| Postprint Version | 1.0 |
|---|---|
| Journal website | http://dx.doi.org/10.1016/j.pec.2011.01.005 |
| Pubmed link | http://www.ncbi.nlm.nih.gov/pubmed?term=21292424 |
| DOI | 10.1016/j.pec.2011.01.005 |

# A standardized approach to qualitative content analysis of focus group discussions from different countries

FRANCESCA MORETTI[A], LIESBETH VAN VLIET[B], JOZIEN BENSING[B], GIUSEPPE DELEDDA[A], MARIANGELA MAZZI[A], MICHELA RIMONDINI[A], CHRISTA ZIMMERMANN[A] AND IAN FLETCHER[C]

a Department of Public Health and Community Medicine, University of Verona, Italy
b NIVEL, Netherlands Institute for Health Services Research, Utrecht, The Netherlands
c Division of Clinical Psychology, University of Liverpool, Liverpool, UK

**ABSTRACT**

Objective: To describe the methodological procedures of a multi-centre focus group research for obtaining content categories also suitable for categorical statistical analyses.

Methods: Inductive content analyses were performed on a subsample of 27 focus groups conducted in three different countries, the Netherlands (Utrecht), the UK (Liverpool) and Italy (Verona). The analyses of the subsample of focus group discussions were performed in five steps: (1) independent development of content categories in each of the participating centres, (2) obtaining consensus categories, (3) creation of a manual with coding rules and defining criteria for categories and subcategories, (4) assessment of inter-rater reliability to identify unreliable categories to be revised, and (5) repetition of inter-rater reliability assessment.

Results: The resulting coding system considers five areas: non verbal communication, process oriented expression, task oriented or problem focused expressions, affective or emotional expressions, and physician's personal characteristics. It contains 12 categories of acceptable inter-rater reliability and 41 subcategories.

Conclusion: The coding procedures show how focus group data, obtained in an international multi-centre study can be analysed in a systematic way combining scientific rigour with the richness of data obtainable from qualitative methodologies.

Practice implication: The applied procedures may be helpful for multi-centre focus group research on other topics.

## 1. INTRODUCTION

The advantage of qualitative research is the richness of the collected data. However, for research purposes, these data need to be interpreted and coded in a valid and reliable way, for instance by qualitative content analysis. Qualitative content analysis techniques seek to classify the discussion material into an effective number of categories that represent similar meanings. According to Hsieh and Shannon [1], qualitative content analysis can be defined as a "research method for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns". The validity of the inference is ensured by complying with a systematic coding process. In other words, content analysis allows researchers to interpret subjective data in a scientific manner.

When analysing focus group discussions, researchers have the choice between two main options, the inductive content analysis (or conventional content analysis) or the deductive content analysis (or directed content analysis) [1], [2] and [3]. The specific type of content analysis approach chosen varies according to the purpose of the research and the problem being studied. This distinction is primarily based on the different way (inductive or deductive) the categories are derived from the text.

In inductive content analysis coding categories are derived directly and inductively from the raw data. Researchers avoid using preconceived categories, allowing the categories and names for categories to 'flow from the data' instead. They immerse themselves in the data to allow new insights to emerge. Similar to a grounded theory approach, the main purpose is to develop theories. The advantage of the conventional approach of content analysis is that direct information is gained from the study participants without preconceived theoretical perspectives having been imposed [1], [2] and [3].

Deductive content analysis is guided by a more structured process than in an inductive approach. The deductive approach is based on previously formulated, theoretically derived categories and the initial coding starts with a theory or relevant research findings. Using existing theory or prior research, researchers begin by identifying key concepts or variables as initial coding categories of analysis, bringing them in connection with the text [1], [2] and [3].

Hsieh and Shannon [1] indicate as appropriate the use of a conventional content analysis when "existing theory or research literature on a phenomenon is limited". On the other hand a deductive approach should be preferred when the purpose is "to validate or extend conceptually a theoretical framework or theory".

We decided to use the inductive approach for analyzing focus group discussions on the quality of physician communication from a patient perspective which were held in four western European countries. The analysis of these focus group discussions should offer new insights into lay people's perspective on doctor–patient communication in different countries and contribute to the development of patients' theory on physicians' communicative performance. Citizens of four different countries were involved in this large international multi-centre focus group study (GULiVer). The study draws its name GULiVER from the four participating centres: Ghent University (Belgium), Utrecht University/NIVEL (the Netherlands), Liverpool University (United Kingdom) and the University of Verona (Italy). A detailed description of the research protocol is provided elsewhere [4].

Cross country comparisons of focus group discussions as in GULiVer present linguistic, procedural, methodological challenges which have to be resolved in order to obtain a reliable, consensus based coding system with which to analyze the huge amount of qualitative data.

