| Content Analysis            | Systematically analyse textual data, making replicable and valid inferences with the hope of generating knew knowledge, insights as well as condense the data (Elo & Kyngäs, 2008) | Participant observation used sometimes, in-depth inter views and focus groups suitable for use |

Adapted from Moser and Korstjens (2018).
researchers and diverse methodologies too (Constantinou et al., 2017; Guest et al., 2020; Hennink et al., 2019; Tran et al., 2017) to address the paucity in guidelines of determining the appropriate sample size to reach saturation point and the methodologies to be employed to accomplish saturation. The shortcomings and lack of clarity in the saturation definition and its dimensions still pose a challenge for researchers perhaps limiting the adoption of the proposed methods. For example, Constantinou et al. (2017, p. 2) avails the Comparative Method for Theme Saturation (COMeTS). It is argued to be easy, comparative and inclusive. The limitation, being that, it might be time consuming and complex for larger and more qualitative studies. It could also be challenging to adopt in studies that largely rely on observation and unstructured sources of data when collecting data.

Questions persist on, when do we estimate the sample size, what is the right sample size (how many participants?), what guidelines do we follow in estimating the sample size that enables saturation point to be attained. Random sampling is difficult and “impracticable” for qualitative research, but purposive, judgemental and theoretical sampling are more feasible (Sim et al., 2018). The various sampling methods that are available to researchers also perpetuating the controversy in addressing saturation in the absence of any guidelines. Some of the other most frequently raised question is, when do we reach saturation in a sample (a priori, during data collection or analysis)? An often advanced edict of qualitative researchers is to collect data until saturation point is attained, but very little rationale has been given for this assertion in regards to the principles that underlie saturation (Blaikie, 2018; Low, 2019; Morse, 2000). Perhaps the issue of the widthness of the concept of saturation explains it problematic application in sampling. For example, which form of saturation is appropriate for which type of qualitative sampling, considering the multiplicity of the sampling techniques. In consonance, Saunders et al. (2018, p. 1899) express that “there is a risk that saturation is losing its coherence and utility if its potential conceptualisation and uses are stretched too far”. Moser and Korstjens (2018) outline the different sampling methods that are at the disposal of qualitative researchers to employing in choosing a sample depending on the purpose of the research and the characteristics of the target population and these are summarised in Table 3.

| Type of Sampling           | Description                                                                 |
|----------------------------|-----------------------------------------------------------------------------|
| Purposive Sampling         | Selection of participants based on the researcher’s personal judgement, based on the informative nature or “information power” of participants. For example experience, institutional memory, specificity, purpose of the study and their relevance to the study. |
| Criterion                  | Choosing participants on the basis of a pre-determined criteria of importance. |
| Theoretical                | Choice of participants is driven by emerging findings to ensure sufficiency in addressing theoretical concepts that are key to the research. |
| Convenience                | Sampling is based on availability, the readily and easily available.         |
| Snowballing                | Selection is informed by referrals by participants previously selected. For example, one tax consultants refers you to two other more knowledgeable and experienced tax consultants that he knows from their association in the tax field. |
| Maximum Variation          | Choice of participants based on a broad range of variations in the backgrounds of these participants. |
| Extreme Case               | Purposeful choosing of the most unusual cases.                              |
| Typical Case               | Most typical and average participants are chosen.                           |
| Confirming and Disconfirming| Sampling that is meant to support checking or challenging of emerging trends or patterns in the data |
The various methods of sampling would entail different points of saturation and equally different sample sizes, but the challenge is the lack of appropriate guidelines in literature to say in regards to sampling. Other researchers suggest that sampling adequacy should be driven by saturation and replication (Low, 2019) yet others suggest that it must be based on whether enough data to explain all the key components of the phenomenon under study can be collected (Mason, 2010). Roy et al. (2015) emphasise the comprehensiveness of the field work as what determines the sufficiency of the sample not the sample size per se.

