Understanding Thematic Analysis and its Pitfall

Mostafa Javadi¹, Koroush Zarea²*

¹Research Center for Nursing and Midwifery Care in Family Health, School of Nursing and Midwifery, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
²Chronic Disease Care Research Center, Nursing & Midwifery School, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

*Correspondence should be addressed to Koroush Zarea, Chronic Disease Care Research Center, Nursing & Midwifery School, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran; Tell: +989166154943; Fax: +983538249705; Email: Korosh1685@hotmail.com.

ABSTRACT

The best meaning for the term “research” is in the term itself. Research methods are under the impact of philosophy and ideology. As the type of view and understanding is various in different individuals, different research methods have been developed and used so far. Qualitative research methods are one type of research and emphasize facts and relative knowledge and the knowledge is formed in the framework of time. Thematic analysis is one of the types of qualitative research methods which has become applicable in different fields. This study explores different types of thematic analysis and phases of doing thematic analysis. Then the issues and advantages of thematic analysis are discussed.

Key words: Thematic Analysis, Qualitative Research, Theme, Inductive Analysis.

Copyright © 2016 Mostafa Javadi et al. This is an open access paper distributed under the Creative Commons Attribution License. Journal of Client Care is published by Lexis Publisher; Journal p-ISSN xxxx; Journal e-ISSN 2476-6682.

1. INTRODUCTION

Thematic analysis (TA) is one of the most common forms of analysis in qualitative research (1). The history of use of TA is unclear. But it was used interchangeably with content analysis (Christ 1970), phenomenology (Benner 1985), and ethnography (Aronson 1994) (2). TA is an approach for extraction of meanings and concepts from data and includes pinpointing, examining, and recording patterns or themes (3-5). Data can be in any form including: transcription of an interview, notes in the field, political documents, pictures, and videos (6, 7). TA is a method for detection, analysis and reporting the themes in data (4, 5). It is the minimum organization and description of a set of data that is widely used in qualitative data analysis (4, 8). Rubin and Rubin suggest that this analysis is very exciting as you discover themes and concepts from the interviews you have had (9, 10). This state regarding themes has been referred to as theme emerging by Ely (1997) (11). One of the advantages of content analysis is its flexibility (5). A good TA can highly help in both reflecting and clarifying the reality. TA is considered as a basic method for qualitative analysis (4, 10). This method is widely used in Interpretative Phenomenological Analysis (IPA) and even in other methods for qualitative research such as grounded theory (1). Of course, it is better to use discourse analysis and narrative analysis in grounded theory (12). Two major concepts regarding data should be clarified before discussing analysis method. These two concepts are data corpus and data set. Data corpus refers to all of the collected data for a special research subject and data set refers to all the data employed for a special analysis (5, 12).

2. What is Theme?

In fact theme mentions some important points regarding the research data and shows a pattern or meaning related to data sets (4). Theme is a kind of agreement that, in comparison to the main text from which the theme is extracted, is more concise, accurate, simpler and shorter. Most themes are expressed in a more explicit and tacit way rather than a clear and an explicit way (4, 13). A theme must be differentiated from a code. Several authors recommend that researchers "code for themes" This can be illusory because the TA is not only coding. Theme is the outcome of coding. The code is the label referred to special
Some variables limited the way the method is applied in a range of theoretical and epistemological paradigms in psychology. Through this theoretical freedom, TA is a flexible research tool which is also useful and can do data calculations in a potentially rich way with details and even with complexities. Therefore, flexibility is one of the advantages of TA and of course result in some criticisms of this type of analysis (4, 12).

3. How are Themes Analyzed and Reported?

3.1. Rich or detailed description?

Decision-making on whether to make the reported themes in the form of rich description of the data set or detailed account of one particular aspect is upon the researcher. Depth and complexity are impacted if the researcher wants to create a paper with a limited number of words or around specific topics on which few number of papers have been published, or if the research question is broad or if the codes and themes have to be an accurate reflection of the whole content of data set. In contrast, sometimes the researcher may decide to provide a report of a special theme or several themes with good details, for example, provide a part of the results of a thesis (5, 21).

