(Mis)conceptualising themes, thematic analysis, and other problems with Fugard and Potts' (2015) sample-size tool for thematic analysis

Virginia Braun and Victoria Clarke

Commentary published in *International Journal of Social Research Methodology*, *19*(6), 739-743.

One of us (VC) was having a conversation with a student recently about the origins and history of thematic analysis (TA). The student had read *Qualitative Research in Counselling and Psychotherapy* (McLeod, 2011), a text which presents TA as a variant of grounded theory. Victoria commented that she thought that TA evolved from content analysis, and therefore predated grounded theory, and discussed her recent discovery of the use of a variant of TA in psychotherapy research in the 1930s-1950s. The student let out a heavy sigh and slumped in her chair, bemoaning her ability to ever fully grasp qualitative research in all its complexity. This reaction is not uncommon. Students learning and implementing qualitative research at times find it bewildering and challenging; simple models of 'how to do things' can appear to offer reassuring certainty. But simplified models, especially if based in confidently-presented-yet-partial accounts of the field or an approach, at best obfuscate and at worst lead to poor quality research.

In our discipline (psychology), students typically learn about qualitative research only *after* they have been fully immersed in the norms, values and methods of scientific psychology. Many find it difficult to let go of what we call a 'quantitative sensibility'. For such students, and others not well versed in a *qualitative* sensibility, Fugard and Potts' (2015) tool for determining sample sizes in TA research has great intuitive appeal; it provides a life-raft to cling to in the sea of uncertainty that is qualitative research. Thus, we share Hammersley's (2015) concerns that their tool will be used by funding bodies and others (e.g. editors, reviewers) to determine and evaluate sample sizes in TA research. We fear it will result in further confusion about, and further distortion of, the assumptions and procedures of qualitative research. We here build on concerns expressed by others (Byrne, 2015; Emmel, 2015; Hammersley, 2015) to briefly highlight why this quantitative model for qualitative sampling in TA is problematic, based on flawed assumptions about TA, and steeped in a quantitative logic at odds with the exploratory and *qualitative* ethos of much TA research.

(Mis)conceptualising thematic analysis: Thematic analysis is not one approach

Despite acknowledging diversity through reference to 'thematic analyses' (and citing authors with different orientations, such as ourselves – Braun & Clarke [2006, 2013] – Boyatzis [1998] and Guest, MacQueen & Namey [2012]), Fugard and Potts implicitly treat TA as a homogenous entity. They do not highlight differences in underlying assumptions and procedures between these approaches,

differences which are not insignificant, and are consequential for the applicability of their model. Although most TA proponents agree that TA is theoretically independent or flexible, and that coding occurs at two levels – semantic or manifest meaning; latent or implicit meaning – there is less TA scholars and researchers share in common or agree on. Claimed theoretical independence of TA is often limited by unacknowledged theoretical assumptions. There is no widely agreed on definition of a theme, with conceptualisations of a theme varying widely; procedures to identify themes also vary. As we discuss these problems, we contrast two broad approaches to TA, which we call 'coding reliability' (authors such as Boyatzis, 1998; Guest et al., 2012; Joffe, 2012) and 'organic' (ourselves and others) – these and other models of TA we discuss further elsewhere (Clarke & Braun, forthcoming).

(Mis)conceptualising themes: not everyone views themes as diamonds

For Fugard and Potts' model to work, a very particular idea of what a theme is, and how it can be identified, is required. Essentially, the model has to conceptualise themes as ontologically real, discrete things, out there in the world (or the data), identifiable by researchers – like diamonds scattered in the sand, waiting to plucked-up by a lucky passer-by (though within the paper, there are varied definitions of 'a theme'; Emmel, 2015). That Fugard and Potts implicitly (as well as explicitly) regard analysis as a process of theme-discovery is evidenced in the language used: "to have a chance of *capturing* themes" (p. 7; our emphasis); "in order to aid the *recognition* of a theme" (p. 9; our emphasis); "if a theme only has a 50% chance of being expressed by the participant and *noticed* by the researcher" (p. 10; our emphasis). This idea of discovery is deeply problematic to many qualitative scholars, who rather view themes as actively crafted by the researcher, reflecting their interpretative choices, instead of pre-existing the analysis. They are offered to the reader as a compelling and coherent reading of data, rather than (more or less) accurate identification of a decontextualized or pre-existing truth.

If themes are conceptualised as meaningful entities that are constructed from codes that unify disparate data, and capture the essence of some degree of recurrent meaning across a dataset (Braun & Clarke, 2013; DeSantis & Ugarriza, 2000), rather than things in the world that the researcher unearths, the idea of discovery does not work. The end-product from *baking* – e.g., a cake – offers a better metaphor than diamonds. A whole combination of materials (ingredients), processes and skills combine to produce a cake. Before baking, the cake isn't waiting to be 'revealed' – it comes into being through activity and engagement, within set parameters. Fugard and Potts' model, which relies on themes-as-diamonds, requires a series of conceptual positivist-empiricist assumptions (about that nature of reality, about the nature of research, about what our data give us

access to) that don't hold up across much qualitative researching, and which are discarded by many qualitative researchers. Where is the open exploration of new ideas, understandings and constructs that qualitative research excels at?

