Lynch Fiona (Orcid ID: 0000-0002-6386-6739)
Vears Danya F (Orcid ID: 0000-0002-6290-545X)

# Alleviating the confusion around content analysis: A comment in response to Wainstein, Elliott & Austin 2023

Fiona Lynch[1,2], Lynn Gillam[3,4], Danya Vears[1,3]

[1]Biomedical Ethics Research Group, Murdoch Children's Research Institute, Parkville, Victoria, Australia

[2]Melbourne Law School, The University of Melbourne, Parkville, Victoria, Australia

[3]Department of Paediatrics, The University of Melbourne, Parkville, Victoria, Australia

[4]Children's Bioethics Centre, The Royal Children's Hospital Melbourne, Parkville, Victoria, Australia

**Corresponding author:** Danya Vears

+61 (3) 8341 6200

danya.vears@mcri.edu.au

**Suggested running head:** Alleviating the confusion around content analysis



This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jgc4.1829

## Introduction

We came across, and applaud, the article from Wainstein, Elliott & Austin recently published in the *Journal of Genetic Counseling* (2023) for providing a much-needed overview of the considerations for the use of qualitative methodologies in genetic counselling research. We agree wholeheartedly that sufficient training and rigour is often lacking in the qualitative research conducted in this field and that this article offers a good overview of many of the most useful qualitative analysis methods. However, as researchers with extensive experience in both qualitative analyses and supervision of those undertaking such analyses, we were concerned about the description of a particular method of analysis: content analysis. We are particularly familiar with this method, having used it frequently in our own research, and also recently published an article clearly describing this methodology and its utility (see Vears & Gillam (2022)). As such, we felt compelled to add to Wainstein and colleagues' description of what content analysis is and outline, from our shared experience, the potential uses of this incredibly valuable qualitative analysis method.

## The confusion around content analysis

In their article, Wainstein and colleagues claim that "a common use of content analysis is for the evaluation of open-ended questions in an otherwise quantitative survey," and go on to describe a form of content analysis which considers "frequencies with which a topic or idea is discussed" (Wainstein et al., 2023, p. 306). What the authors describe is, in fact, only one type of content analysis, known as *quantitative* content analysis, rather than representative of content analysis as a whole.

**What is content analysis?**

The term 'content analysis' broadly describes a family of approaches to analysing textual data, from interpretative and intuitive approaches to strict, systematic processes (Hsieh & Shannon, 2005). Content analysis was first used in the 19th century to analyse text data such as hymns, media articles, advertisements, and political speeches (Elo & Kyngäs, 2008). Early attempts to differentiate content analysis classified different approaches as either quantitative or qualitative (Elo & Kyngäs, 2008; Hsieh & Shannon, 2005). Quantitative content analysis traditionally comes from the field of media research, whereas qualitative content analysis arose in the social sciences (Bengtsson, 2016). Several definitions and descriptions of content analysis (both quantitative and qualitative in nature) have subsequently been presented over the time of its use (Bengtsson, 2016) (Table 1). The methodological literature on content analysis employs a wide variety of terminology and poorly articulated descriptions of how to conduct such methods; in particular, a number of terms are used to refer to qualitative content analysis across the literature, including "qualitative content analysis" (Cavanagh, 1997; Graneheim & Lundman, 2004; Krippendorf, 2004; Polit & Beck, 2004; Vaismoradi et al., 2013), "interpretive content analysis" (Ahuvia, 2001), "inductive content analysis" (Elo & Kyngäs, 2008; Vears & Gillam, 2022), or even simply "content analysis" alone (Downe-Wamboldt, 1992; Weber, 1990). To add to the confusion, in their widely cited paper, Hsieh and Shannon introduce new terminology which does not reference directly the qualitative/quantitative distinction. They describe three distinct types of content analysis: conventional, directed, and summative (Hsieh & Shannon, 2005), giving explanations that mix qualitative and quantitative analysis techniques.