3.2. Inductive or Theoretical Thematic Analysis?

3.2.1. Inductive thematic analysis

An inductive analysis means that the recognized themes are strongly made related to the data (4, 22). In this method in which the data are collected for a specific research subject, for example with focus group method, the recognized themes may have little relationship with the questions asked from the participants (4, 5, 12).

3.2.2. Theoretical TA

Theoretical TA is mostly done based on the theory or the analysis liked by the researcher (22). Thus, it is explicitly extracted by the researcher in the form of that specific theory. This method of analysis presents a description of the data with less richness and the details are presented based on the initial theory (4, 5). An issue here would be whether the researcher studies the findings of similar studies or not? There is no definite answer to this question yet as other qualitative studies are also disputed, if the results of similar studies are considered before the analysis there is a risk of the researcher’s scope being limited and the researcher may ignore some the critical aspects of the data or pay too much attention to some specific parts of the data (18, 22). On the other hand, a review of some similar studies can result in the researcher being sensitised to some intangible aspects. A review of texts before the analysis is recommended in theoretical method (15, 16, 18, 22).

[Data Analysis Methods Can Roughly be divided into Two Groups:

In the first group are those methods which exist or epistemologically placed like Conventional Analysis (CA) and Interpretative Phenomenological Analysis (IPA) and some variables limited the way the method is applied in a framework. In these methods, the analysis is done based on

1. Overall nature: experience
2. Structure: nature or basis of the experience
3. Performance: capturing and uniforming the nature or basis of the experience into a meaningful whole
4. Shape: stable or multitude of multiple experience instances
5. State: recurrent of the experience (16)

Based on a study by DeSantis and Ugarriza (2000) conducted on qualitative papers between 1979 and 1998, 40% of the papers had used the word “theme” in their studies. According to them no specific definition of them was found in the aforementioned papers. However, several definitions of the word “theme” which exist in different sources are as follows:

- Brink, Wood (1997): The term “theme” is used for describing the fact that the data are grouped around a main issue (17)
- Speziale, Streubert (2011): theme is a structural meaningful unit of data which is necessary for providing qualitative findings (18).
- Polit, Hungler (1999): a recurrent and systematic occurrence which appears in qualitative data analysis (19).

Data Analysis Methods Can Roughly be divided into Two Groups:

In the first group are those methods which exist or epistemologically placed like Conventional Analysis (CA) and Interpretative Phenomenological Analysis (IPA) and some variables limited the way the method is applied in a framework. In these methods, the analysis is done based on
3.3. Semantic or Latent?

3.3.1. Semantic themes (explicit)

In semantic approach the themes are detected at “the surface or semantic appearance” and the researcher is not after something beyond what the participant has said or what is written in the text. This is simplest and the most evident type of theme. In this method the data are explained and it is simply for showing patterns that exist in the data and are organized in the forms of content, summarized or interpreted meanings. Here efforts are made to theorize the importance of patterns and their wider meanings (4, 16, 18).

3.3.2. Latent themes (interpretative)

In this level of analysis we go beyond what is obtained in the semantic method. This level is the beginning of efforts for detecting and testing beliefs, presumptions and conceptualization for forming semantic content of the data and with a level of the researcher’s interpretation (4, 16, 18). In fact, it can be said that the semantic approach is after the literal meaning while the latent or analytical approach requires going from description in which the data are just organized to reveal some patterns in semantic content and make concise, to interpretation in which efforts are made to create a theory based on the importance of the patterns and a wider framework of meanings and connotations (5).

4. Sample Size

There is no unique guidance on what sample size is needed for a TA (23, 24). It's depending on the type of data collection, size of the project, and how are themes analyzed and reported (25). Sampling is continued until no further codes are found (data saturation) (25, 26).