Fugard and Potts' model, at least, does not reiterate another very common and in our view problematic conceptualisation of a 'theme': the reporting not of themes, but of topics or domains of discussion albeit claiming them *as* themes. Such analysis often effectively provides descriptive summaries of the responses around the topic or focus of the so-called theme – combining a wide range of, potentially radically different, meanings – for example, analysis which identifies themes such as "perceived outcomes for children" or "perceived impact on rehabilitation" (Kinsella & Woodall, 2016) – sometimes clustered around the questions participants have been asked to discuss. If we understand themes as reflecting data extracts all related to a core, shared, meaning, domain-summaries constitute under-developed or poorly-conceptualised themes (Connelly & Peltzer, 2016; Sandelowski & Leeman, 2012). The logic behind Fugard and Potts' model cannot apply to this 'domain' approach: in purposively sampled participants, everyone will likely have some kind of view on things they are asked to discuss, like 'perceived outcomes' or 'perceived impacts,' making the idea of 'theme prevalence' irrelevant.

Themes: Identified or developed?

Clarity around what a theme is, and what it represents, is vital for quality TA. The 'diamond' model of a theme does potentially fit with 'coding reliability' approaches to TA – if themes can be 'captured', 'recognised' and 'noticed' (see Guest, Bunce & Johnson, 2006), they conceptually pre-exist the analytic and interpretive efforts of the researcher. In these approaches, which effectively do qualitative analysis within more or less quantitative logic, themes are developed early in the analytic process, through engagement with data and/or theory. Coding is conceptualised as a process of searching for evidence of identified themes. A structured code-book guides the coding process, which is best undertaken by more than one researcher – high inter-rater reliability offers quality assurance that coding has successfully captured salient themes, which really are there. This consensus coding approach assumes a reality we can agree on, and reveal, through our TA endeavours: the diamonds can be identified, collected, and sorted into piles of like-type.

In contrast, Fugard and Potts' model *doesn't* work for the fully qualitative logic and procedures of 'organic' TA, where coding and theme development processes are organic, exploratory and inherently subjective, involving active, creative and reflexive researcher engagement. The process of analysis – rigorous coding followed by a recursive process of theme *development* – involves the researching 'tussling with' the data to develop an analysis that best fits their research question

(which often will evolve and become refined throughout the analytic process, as Hammersley, 2015, notes). Imagine the wannabe cake baker: standing in their kitchen, surveying the array of ingredients (as well as skills and other factors) at hand, their decision of what *sort* of cake to bake reflects the intersection of many factors. The same goes for analysis in organic TA.

(Mis)conceptualising the logic of samples

Fugard and Potts produce a model where theme relevance is predicated on frequency – and so you determine the frequency of the least-prevalent theme, to determine the sample size you will need. How far along that beach will you need to walk, before you find all six types of diamonds randomly scattered there? It's an inputs-outputs model, implicitly located within the logic of generalisability and replicability. But in organic TA, frequency is not the only (or even primary) determinant for theme development: patterning *across* (some) data items is important, but relevance to addressing the research question is key. What is fundamental is the recognition of TA as a method of identifying patterned meaning *across* a dataset – it's not intended as an idiographic or case study method (although it *has* been used in case study research, Cedervall & Åberg, 2010). And a single instance is not evidence of a theme – which seems to be the logic behind a model based on likelihood of identifying an instance of the least common 'theme'.

Fugard and Potts' noted that our recommendations (Braun & Clarke, 2013) for sample size in TA range from 2 to over 400, and state "it is unclear how to choose a value from the space between" (p. 669). Quite apart from any concerns we might have about the rhetorical decontextualisation of using sampling guidance provided in a student *textbook* to create the impression that established TA practitioners are floundering in the dark when it comes to estimating sample sizes when writing funding and other types of research proposals, there is much to *guide* how one chooses a value from the space between. There is robust and rich discussion around 'sample sizes' in qualitative research (e.g. Coyne, 1987; Malterud, Siersma & Guassora, 2015; Morse, 2000), as well as the logic behind sampling (even the idea of 'sampling' itself is contested), and challenges to the (positivist) implication or claim that larger samples are defacto better. However, the criteria for choosing a sample are not determined by the logic of (post)positivism, and cannot be. Moreover, most would agree that sample size "cannot be predicted by formulae or perceived redundancy" (Malterud et al., 2015, p. 2), and is something qualitative researchers often revisit during data collection, in a live and critically-reflexive, evaluative way. With an organic and flexible approach to TA, and a very wide range of potential project sizes, and data sources, it is expected and appropriate that samples would vary considerably in size. Moreover, if we do not conceptualise themes as diamonds waiting to be discovered, we don't have to rely on the idea of a truth we might miss – and hence do not need to