The reason for the confusion about content analysis, we suggest, is that content actually varies across two dimensions, not just one, but the naming conventions do not capture both. The two dimensions are: (1) the source of the content codes that are identified for analysis (inductively or deductively derived), and (2) the type of analysis that is used on these codes (quantitative or qualitative). The names given to various types of content analysis do not clearly distinguish or refer to both dimensions. Here we briefly discuss the different types of content analysis in an attempt to show the range of different approaches, beyond the account given by Wainstein and colleagues, and clarify some of the terminology used.

*Quantitative content analysis*

Content analysis *can* be used primarily as a quantitative method, with text data coded into categories and described using statistics (such as frequencies) (Hsieh & Shannon, 2005). This is the type of content analysis that Wainstein and colleagues seem to have in mind in their description of content analysis. This type of content analysis involves searching the text for, and counting of, recurring words or concepts (Bengtsson, 2016; Patton, 2001); it is more standardised and formulaic than qualitative content analysis (Elo & Kyngäs, 2008). As Wainstein and colleagues say, it is commonly used for analysing free text responses in surveys and questionnaires because it marries well with quantitative analysis of the closed-ended (forced choice) questions.

Hsieh and Shannon's description of "summative content analysis" (Hsieh & Shannon, 2005) aligns most closely with our understanding of "quantitative content analysis" (Bengtsson, 2016). Some might disagree that this summative method is quantitative, on the grounds that Hsieh and Shannon describe it as moving beyond simple frequency statistics of

particular codes or categories; they suggest it includes some interpretation of underlying content (Hsieh & Shannon, 2005). We suggest that this is not enough to make this a qualitative form of content analysis, because all research (including quantitative research) involves some level of interpretation. In support of this, Elo and Kyngäs highlight that *qualitative* content analysis is far more than "a simplistic description of the data or a counting game" (2008, p. 108).

*Qualitative content analysis*

In qualitative content analysis, the aim is to describe (rather than quantify) the phenomenon in question, using a rigorous, systematic approach in which data are reduced to concepts (or categories) (Elo et al., 2014; Elo & Kyngäs, 2008). While qualitative content analysis aims to provide "a condensed and broad description of the phenomenon" (Elo & Kyngäs, 2008, p. 108), this method can also result in the creation of a model or conceptual system describing the phenomenon (Elo et al., 2014). The analysis aims to produce an understanding of the meaning of the content of the data: a rich answer to the research question, that is also practically relevant to the real-world context of the research (Vears & Gillam, 2022). There is a definite lack of published procedures for conducting qualitative content analysis (Elo et al., 2014; Elo & Kyngäs, 2008), but in broad terms, it involves identifying pieces of text that have common content, grouping them into content categories, then focusing on understanding the meaning of content by an iterative process of comparing and asking questions about the content. There is no numerical or statistical analysis. This is the type of content analysis that Wainstein and colleagues have missed, even though it is widely used.

This process is very similar to thematic analysis, so additional confusion arises when trying to elucidate the difference between forms of qualitative content analysis and the apparently similar methods (Vears & Gillam, 2022). We (DV and LG) recently described our use of qualitative inductive content analysis in recognition that for the uninitiated, the methodological literature on content analysis can be confusing and even contradictory (Vears & Gillam, 2022).

Cutting across the division between qualitative and quantitative forms of analysis is the distinction between deductive and inductive approaches to coding (labelling of pieces of text). Understanding this difference in the ways in which content codes are identified *before* they are analysed helps to further appreciate the diversity of forms of content analysis. It is not clear from Wainstein and colleagues' brief description whether they have in mind an inductive or deductive approach to identifying content categories when they describe counting frequencies as the method of analysing the content: either would be possible, as we explain below, but for methodological rigour this needs to specified.