5. Phases of Thematic Analysis (Braun and Clarke, 2006)

TA phases have many similarities with other phases of qualitative studies and they are not specific to thematic analysis (5, 22). The analysis is done in a recursive process and not a linear way. The codes are extracted and these codes are transformed into themes. The researcher frequently refers to the extracted codes and the entire data set and validates them (2, 5). Also, the researcher does the same between the analysis phases which are mentioned below. Writing is an indispensable part of the analysis and thus the researcher should not do it the last phases. The researcher should continuously take notes from his/her analysis and write the ideas that come to his/her mind regarding the codes in the first step (18, 22). Flexibility as a principle should be considered in the analysis and what are recommended as analysis do not that are rules (5).

5.1. Familiarizing Yourself with Your Data

First, each word in the content of the interviews and speeches should be transcribed with correct spelling and this phase is one of the most significant phases in interpretative qualitative studies (9). It should be noted that even the way a comma is placed can changed the meaning perceived from a content. This is very difficult and time consuming but highly valuable and familiarization with the data occurs during it (4, 27). If your text is transcribed for you by someone else, you should read the text completely. In order to find out the content depth, the immersion of the researcher in the data is necessary. Researchers recommend active repeated reading so that you become familiar with all aspects of your data. It is necessary to read the whole set of data, before coding, in order to obtaining an overall understanding. The preliminary patterns will be formed by reading the ideas and possible similarities (9, 14). You should remember that all parts of the data are important and if you study some parts selectively, you may ignore other parts. In fact, it is through examining the data that specific patterns and meanings in the writings gradually emerge. Then the verbal data should be turned into a manuscript that has minimum ambiguity grammatically and you know what you mean whenever you refer to it (5).

5.2. Generating Initial Codes

Create a preliminary list of ideas related to the data. Organize your data into significance groups and give the initial codes to the data (28). The codes can be explicit or implicit meanings (semantic or latent) that are related to the most basic part of the data or raw information and can be evaluated in a meaningful way with regard to a phenomenon. The codes can be formed depending on the analysis type i.e. inductive or theoretical, or depending on the specific type of the question that is formed in your mind (4, 14, 28).

Begin the process systematically from the overall set of data. Pay complete and equal attention to all data and identify the important aspects in the data that may or may not be repeated in the data. One thing that should be considered in this phase is that the code will be as a high number of potential themes or patterns (4, 14). Give a related code to the data itself and pay less attention to its surroundings. Remember that you can give highly different codes to an extracted content and it is possible that you do not give code to a content at all, give one code or many codes. You can use different methods for writing codes. For example, some softwares have been designed for this purpose. Use margin notes and with different colors or cards. It is better to have access to raw data whenever you refer to the text for coding later. For example, with copying the previous content before note taking in paper method or hiding cases in the software. This helps you to consider equal importance for all parts of the text (5).

5.3. Searching for Themes

The themes are sought from the codes whenever the initial codes are formed. For this purpose, you should know the codes. You have a long list of different codes. You can
gradually bring similar codes under a set. You can give a name to each set and write a concise explanation for that name separately (5, 14, 28). Then try to organize the code sets meaningfully. Some codes form theme, some others are subthemes and some are codes that do not belong to a theme yet and they are necessary to be written temporarily to later determining the themes they belong to; or it may be necessary to extract a theme from them. Consider how different codes can be combined to form an overarching theme. At the same time, you should think about the relationship between different codes, themes and theme levels. Using the designs in the form of concept map on paper, software and schematic diagrams is highly helpful (5, 28). Now your preliminary thematic map is emerging! But it is not definite and final and is developed in the analysis process. Look at the thematic map of the study by Ghiyasvandian (2014) (29) (Figure 1, Figure 2).

![Thematic map of the study by Ghiyasvandian (2014)](image1)

And the developed themes from the same study:

![Final themes in the aforementioned study](image2)