The present paper, which is the second of a series, aims to give an example of a standardized approach to analyze qualitative data from a multi-centre study and describes the step-by-step procedures leading to the final coding system, to be applied, eventually, to the entire set of focus groups. The procedures described are based on the data collected in Utrecht, Liverpool and Verona. Ghent joined the project later and did not contribute to the development of the coding system.

## 2. METHODS

### 2.1. Study sample and focus group task

Recruitment of the lay people took place in and around Liverpool (UK), Verona (Italy) and Utrecht (Netherlands) in the early summer of 2008; in 2009/2010, Ghent (Belgium) joined the project. People were approached in public areas, such as shopping centres, and via calls in free local papers. The same recruitment procedure was used in all countries. Inclusion criteria were age over 18 years; at least one GP-visit over the last 12 months; speaking the country's language; not being involved in a medical lawsuit or formal complaint during the last two years. The selection of participants was stratified by gender (separate male and female panels) and age (18–30; 31–49; >50) in order to ensure a heterogeneous distribution of the sample and comparable results. A total of 259 lay people participated in the study, evenly distributed over 35 focus groups, with six to eight persons per group and nine focus groups per centre, except Ghent with eight focus groups only.

The groups watched the same set of videotapes of standardized medical consultations (OSCEs) conducted by 4th-year medical students from Liverpool Medical School. The videotapes were selected to provide a maximum variation in the quality of student doctors' communication as rated by their examiners. The videotapes presented gynecological consultations about clinical problems that are associated with high levels of emotional distress. Students were expected to use their communication skills to recognize and

manage such a distress. Specifically, female simulated patients presented one of the following scenarios: (a) vaginal discharge, related to unprotected sex with an unfamiliar partner, and a possible diagnosis of a sexually transmitted disease; (b) severe menstrual pain, resulting in frequent absence from work, with a fear of job dismissal due to the presenting complaint.

The video tapes were either dubbed (Italy) or subtitled (the Netherlands) to conform to the accepted practice of displaying English language television in the two countries. Transcripts of the consultations in the language of the participants were also provided. The participants were invited to evaluate the student doctors' performance using rating scales and writing comments to dialogue pieces on the transcripts before discussing and sharing their views in the group. The focus group meetings lasted about 6 h (including coffee and lunch breaks), of which 2 h were actually spent on group discussions. The study procedures were described in great detail and similar in all countries. A detailed description of participants inclusion and exclusion criteria, recruitment process, allocation in the focus groups and socio-demographic characteristics are described elsewhere [4].

Ethical approval was granted by the Faculty of Medicine Education Research Committee of the University of Liverpool. The students of the selected OSCE examinations gave informed consent to participate in a European study where their OSCE videos would be viewed and rated by panels of experts in medical communication and by focus groups of lay people in the different countries. In the Netherlands, no ethical approval is needed when (a) no patients are involved, and (b) participants do not receive a medical intervention which can influence their health. The Belgium and Italian centre individually asked and obtained approval from the local Ethics Review Committee. The lay participants of all centres signed an informed consent before entering the study.

### 2.2. Inductive content analysis procedures

To establish trustworthiness of data and to obtain valid and reliable inferences, each centre had to adhere to a set of systematic and transparent procedures for arranging and processing the raw data. The entire process lasted about 9 months during which monthly telephone meetings as well as two life investigator meetings were organized in order to discuss all emerging issues (see below). Overall, two researchers from Netherland (J.B., L.V), three from Italy (F.M., M.R., G.D.), and one from the UK (I.F) were involved in the development of the GULiVer framework.

### 2.2.1. Arranging focus group material and defining the unit of analysis

First, the video registrations of the focus groups were transformed into written text format by their complete transcription into the country language. Questions and comments by the group facilitators were included to check their neutrality. Non verbal behavior of participants and sounds or pauses were not transcribed. Italian and Dutch focus group discussions were translated into English.

Second, the unit of analysis had to be defined. The unit of analysis is a "segment of text that is comprehensible by itself and contains one idea, episode, or piece of information" [5]. This is one of the most important decisions for content analysis [6] and [7]. A meaningful unit in qualitative content analysis is usually a theme [7], particularly when the researcher is looking for the expression of an idea [8]. Codes will be assigned to text pieces of any size as far as they relate to the theme. The theme which in GULiVer defined the units of analysis included therefore all issues referring to any participants' comment on a particular element of the student doctor's communication.

### 2.2.2. From units to categories: rules for the content analysis within a country

Researchers of each centre analysed the same three randomly selected focus group transcripts (one from each country) and elaborated their country specific coding system—British, Dutch, and Italian. In line with the guidelines in the literature [1], [2], [9] and [10], the procedure to ensure a consistent and uniform approach among the centres considered five steps (Fig. 1).