5.2.1. What shapes sample sizes determination with regards to saturation?
Sampling decisions must be guided by the research objectives and the need to collect thick and rich data that is data of appropriate quality and of the right quantity, respectively (Fusch & Ness, 2015). The sample selected must enable the research to collect adequate information “to produce a corpus from which they can draw qualitative conclusions”. (Rowlands et al., 2016, p. 43). Sim et al. (2018) allude to the a priori sample size estimation. A priori sample size determination is generally driven by the need for researchers to address the demands from funders, reviewers and ethical clearance bodies including planning resource allocation. These sample sizes are used to assess the practical aspects of the standard, subjectivity or objectivity of the study and the likely issues of validity and ethical consideration that might originate from the study. This to some extent justifies sampling a priori, but despite this justification, Saunders et al. (2018, p. 630) considers a priori sample size adjudging implausible especially in inductive exploratory research.
In a similar opinion, Sim et al. (2018) acknowledge that estimating sampling sizes a priori is inherently complicated as sample size determination is an “adaptive and emergent” process influenced by the stage of “information redundancy” (Braun & Clarke, 2016, 2019; Saunders et al., 2018) or the theoretical standpoints that originate as data goes through evaluation (O’Reilly & Parker, 2013). Sim et al. (2018, p. 630) assererate that “... a firm judgement on the number of participants ultimately required to reach saturation can only be reached once the study is under way”. Malterud et al. (2016, p. 1757) warn against definitively and conclusively determining sample sizes a priori and assert that instead, it must be a matter that is taken upon as journey, with revisiting, revising, redefining and refining the sample size throughout the research leaning on issues such as saturation point as well as thickness and quality of data. This points to perhaps a “posteriori” sample size determination (Sim et al., 2018, p. 620). The researchers put emphasis on that sampling should not be a matter of how many interviews are held or how many participants were interviewed but that of who are they? What knowledge and competences do they possess which is relevant to the study as well as to the drawing of credible conclusions?

Re-affirming that just the number of participants is an insufficient basis of sample size, Hennink et al. (2017) underscore the need to pay attention to both “code saturation and meaning saturation”. Sandelowski (2008) states that sampling is much more than the number of participants but their experiences as well. On a similar vein, Hammersley (2015) avows that researchers must consider the relevance and richness of participants’ knowledge or information to the development of the research and theoretical insights. Therefore, it is evident that sample size estimation should be a process considered throughout the study, as the decision can be altered over the process of data collection and analysis. The number of interviews you will need will change day to day as you learn more and revise your ideas” (Baker & Edwards, 2012, p. 15). Epistemological view, aims and objectives of the study will guide the sampling process also. It not just about the number of participants but the appropriateness of the data that has been collected in the context of the angle of the research.

Malterud et al. (2016, p. 1756) advance that researchers must consider “information power” when selecting the participants and sample sizes to avoid “producing that which is already known”. Information power is built on “(a) the aim of the study (b) sample size specificity (c) use of established theory (d) quality of dialogue (e) analysis strategy”(Malterud et al., 2016, p. 1756). “A study will need the least amount of participants when the study aim is narrow, if the combination of participants is highly specific for the study aim, it is supported by established
theory, if the interview dialogue is strong and if the analysis includes longitudinal in-depth exploration of narratives or discourse details” (Malterud et al., 2016, p. 1757). The more the information power the sample holds the lower the number of participants needed. The more knowledgeable the participants, the richer the discussion and the lower the sample size needed. Complexity is visible in the information power suggestion, for example on quality of dialogue, there is an element of subjectivity as the quality of communication is not only dependent on the knowledge and competence of participants but also on the creation of rapport and interviewer skills. It is challenging to foretell the articulateness of participants in advance and the interviewer skills and bias can compromise the whole process of data collection even with the appropriate sample. The above arguments imply that the sample size resolution is a continuous process throughout the research, a priori, evaluated on an on-going basis and appraised for its adequacy in terms of analysis and publishing of results in the final state, it is thus a stage by stage process (Guest et al., 2006)

5.2.2. How to choose a sample that enables the accomplishment of saturation
How saturation was reached and demonstrated in a study provides justification for methodology as well as clarity of reasoning (Sim et al., 2018). The lack of transparency and rationalisation of sample sizes, sampling techniques and underlying presumptions compromise the credibility and validity of most qualitative researches. Researchers sometimes proffer unsubstantiated opinions that saturation was accomplished, but the how and when remain unaddressed (Malterud et al., 2016). The sample size justification and the choice of sampling techniques must strike a balance with other procedures of data collection to avoid the uneven prominence over others. Sim et al. (2018), p. 630, citing Emmel (2013, p. 154) pronounce that “it is not the number of cases that matters, it is what you do with them”.