5.4. Reviewing Themes

Two basic principles regarding the characteristics of theme should be pointed out here: internal homogeneity and external heterogeneity. This means that the data inside the themes should be meaningfully related to each other and the themes should be explicitly and expelled differentiable. When you refer to your initial themes you will see that some of them are not really theme or do not have enough supportive data. Some themes may be merged with other themes due to overlap. Or they may create new themes in combination with other themes due to having homogeneity or common roots. Even subthemes may be separated from
each other and placed under a new theme. It is not unexpected for new themes that did not exist before to emerge. This shows a part of flexibility (5). This phase is done in two levels: in the first level, you should go back to the extracted codes of each theme and see whether these codes form a consistent pattern. If they did, you go to the second phase in which a process similar to the one in the previous phase is done but we consider the validity of themes regarding the whole data in the entire data set. In this state our thematic map should be an accurate document for our data set as a whole. At the end of this phase you should have a good idea on what differentiates the themes, how they are matched and the whole story they tell about the data. Otherwise you need more refining and reviewing (5). As pointed out before, analysis is done in a cyclic process and there may be a need to refer to previous phases at each phase (4, 9, 14); of course, to the extent that you are not lost in the never-ending cycle of analysis! Stop the refinements whenever you reached the conclusion that refinements do not add an important thing. You are transferred to the next phase when you reached a convincing thematic map.

5.5. Defining and Naming Theme
You define theme in this phase and review and refine them while as you are analyzing. You reach the theme essence by defining and refining. You reach what the theme says and what it is about and what aspects of the data are covered by the theme. Here, in addition to interpretation of the data content, you should determine the things that are interesting regarding the data and the way they become important. In addition, during refinement you should determine whether each theme has subtheme (s) or not. Subthemes are in fact themes inside themes and a set of subthemes make a complex and big theme and show the meaning hierarchy in the data. You should be able to define what themes are and what they are not at the end of each phase. This means that you should summarize the scope and contents of each them in about two sentences. Naming theme is done after defining. The names should be clear, accurate and evident and transfer to the reader the thing the theme is about quickly (2, 5).

5.6. Producing the Report
The sixth phase begins when you have a good set of themes and you do the final analysis by writing and reporting them. It is important to note that the story of themes is expressed accurately, consistently, logically, without repetition and with attraction through or from within the themes. The provided essence should be identifiable easily (5). Morse (1995) provide 5 steps for thematic analysis:

a) Recognition and listing cognitive data (parts of patterns) or nursing observations and experiences
b) Combination of related data and patterns into meaningful units based on having relationship with bigger units that are known as theme.

c) Recognizing subthemes and subpatterns and determining the way they become related to patterns and themes
d) Synthesis of several small themes for obtaining a general, comprehensive and broad view
e) Formulation of phrases of themes or patterns for more retesting or reconfirmation of nursing phenomena (16).

6. Thematic Analysis Issues
TA is a clear, uncomplicated and straightforward qualitative study which does not need some theoretical details and technical knowledge such as discourse analysis or conversational analysis. Therefore, it can be said that doing a good TA in data is a simple, enjoying and flexible work. However, like other qualitative methods, some potential pitfalls may result in weak analysis and they should be dealt with.

a) One should be unbiased in doing TA as a research work. Unprofessional and simplistic view sometimes destroys the value and validity of TA in a way that the result becomes desired and positive and thus leads to serious damage (5).

b) TA is not just collection of a series of similar or organized data together with a little a low level of analytical narrative that simply paraphrases the data content and turns them into their initial phrases. The essences in TA indicates the analytical points that are processed by the researcher about the data and should be used for making sense and supporting the analysis which is beyond a specific content and to tell the reader what the data content means or may mean and not for providing a summary of data in the form of a series of extracted words (5, 12).

c) Do the researchers reach what the data really tell in their explored subject? (30). As in any scientific study the conclusions and judgments should be based on the data from the study, the TA is not an exception and should refrain from personal inferences and specific prejudgments by the researcher on the research content and should pay attention to explicit or latent content of the text or message as it is. What has been mentioned in the text or message should be analyzed and investigation of the content based on the existing information related to the aim of the author should be avoided. For this reason, sometimes a lack of proportionality between the data and analytical claims created for it is seen. In such cases there is no coherence and consistency between the claims and the data; and in the worst scenario, the data extract requires another analysis or is even in contrast with the claims mentioned (5). The researcher should find out whether his/her interpretations and analytical points are compatible with the data extract or not.