chase the relatively large sample sizes (for interview-based qualitative research) that Fugard and Potts' model produces. Bigger isn't necessarily better. The bigger the sample, the greater the risk of failing to do justice to the complexity and nuance contained within the data. The student researchers we supervise, as well as published researchers, routinely generate themes and develop complex analyses from smaller samples. This isn't just because themes in organic TA are *constructed* rather than found. It is because a process of fine-grained coding captures diversity and nuance, and provides a foundation for conceptualising possibly *significant* patterns (for research questions) of shared meaning. What we have to have is a clear conceptualisation of what those themes *represent*, and *how* and *why* we treat them as significant. This is more important than some predetermined sample size.

Why should we try to fix what isn't broken?

Qualitative researching is a rich and robust field, with criteria that differ from those in quantitative studies. Attempts to 'fit' qualitative research into a quantitative standards and processes are not just unnecessary – the paradigm itself has done well both at developing 'quality standards' *and* at keeping conversations about things like quality and sample sizes live (e.g. Madill, Jordan & Shirley, 2000; Reicher, 2000; Tracy, 2010) – they are also risky. As *qualitative* researchers, we find Fugard and Potts' model not only essentially meaningless (we *do no*t recommend its use with our 'organic' version of TA… though it *may* offer something of value to 'coding reliability' TA), but also deeply troubling – especially if it becomes a voice of authority that trumps the voices and internal logic of researchers operating within a *qualitative paradigm*.

References

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.

Braun, V. & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. London: Sage.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Byrne, D. (2015). Response to Fugard and Potts: Supporting thinking on sample sizes for thematic analyses: A quantitative tool. *International Journal of Social Research Methodology*, *18*(6), 689-691.

Cedervall, Y., & Åberg, A. C. (2010). Physical activity and implications on well-being in mild Alzheimer's disease: A qualitative case study on two men with dementia and their spouses. *Physiotherapy Theory and Practice, 26*, 226-239.

Clarke, V. & Braun, V. (forthcoming). Applied thematic analysis. In P. Brough & S. Occhipinti (Eds.), *Research methods for applied psychologists: Design, analysis and reporting*. London: Routledge.

Connelly, L. M. & Peltzer, J. N. (2016). Underdeveloped themes in qualitative research: Relationships with interviews and analysis. *Clinical Nurse Specialist*, January/February, 51-57.

Coyne, I. T. (1997). Sampling in qualitative research. Purposeful and theoretical sampling: Merging or clear boundaries? *Journal of Advanced Nursing*, *26*, 623-630.

DeSantis, L. & Ugarriza, D. N. (2000). The concept of theme as used in qualitative nursing research. *Western Journal of Nursing Research*, *22*(3), 351-372.

Emmel, N. (2015). Themes, variables, and the limits to calculating sample size in qualitative research: A response to Fugard and Potts. *International Journal of Social Research Methodology*, *18*(6), 685-686.

Fugard, A. J. B. & Potts, H. W. W. (2015). Supporting thinking on sample sizes for thematic analysis: A quantitative tool. *International Journal of Social Research Methodology*, *18*(6), 669-684.

Guest, G., MacQueen, K. M. & Namey, E. E. (2012). *Applied thematic analysis*. Thousand Oaks, CA: Sage.

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.

Hammersley, M. (2015). Sampling and thematic analysis: A response to Fugard and Potts. *International Journal of Social Research Methodology*, *18*(6), 687-688.

Joffe, H. (2012). Thematic analysis. In D. Harper & A. R. Thompson (Eds), *Qualitative methods in mental health and psychotherapy: A guide for students and practitioners* (pp. 209-223). Chichester: Wiley.

Kinsella, K. & Woodall, J. (2016). 'Pure bonding time': Prisoner and staff perceptions of the impact of taking part in a drama project for imprisoned men and their children. *Prison Service Journal*, Issue 223, 11-17.

Madill, A., Jordan, A., & Shirley, C. (2000). Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies. *British Journal of Psychology, 91*, 1-20.

Malterud, K., Siersma, V. D. & Guassora, A. D. (2015). Sample Size in qualitative interview studies: Guided by information power. *Qualitative Health Research,* DOI: 10.1177/1049732315617444.

McLeod, J. (2011). *Qualitative research in counselling and psychotherapy*, 2nd ed. London: Sage.

Morse, J. M. (2000). Determining sample size. *Qualitative Health Research*, *10*(1), 3-5.

Reicher, S. (2000). Against methodolatry: Some comments on Elliott, Fischer, and Rennie. *British Journal of Clinical Psychology, 39*(1), 1-6.

Sandelowski, M. & Leeman, J. (2012). Writing usable qualitative health research findings. *Qualitative Health Research*, *22*(10), 1404-1413.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837-851.