*Deductive approach to coding*

In a deductive approach to content analysis, researchers analyse the textual data they have collected (e.g., free text survey responses, media articles, or even long text such as an interview transcript) using a predetermined list of codes, often developed based on the existing literature. The code list is static and is not refined throughout the coding process; it is simply applied to the data set, and all data is sorted into the categories it specifies. Hsieh and Shannon's description of "directed content analysis" (Hsieh & Shannon, 2005) aligns most closely with our, and others', understanding of "deductive content analysis" (Elo & Kyngäs, 2008). This method is most useful when there is existing theory or prior research

about the phenomenon of interest (allowing content codes to be set in advance of looking at the text data), and the aim of the research is to validate or extend such theories (Hsieh & Shannon, 2005), or to compare findings across different time periods (Elo & Kyngäs, 2008).

A deductive approach to coding typically goes with quantitative analysis of the codes, so this may well be the type of content analysis that Wainstein and colleagues are referring to.

*Inductive approach to coding*

In contrast to a deductive approach to coding for content analysis, an inductive approach involves developing codes from the data, rather than having a predetermined list. This is commonly described as the codes "emerging" or "arising" from the data, through an iterative process of close reading and comparison. These codes can be refined as analysis progresses (Elo & Kyngäs, 2008; Vears & Gillam, 2022). Inductive content analysis is suited to situations where little is known about the phenomenon of interest (Elo & Kyngäs, 2008; Hsieh & Shannon, 2005). Hsieh and Shannon's "conventional content analysis" (Hsieh & Shannon, 2005) aligns closely with inductive content analysis.

The inductive approach to coding content is compatible with both qualitative and quantitative methods of analysis. For inductive quantitative content analysis, the codes are developed inductively, and once the data is coded, frequencies are counted and the data analysed quantitatively. This approach is often seen with content analysis of free text in questionnaires and surveys. So, rather than deductive content analysis, this could be what Wainstein and colleagues are describing. Their description does not provide detail about the approach to generating codes, only about the kind of analysis that follows on the coded content.

The type of content analysis that Wainstein and colleagues are clearly not referring to, however, is inductive qualitative content analysis. Qualitative content analysis using an inductive approach to coding does not involve counting instances of content (codes) to determine frequencies or statistical correlation (Vears & Gillam, 2022). It is about holistic description and meaning, so that even if a piece of content representing a particular thought or idea is coded only once in a whole data set, it may still be important to report such a finding (Vears & Gillam, 2022) and account for it in the interpretation of the data.

Qualitative content analysis which uses an inductive approach to coding is a major form of content analysis, recognised in the literature and seen in many published studies.

*A note on thematic analysis*

Thematic analysis is not a form of content analysis. However, Hsieh and Shannon highlight that inductive qualitative ("conventional") content analysis may be easily confused with other qualitative methodologies such as grounded theory or phenomenology (Hsieh & Shannon, 2005) which employ forms of thematic analysis. Indeed, there are many similarities between inductive qualitative content analysis and thematic analysis (Braun & Clarke, 2006; Vears & Gillam, 2022). Inductive content analysis, much like thematic analysis, is a non-linear, reflexive process (Elo & Kyngäs, 2008). However, inductive content analysis differs from thematic analysis in its aim to stay closer to the phenomenon of interest (Vears & Gillam, 2022). Themes and categories are similar, but represent different levels of detail of analysis; a category provides a description of content, whereas a theme delves deeper into meaning of the data (Vaismoradi et al., 2013). Inductive content analysis therefore provides a description of a phenomenon, whereas thematic analysis provides a detailed, nuanced account of a phenomenon (Vaismoradi et al., 2013).