d) Sometimes a part of the questions for data collection or interview guidance is introduced as theme. It is obvious that in such cases the researcher has not done any analytical work for identifying themes in the data sets and the themes are made of the researcher’s
assumptions and not data analysis. Also, the interview questions may be impacted by the researcher’s presumptions and thus, the researcher presents his/her presumptions instead of the data tell what they mean. The involvement of direct view of the researchers that may be originated in the mental background should be avoided in every phase of the analysis (5).

c) The analysis is weak or unconvincing. This is revealed when the themes have high levels of overlap or the themes internally lack coherence and consistency; in a good TA all aspects of a theme should be focused around a main idea or concept. This issue happens when all the aspects of the data are not analyzed or there is a defect in providing a rich description or interpretation of an aspect or several aspects of the data. Sometimes it is seen that the researcher’s attention is diverted from the path that results in a lot of emphasis on a part or viewpoint and an overall understanding of the text is ignored. Unconvincing or weak analysis can be due to failure to provide enough examples of the data, for example one or two extracts for each theme. Material analysis (analysis of a subject) is an assessed, self-conscious and artistic creation that is made coherent by the researcher for convincing the reader in presenting a discussion as justifiable. Of course the researcher should avoid anecdotalism which means that one or several limited samples of the phenomenon-when they are idiosyncratic- are manifested in a theme. This does not mean that a few examples cannot be interesting or revealing but it is important not to depict them as overarching theme (5).

e) Mismatch between theory and analytical speech or between research and TA form that is used. It is necessary that the data interpretations match theoretical framework in a performing a good thematic analysis. Even if an analysis is good and interesting but it does not explain what its theoretical presumption or purpose is, it will lack crucial information and thus it is defective in one aspect.

f) Gibson (2006) points out three issues for thematic analysis, the main part of which is theoretical issue: it is the interpretativism which is in fact the interpretation of others’ actions through our understanding. The second issue is language which is one of the ways from which we make sense of our experiences from the world. Wittenstein has a famous expression: the limitation of my language is the imitation of my world. This points out the shortcomings of language and that each individual has its own world and attention should be paid to context for entering it. Some questions can be asked in this section which indicate the issues in this type of analysis: do the researcher’s activities make sense of others’ actions? The second issue that is pointed out by him is quantification in TA for creation of a set of patterns and the researcher should not just pay attention to repetition of specific terms (31).

It is important to note this disadvantages occur due to inappropriate research questions or poor analyses and not TA itself (Hollardson, 2009; Hayes, 2000).

7. Advantages of Thematic Analysis

Though there are some criticisms of thematic analysis, this method is simple and simpler than other qualitative research methods. High level of flexibility and simplicity and tangibility of analysis phase have made less-experienced researchers in qualitative studies not hopeless and have attracted them towards this method. On the other hand, the results of this method are understandable for the public who have a low education level. Overall, considering the advantages and limitations of this method, the researcher decides whether to employ this method or not.

8. CONCLUSION

Thematic analysis is the most common and the simplest form of analysis in qualitative research. TA is an approach for extraction of meanings and concepts from data and includes pinpointing, examining, and recording patterns or themes. TA not only provides a flexible method of data analysis in qualitative research, it establishes the more systematic and explicit form of it without threatening depth of analysis. Overall, considering the advantages and limitations of this method, the researcher decides whether to employ this method or not.

FUNDING/SUPPORT

Not mentioned any funding/ support by authors.

ACKNOWLEDGMENT

Not mentioned any acknowledgment by authors.

AUTHORS CONTRIBUTION

Dr. Javadi prepared draft of manuscript and supervised revision. Dr. Zarea designed the study and supervised data collection.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this paper.