Sim et al. (2018) submit four methods of choosing sample sizes: the rules of thumb, conceptual models, numerical guidelines derived from empirical studies and statistical formulae. These have their advantages and shortcomings.

5.3. Qualitative data collection methods and saturation
There are various methods that can be used to collect qualitative data such as observation, literature review, interviews and focus groups. Different forms of saturation as well as different degrees will go along with different research methods. Questions that arise include those such as the number of interviews to be held before saturation point can be reached or how many participants to be sampled as well as perhaps how many focus group discussions to be held and how many participants per focus group. Kuzel (1992) suggests 6 to 8 interviews when researching on a homogenous sample. Hammersley (2015) argues that it is not the number of participants that is vital, the issue is which informant make up the sample. The bottom line is, Can we get enough information or do we have enough information from the sample to fully capture the complexity of the phenomenon under investigation? (Mason, 2010). What is of importance is that the interviews must be adequate to rationalise the claims and conclusions drawn by the researcher. Several research have been conducted on sample sizes especially with regards to adequate sample sizes in interviews and focus groups (Baker & Edwards, 2012; Guest et al., 2017; Hennink et al., 2017, 2019). Section 5.3.1 and 5.3.2 will summarise some these studies respectively.

5.3.1. Adequate sample sizes to reach saturation in interviews controversy
Questions have been raised concerning, how many interviews are enough and with varying answers. Baker and Edwards (2012) table that the appropriate answer is, “it depends”, when asked, “How many interviews are enough in qualitative research?” The next question will be, it depends on what? Baker and Edwards (2012) avow that it is dependent upon several factors are key among them: saturation, minimum requirements of sample sizes in qualitative research, theoretical underpinnings of the study, heterogeneity of the population and the breadth and scope of research questions. Diverse sample size for interviews to enable saturation have been suggested by various researchers (Constantinou et al., 2017; Guest et al., 2006;
Mason, 2010; Roy et al., 2015; Tran et al., 2017). Within these many studies other researchers argue for small samples (Creswell, 2014; Guest et al., 2006; Roy et al., 2015) whereas other are in favour of bigger samples and question the credibility of smaller samples (Mason, 2010; Tran et al., 2017). Guest et al. (2006) and Roy et al. (2015) discourage large samples pointing to the complexity in data analysis that could end up compromising meaningful exploration of the collected data and lead to the failure of addressing research questions fully or worse still the researcher can fail to appropriately contextualise the data to the research. Those in favour of large samples argue that they allowed for a wider investigation, ensure diverse opinions are collected and provide thick and rich data yet others argue that the richness and thickness depend on the quality of informants and interviewer’s professionalism (Malterud et al., 2016) and skilfulness in creating rapport and gathering data or intensity of the data gathering (Roy et al., 2015). Some researchers argue that the size of the sample depends on whether the sample is from a homogeneous or a heterogeneous population, the more uniform the lower the number required and the more diverse the wider the sample. Guest et al. (2006, p. 78) state that “a sample of six interviews may be sufficient for the development of meaningful themes and useful interpretations”, though they acknowledge that this is not always the case. Researchers should not lean on this suggestion to justify the conducting of flawed research with poorly identified samples. Reiterating the relevance of smaller samples especially when employing in-depth interviews, Rosenthal (2016) states that generalisability is not the fundamental target of in-depth interviews, but the major aim is to build a deeper understanding of the meaning behind behaviour, through an appreciation of the experiences, views and perceptions of participants, thus smaller samples are more ideal.