Step 1. Repeated reading to gain a sense of whole and highlighting phrases that appear to capture the theme connected with the research question "lay persons' view of doctor patient communication".
Step 2. Taking notes of the content area to which the highlighted phrases referred.

Step 3. Grouping of content areas expressing similar concepts into mutually exclusive categories and giving a first label, such as 'being straightforward/pertinent'.

Step 4. Revising the categories and subsuming previous or formulating new categories if necessary and identifying linked categories, obtained by the same procedure.

Step 5. Organizing linked categories into a hierarchical structure of areas, categories and subcategories and final checking of category overlaps to merge or to divide into subcategories.

[FIGURE 1]

### 2.2.3. The development of the GULiVer frame work

All authors, through a series of investigator meetings, collaborated on developing the preliminary framework and arrived at consensus on the final coding framework, after having worked through the following procedure.

The coding systems elaborated by each centre were compared in a plenary meeting, starting from discussing concordant and discordant categories. Categories which identified similar themes were retained for the GULiVer coding system and relabeled. Disagreement on coding categories was approached by analyzing and discussing the procedure which had led to the identification of the critical category and ended when new agreement was found. Since too many categories may lead to difficulties in organizing text data in an effective way [11] we agreed on a maximum number of 15 categories.

The majority of the categories and subcategories of step 5 identified in each centre (e.g. 'Commitment', 'Listening', 'Self confidence', and 'Summarizing') coincided. The Dutch centre identified three additional subcategories not considered by the others: 'Introduction or ending of conversation', 'Asking permission', 'Using stock phrases'. Additional subcategories by the Italian centre were: 'Being informal', 'Being prepared/skilled/experienced', 'Inspire trust', and 'Verifying'. All were considered relevant and can be found in the final GULiVer system. The English centre did not identify subcategories which were not also identified by the other centres.

Most of the differences between the three countries regarded the hierarchical organization of areas, categories and subcategories. For example, the Italian centre grouped the subcategories 'Being reassuring', 'Being empathic', 'Creating an inviting atmosphere' and 'Putting the patient at ease' into a main category 'Supporting and dealing with patient emotions', whereas the Dutch centre identified comparable subcategories ('Reassurance', 'Emphatic', 'Inviting attitude', and 'Pleasant attitude') but grouped these into a main category 'Affective attitude of the doctor'.

To level this difference we agreed to include the subcategories 'Reassurance/trust', 'Empathic', 'Inviting attitude' and 'Pleasant attitude' (see Fig. 2) in the category 'Affective attitude' to become part of a newly created area 'Emotion oriented expressions'. In addition, all country specific subcategories fitting the category 'Affective attitude' were removed from their previous allocation and included here. These were 'Showing interest/committed', 'Facilitating', 'Neutral/No personal remark' and 'Listening'. The procedure was repeated where needed until a final structure of the GULiVer coding system was obtained. As recommended by Lincoln and Guba [12] the categories should be defined in a way that they were internally as homogeneous as possible and externally as heterogeneous as possible. This group process led to the final GuLiVer framework shown in Fig. 2.

[FIGURE 2]

It considers the areas 'Non verbal communication', 'Process oriented expression', 'Task oriented or problem focused expressions', 'Affective or emotional expressions' and 'Physician characteristics', comprising 12 categories and the specific doctor behaviors for each category.

An additional area/category 'other' was included for participants' expressions not attributable to any of the existing categories and subcategories (e.g. "I'd like my doctor knew me. I've been having the same doctor for 30 years". This phrase has to do with the continuity of care which is not included in the coding system and was classified as 'Other').

### 2.2.4. Development of category defining criteria and coding rules

Since multiple coders had to be involved, a coding manual was developed to ensure coding consistency. The areas and their respective categories and subcategories, if any, were defined and rules for assigning codes were fixed, backed up with examples [6].

Moreover, coding rules for verbal turn divisions were established. A verbal turn was defined, according to the definition by Roter and Larson [13], as a complete thought, expressed by one participant without any interruptions by other speakers. All verbal turns of participants had to be coded, including uninformative, irrelevant, incomprehensible or nonsense comments which were given a special code (non classifiable). Turns often contained multiple information regarding participants' preferences/opinion about student doctors' performance. In this case phrases referring to different content categories/subcategories were to be divided and classified separately. Accordingly the phrase "She reassured the patient in the end and she explained what the doctor is going to do" was divided into two different parts (Table 1). The example shows also that each segment has to be given the area, category and subcategory code.

[TABLE 1]

### 2.3. Inter-rater agreement

#### 2.3.1. Phase 1

A pilot inter-rater reliability test of the coding system was carried out on three randomly selected focus group transcripts. Three independent raters, one from each centre (L.V. from Utrecht, G.D. from Verona and I.F from Liverpool), coded the group discussions according to the established coding rules and definitions. Inter-rater agreements were calculated for categories and subcategories adopting Krippendorff's alpha reliability coefficient. It ranges from 0 (complete disagreement) and 1 (complete agreement) and its use is recommended when data regard complex categories derived from content analysis [14] and [15]. A specific procedure was applied using the free statistical software R and implementing the package "concord" [16].