REFERENCES

1. Guest GM, KM and Namey, EE Applied Thematic Analysis. Thousand Oaks California: Sage; 2012.
2. Braun V, Clarke V, Terry G. Thematic analysis. Qualitative research in clinical and health psychology. 2015;5:113. [Crossref]
3. Tjandra NC, Osei C, Ensrin J, Omar M. Exploring the influence of country-of-origin information to Generation Y’s perception towards international fashion brands. 2013. 4. Boyatzis RE. Transforming qualitative information: Thematic analysis and code development; Sage; 1998.
5. Braun V. Clarke V. Using thematic analysis in psychology. Qualitative research in psychology, 2006;3(2):77-100. [Crossref]
6. Joffe H, Yardley L. 4. Content and thematic analysis. Research methods for clinical and health psychology California: Sage. 2004;56-68. [Crossref]
7. Guest G, MacQueen KM, Namey EE. Applied thematic analysis; Sage; 2011.
8. Mitchell SA, Fisher CA, Hastings CE, Silverman LB, Wallen GR. A thematic analysis of theoretical models for translational science in nursing: Mapping the field. Nursing outlook. 2010;58(6):287-300. [Crossref]
9. Rubin HJ, Rubin IS. Qualitative interviewing: The art of hearing data: Sage; 2011.
10. Rubin H, Rubin I. S. (1995). Qualitative interviewing: The art of hearing data. Thousand Oaks, CA: Sage.
11. Ely M. On writing qualitative research: Living by words. Psychology Press; 1997.
12. Vassmorod M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nursing & health sciences. 2013;15(3):395-405.
13. Wilkinson S. Women with breast cancer talking causes: Comparing content, biographical and discursive analyses. Feminism & Psychology, 2000;10(4):431-60. [Crossref]
14. Saldaña J. The coding manual for qualitative researchers: Sage; 2015.
15. DeLamont L, Ugurcay DN. The concept of theme as used in qualitative nursing research. Western Journal of Nursing Research. 2000;22(3):351-72. [Crossref]
16. Morse JM. Qualitative research methods for health professionals. 1995.
17. Brink P.J, Wood MJ. Advanced design in nursing research: Sage Publications; 1997.
18. Speziale HS, Stribert HJ, Carpenter DR. Qualitative research in nursing: Advancing the humanistic imperative: Lippincott Williams & Wilkines; 2011.
19. Politi D, Hungler B. Essentials of nursing research: Methods, appraisal, and utilization. 1994.
20. Taylor SJ, Bogdian R, DeVault M. Introduction to qualitative research methods: A guidebook and resource: John Wiley & Sons; 2015.
21. Wittel J, Entner T. [Botulinum toxin treatment of hip adductor spasticity in multiple sclerosis]. Wiener klinische Wochenschrift. 2001;113 Suppl 4:20-4. PubMed PMID: 15906048. Epub 2004/10/28. Botulinum Toxin Typ A in der Behandlung der Adduktorenpastizität. ger.
22. Corbin J, Strauss A. Basics of qualitative research: Techniques and procedures for developing grounded theory: Sage publications; 2014.
23. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field methods. 2006;18(1):59-82. [Crossref]
24. Emmel N. Sampling and choosing cases in qualitative research: A realist approach: Sage; 2013.
25. Fugard AJ, Potts HW. Supporting thinking on sample sizes for thematic analyses: a quantitative tool. International Journal of Social Research Methodology. 2015;18(6):669-84. [Crossref]
26. Baker SE, Edwards R, Dodge M. How many qualitative interviews is enough?: Expert voices and early career reflections on sampling and cases in qualitative research. 2012.
27. Nikbakht-Nasrabadi A, Mohammadpour A, Abbasi M, Javadi M. Iranian Nursing Students Perspective of their Rights in Clinical Evaluation: A Thematic Analysis Study. Life Science Journal. 2012;9(4).
28. Clarke V, Braun V. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. The psychologist. 2013;26(2):120-3.
29. Ghiyasvandian S, Nikbakht-Nasrabadi A, Mohammadpour A, Abbasi M, Javadi M. Iranian nursing students’ perspectives of educational equity. Acta Medica Iranica. 2014;52(12).
30. Vaezi AA, Tavangar H, Javadi M. Factors Influencing the Process of Continuous Nursing Education Based on the Experiences of Nurses: A Qualitative Study. Strides in Development of Medical Education. 2015;11(4):508-22.
31. Gibson W. Thematic analysis. Retrieved August. 2006;2:2008.