Hennink et al. (2017) also suggest that the number of interviews or participants will also differ in in relation to the type of saturation targeted by the research, is it perhaps meaning or code saturation? Constantinou et al. (2017) posit that saturation or the adequacy of sample size to reach saturation point can differ depending on the order with which the interviews were conducted and analysed. The researchers firstly analysed their 12 interviews in the order of how they were conducted, starting with the first and reached saturation at the 5th interview. They then reordered the analysis using reverse order and attained saturation when analysing the 7th interview. They therefore concluded that saturation point varies and is anchored on the order of interviews during analysis. Table 4 presents a summary of some selected studies conducted on saturation and suggested sample sizes.

| Sample size                          | Rationale                                                                 | Researcher               |
|-------------------------------------|---------------------------------------------------------------------------|--------------------------|
| 12 semi-structured interviews       | Convenient sampling, initial sample based on suggestions by Mason (2010) | (Constantinou et al., 2017) |
| 6 to 12 interviews                   | More comprehensive data can be collected from in-depth. Small samples more appropriate in homogeneous sample and experts in the field | (Guest et al., 2006; Kuzel, 1992) |
| 5 interviews for code saturation   | Code saturation Meaning Saturation                                        | (Hennink et al., 2017)   |
| and more interviews in order to     |                                                                           |                          |
| achieve saturation on meanings      |                                                                           |                          |
| 25 to 200 interviews                | Open ended surveys                                                       | (Tran et al., 2017)      |
| 6 interviews                        | Large samples could compromise credibility and contextualisation of data and quotations | (Morse, 2000)            |

Source: Author’s compilation from various sources.
Table 5. Studies on Focus groups, suggested numbers and Justification

| Number of focus groups | Justification                                                                                                                                                                                                 | Researcher(s)                      |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| 6 focus groups         | Code saturation can be easily attained (By 6th interview 94% of all codes had been attained and 96% of high prevalent codes had been identified), yet meaning saturation entails comprehending issues fully which is not easy. | (Hennink et al., 2019)            |
| 6 Focus groups         | 64% of codes generated at 1st interview, 84% of the code at 3rd interview and 80 to 90% of the thematic codes emerged at 6th focus group.                                                                    | (Guest et al., 2017)              |
| 5 Focus groups         | Used maximum variation sampling to create a diverse sample. Employed inductive approach to generate themes and deductive approach in applying the themes.                                                      | (Coenen et al., 2012; Hancock et al., 2016) |

Source: Various Studies.

5.3.2. The intricacy on the right sample sizes to achieve saturation in focus groups

Different numbers of focus groups to achieve saturation have been suggested by researchers (Guest et al., 2017; Hennink et al., 2019) as well the number of participants per focus group. Hennink et al. (2019) suggests six parameters influencing saturation in focus groups: study purpose, type of codes, group stratification, the number of groups per stratum, type and degree of saturation. Hancock et al. (2016) suggested that saturation could be sought to be achieved in identifying themes in three different ways in focus groups: by individual participant, by focus group and day of data collection. Hennink et al. (2019, p. 4) brings out controversial factors (group dynamics, group format, demographic stratification, group composition) that are often overlooked when discussing saturation or proposing the number of focus groups to be used to reach saturation. The researcher table that group format might influence viewpoints or equally compromise “narrative depth and understanding of issues”. Demographic stratification affects saturation and samples sizes. (Hennink et al., 2019, p. 4). Such issues are neglected in the literature that expounds on saturation. Table 5 below summarised studies on proposed number of focus groups to be held that can help achieve saturation. Researchers must always bear in mind that it is not just about how many focus groups but how these are conducted, the skills of the moderator and many other considerations (Hennink et al., 2019).

6. Conclusion

Saturation is a very important aspect in qualitative research where samples cannot be estimated with certainty. Controversy surrounds its definition, application and underlying principles. It is viewed as vital for sampling and enhancing the quality of qualitative research. The study explored the intricacies surrounding the concept through a review of published studies on the subject. What became evident in these studies is the convolution in defining the term, the complexity in clearly delineating the different forms of saturation, their interconnectedness and underlying assumptions, the lack of clear methodological guidelines on the application of the concept when sampling, collecting data and analysing it and lastly the intricacy in measuring it. Despite all these complications it was also visible that the concept plays a fundamental role in boosting research quality. Saturation is important in sampling, research process and analysis. An adequate sample must be selected to accomplish saturation of theory, themes, codes, data and meaning. Saturation allows for analysis of both objective and subjective evidence. Analysis of the apparently visible (code and themes) and the hidden (meaning saturation) information is pertinent in data analysis and interpretation and they aid in the understanding of complex phenomenon under research. The researcher can adequately contextualise quotations, combine them with interpretative discussions to fully communicate the research story. Interpretation and conceptualisation
can be balanced. This research recommends that researchers must understand saturation so that they can tell a convincing story when they define the concept of saturation in relation to their own research. They must explain fully which form of saturation they targeted, give the reasons why, elucidate on their journey on how they achieved it and when? Researchers must also strive not to let their pursuit of saturation overshadow other important measures of quality in qualitative research such as: credibility, diversity, conformability, trustworthiness and reliability.