## Conclusion

Supporting and maintaining methodological rigour is essential to ensure trustworthy qualitative research. In a field where this is sometimes lacking, Wainstein and colleagues provide much-needed guidance about the variety of qualitative research methods available to genetic counselling researchers as a vital resource to support good qualitative research. We aim to add to this by sharing our expertise in the field of content analysis. We have described – here and elsewhere (Vears & Gillam, 2022) – a useful form of qualitative content analysis that requires training, skill, and "an enormous amount of work" (Elo & Kyngäs, 2008, p. 113), but repays this investment with robust and insightful research outcomes. We hope that highlighting this type of content analysis, clarifying some sources of confusion, and directing readers to thorough and practical descriptions of the method will support rigorous qualitative research in the *Journal* and beyond.

## Author contributions

FL and DV were responsible for conception and initial drafting of this manuscript. LG contributed to revising the manuscript. All authors gave final approval of this version to be published and agree to be accountable for all aspects of the work.

## Conflict of interest

FL, LG, and DV declare that they have no conflict of interest.

## References

Ahuvia, A. (2001). Traditional, interpretive, and reception based content analyses: Improving the ability of content analysis to address issues of pragmatic and rheoretical concern.

*Social Indicators Research*, *54*(2), 139-172. https://doi.org/10.1023/A:1011087813505

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus open*, *2*, 8-14.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Cavanagh, S. (1997). Content analysis: concepts, methods and applications. *Nurse Researcher*, *4*(3), 5-16.

Downe-Wamboldt, B. (1992). Content analysis: Method, applications, and issues. *Health Care for Women International*, *13*(3), 313-321. https://doi.org/10.1080/07399339209516006

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE Open*, *4*(1), 2158244014522633. https://doi.org/10.1177/2158244014522633

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107-115.

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, *24*(2), 105-112. https://doi.org/10.1016/j.nedt.2003.10.001

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288.

Krippendorf, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage.

Patton, M. Q. (2001). *Qualitative evaluation and research methods* (3rd ed.). Sage.

Polit, D. F., & Beck, C. T. (2004). *Nursing research: Principles and methods*. Lippincott Williams & Wilkins.

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, *15*(3), 398-405.

Vears, D. F., & Gillam, L. (2022). Inductive content analysis: A guide for beginning qualitative researchers. *Focus on Health Professional Education*, *23*(1), 111-127. https://search.informit.org/doi/10.3316/informit.455663644555599

Wainstein, T., Elliott, A. M., & Austin, J. C. (2023). Considerations for the use of qualitative methodologies in genetic counseling research. *Journal of Genetic Counseling*, *32*(2), 300-314. https://doi.org/https://doi.org/10.1002/jgc4.1644

Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Sage.

**Tables**

Table 1. Terminology of different forms of content analysis from the literature

**Table 1. Terminology of different forms of content analysis from the literature**

| Content analysis | | |
|---|---|---|
| **Quantitative content analysis**<br>Known as:<br>• *Quantitative content analysis* (Bengtsson, 2016; Elo et al., 2014)<br>• *Summative content analysis*[1] (Hsieh & Shannon, 2005)<br>• *Content analysis* (Patton, 2001) | **Qualitative content analysis** | |
| | **Inductive content analysis**<br>Known as:<br>• *Inductive content analysis* (Elo & Kyngäs, 2008; Vears & Gillam, 2022)<br>• *Conventional content analysis* (Hsieh & Shannon, 2005)<br>• *Qualitative content analysis* (Bengtsson, 2016; Cavanagh, 1997; Elo et al., 2014; Graneheim & Lundman, 2004; Krippendorf, 2004; Polit & Beck, 2004; Vaismoradi et al., 2013)<br>• *Interpretive content analysis* (Ahuvia, 2001)<br>• *Content analysis* (Downe-Wamboldt, 1992; Patton, 2001; Weber, 1990) | **Deductive content analysis**<br>Known as:<br>• *Deductive content analysis* (Elo & Kyngäs, 2008)<br>• *Directed content analysis* (Hsieh & Shannon, 2005) |

[1]Hsieh and Shannon describe summative content analysis as involving both quantification of words or other content (a quantitative approach), as well as some description of the contextual use of such words or other content (a qualitative approach).