Funding
The author received no direct funding for this research.

Author details
Favourite Y. Sebele-Mpofu1
E-mail: favourite.sebele@nust.ac.zw
1 National University of Science and Technology NUST, Zimbabwe.

Citation information
Cite this article as: Saturation controversy in qualitative research: Complexities and underlying assumptions. A literature review, Favourite Y. Sebele-Mpofu, Cogent Social Sciences (2020), 6: 1838706.

References
Boker, S. E., & Edwards, R. (2012). How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research.
Blaikie, N. (2018). Confounding issues related to determining sample size in qualitative research. *International Journal of Social Research Methodology*, 21(5), 635–641. https://doi.org/10.1080/13645579.2018.1456444
Boddy, C. R. (2016). Sample size for Qualitative Research. *Qualitative Market Research: An International Journal*, 19(4), 426-432. https://doi.org/10.1108/QMR-06-2016-0053
Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: A research note. *Qualitative Research, 8* (1), 137–152.
Braun, V., & Clarke, V. (2016). *Mis* conceptualising themes, thematic analysis, and other problems with Fugard and Potts’(2015) sample-size tool for thematic analysis. *International Journal of Social Research Methodology, 19*(6), 739–743. https://doi.org/10.1177/1364557916639015
Braun, V., & Clarke, V. (2019). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, 1–16. https://doi.org/10.1080/2159676X.2019.1704846.
Coenen, M., Stamm, T. A., Stucki, G., & Cieza, A. (2012). Individual interviews and focus groups in patients with rheumatoid arthritis: A comparison of two qualitative methods. *Quality of Life Research, 21*(2), 359–370. https://doi.org/10.1007/s11136-011-9943-2
Constantinou, C. S., Georgiou, M., & Perdikogianni, M. (2017). A comparative method for themes saturation (CoMeTS) in qualitative interviews. *Qualitative Research, 17*(5), 571–588. https://doi.org/10.1177/1468794116686650
Creswell, J. W. (2014). A concise introduction to mixed methods research. Sage Publications.
Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x
Fetter, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—principles and practices. Health Services Research, 48 (6 Pt 2), 2134-2156. https://doi.org/10.1111/1475-6773.12117
Fugard, A. J., & Potts, H. W. (2015). Supporting thinking on sample sizes for thematic analyses: A quantitative tool. *International Journal of Social Research Methodology, 18*(4), 669–684. https://doi.org/10.1080/13645579.2015.1005453
Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. The *Qualitative Report, 20*(9), 1408.
Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Strategies for qualitative research. *The Sociology Press.*
Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59-82. https://doi.org/10.1177/1525822X05279903
Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PloS One, 15*(5), e0232076. https://doi.org/10.1371/journal.pone.0232076
Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field Methods, 29*(1), 3-22. https://doi.org/10.1177/1525822X16639015
Hammersley, M. (2015). Sampling and thematic analysis: A response to Fugard and Potts. *International Journal of Social Research Methodology, 18*(6), 687–688. https://doi.org/10.1080/13645579.2015.1005466
Hancock, M. E., Amankwoa, L., Revell, M. A., & Mueller, D. (2016). Focus group data saturation: A new approach to data analysis. *The Qualitative Report, 21*(11), 2124.
Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Qualitative Health Research, 27*(4), 591–608. https://doi.org/10.1177/104973231665344
Hennink, M. M., Kaiser, B. N., & Weber, M. B. (2019). What influences saturation? Estimating sample sizes in focus group research. *Qualitative Health Research, 29*(10), 1483-1496. https://doi.org/10.1177/1049732318821692
Kuzel, A. (1992). Sampling in qualitative inquiry. In B. Crabtree & W. Miller (Eds.), *Doing qualitative research* (pp. 31–44). Sage.
Lambert, V., Glacken, M., & Mc Carron, M. (2011). Employing an ethnographic approach: Key characteristics. *Nurse Researcher, 19*(1), 17–24. https://doi.org/10.7748/ir2011.19.1.17.c8767
Low, J. (2019). A pragmatic definition of the concept of theoretical saturation. *Sociological Focus, 52*(2), 131–139. https://doi.org/10.1080/003803272018.1544514
Majid, M. A. A., Othman, M., Mohamad, S. F., & Lim, S. A. H. (2018). Achieving data saturation: Evidence from a qualitative study of job satisfaction. *Social and Management Research Journal, 15*(2), 66–77. https://doi.org/10.24319/smrj/v15i2/4972
Molterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research, 26*(13), 1753–1760. https://doi.org/10.1177/104973231561744
Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research? A review of qualitative interviews in IS research. *Journal of Computer Information Systems, 54*(1), 11–22. https://doi.org/10.1080/08874417.2013.1165667

Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews [Paper presentation]. Forum qualitative Sozialforschung/Forum: qualitative social research.

McGhee, G., Marland, G. R., & Atkinson, J. (2007). Grounded theory research: Literature reviewing and reflexivity. *Journal of Advanced Nursing, 60*(3), 334–342. https://doi.org/10.1111/j.1365-2648.2007.04436.x

McKerchar, M. A. (2008). Philosophical paradigms, inquiry strategies and knowledge claims: Applying the principles of research design and conduct to taxation.

Morse, J. M. (2000). Determining sample size. Sage Publications Sage CA.

Morse, J. M. (2015). *Data were saturated.* Sage Publications Sage CA.

Morse, W. C., Lowery, D. R., & Steury, T. (2014). Exploring saturation of themes and spatial locations in qualitative public participation geographic information systems research. *Society & Natural Resources, 27*(5), 557–571. https://doi.org/10.1080/08941920.2014.888791

Moser, A., & Korstjens, I. (2018). Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis. *European Journal of General Practice, 24*(1), 9–18. https://doi.org/10.1080/13814788.2017.1375091

O’Reilly, M., & Parker, N. (2013). ‘Unsatisfactory Saturation’: A critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research, 13*(2), 190–197. https://doi.org/10.1177/1468794112463106

Rosenthal, M. (2016). Qualitative research methods: Why, when, and how to conduct interviews and focus groups in pharmacy research. *Currents in Pharmacy Teaching and Learning, 8*(4), 509–516. https://doi.org/10.1016/j.cptl.2016.03.021

Rowlands, T., Waddell, N., & McKenna, B. (2016). Are we there yet? A technique to determine theoretical saturation. *Journal of Computer Information Systems, 56*(1), 40–47. https://doi.org/10.1080/08874417.2015.11665799

Roy, K., Zvonkovic, A., Goldberg, A., Sharp, E., & LaRossa, R. (2015). Sampling richness and qualitative integrity: Challenges for research with families. *Journal of Marriage and Family, 77*(1), 243–260. https://doi.org/10.1111/jmfm.12147

Sandelowski, M. (2009). Theoretical Saturation. The Sage Encyclopedia of Qualitative Research Methods, (2), 875–876.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity, 52*(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Sim, J., Saunders, B., Waterfield, J., & Kingstone, T. (2018). Can sample size in qualitative research be determined a priori? *International Journal of Social Research Methodology, 21*(5), 619–634. https://doi.org/10.1080/13645579.2018.1454643

Tran, V.-T., Pocher, R., Tran, V.-C., & Ravaud, P. (2017). Predicting data saturation in qualitative surveys with mathematical models from ecological research. *Journal of Clinical Epidemiology, 82*(71–78), e72. https://doi.org/10.1016/j.jclinepi.2016.10.001

Urquhart, C. (2012). *Grounded theory for qualitative research: A practical guide.* Sage.

Van Rijnsoever, F. (2015). (I Can’t Get No) Saturation: A Simulation and Guidelines for Minimum Sample Sizes in Qualitative Research. *Innovation Studies Utrecht (ISU) Working Paper Series, 15*(5), 1–25.

Walker, J. L. (2012). Research column. The use of saturation in qualitative research. *Canadian Journal of Cardiovascular Nursing, 22*(2).

Wilson, A. (2015). A guide to phenomenological research. *Nursing Standard* (2014+), 29(34), 38. https://doi.org/10.7748/ns.29.34.38.e8821