Overall, the reliability at this stage was relatively low with an alpha of 0.30 (Table 5). An in-depth analysis of the distribution of unreliable categories in the three centres showed that major differences occurred when coders identified within the same turn a different number of relevant contents, that is units of analysis. This difficulty of raters in reliably segmenting turns into units of analysis is a frequent drawback in qualitative research [17]. Distribution differences regarded also the categories 'Structuring', 'Task oriented attitude of the doctor' and 'Collecting information', while the disagreement on the categories 'Socio-demographic characteristics' and 'Other' seemed due to their rare occurrence in all three centres.

#### 2.3.2. Phase 2

The distribution differences were analysed and discussed in a second plenary investigator meeting, leading to a revision of the rules for turn segmentation following the recommendations of Pfeil and Zaphiris [17] (e.g. repeated concepts in a turn were to code only once) and of the coding definitions for the critical categories.

The three raters from the pilot inter-rater reliability test practiced the revised version, re-coding together one of the focus groups. A new set of three focus groups was then randomly selected and coded independently by the same three raters.

### 3. RESULTS

#### 3.1. Inter rater reliability phase 2

As shown in Table 5, the revised rules of turn segmentations and definitions of critical categories improved inter-rater reliability (alpha = 0.48). This value corresponds to a moderate agreement, but given the great number of categories, the use of more than two raters, and the more conservative index of Krippendorff's alpha, such a finding may be considered acceptable [18]. However, in order to further improve the reliability of the subsequent coding process of all focus groups it was decided that any potential coding doubts had to be discussed together in monthly phone-meetings and to be resolved by consensus decisions.

### 3.2. The GULiVer coding system and the coding manual

The procedures described led to the development of a coding system with a specific coding manual containing all areas, categories and subcategories, definitions and rules for verbal turn division. The final version of the manual contains precise operational definitions of the main five areas – 'Non verbal communication', 'Process oriented expression', 'Task oriented or problem focused expressions', 'Affective or emotional expressions' and 'Physician characteristics – and of the 12 categories and 41 subcategories listed in Fig. 2. The coding of subcategories is facilitated by including relevant examples from the focus group discussions. An example of the coding manual is given in Table 2 for the area 'Process oriented expressions', with its four categories 'Structuring', 'Summarizing', and 'Speech peculiarities' and their respective subcategories.

[TABLE 2]

Besides the content of the discussions, we wanted to capture some additional aspects of focus group participants' comments, as is shown in the following examples.

To start with, the two statements in Table 3 indicate a different view on the use of summaries. The first is negative, the second positive (last column). Since it occurred often that the same student behavior was disliked by some but appreciated by others, this important information is taken into account by adding + or − to the respective statements. This allows later on to identify the agreement/disagreement of the participants toward certain behaviors simply by a plus and minus count within each category/subcategory over the focus groups.

[TABLE 3]

The reasons for expressed preferences and dislikes (e.g. "as if she was inattentive") are similarly precious bits of information and therefore starred, as shown in this example; such explanations can be subsequently retrieved, when we analyze the data.

Another relevant distinction is whether participants responded to the doctor on the video, or were prompted by the stimuli provided by the videos to refer to their own experiences with doctors, which led to general considerations, as shown in Table 4. To differentiate between such general opinions (e.g. "I like it when a doctor is easy going") and specific statements referring to the student doctors' performance (e.g. "this doctor was easy going which I liked"), the code = (doctor) in the last column is used for general comments, while +/− indicate appraisal or criticism regarding the student doctors of the videos (Table 5).

[TABLE 4][TABLE 5]

The above example also shows that the discrepancies in +/− codes applied for categories and student doctors inform us whether the student doctor was in some way lacking to perform or performed poorly, a behavior otherwise highly appreciated by the participant, for example to be reassured (reassurance/trust).

A last aspect that can be captured is that sometimes participants will repeat their opinions about the same doctor or on doctors in general. If this happens in a new turn, the statement will be coded but marked with an R in order to be able to calibrate its proper weight in contributing to the discussion. The R will not be used when the same opinion is given about different student doctors.

### 3.3. Excel data set form

Fig. 3 illustrates the set up of the excel file for the coding procedure (Fig. 3) which, in addition to the data above, contains the:
- Number of the focus groups from 1 to 35.
- Identification number of the participant who makes the statement to allow analyses according to participants' different background characteristics, collected at baseline.
- Verbal turn numbers.
- Text of the statement.
- Identification number of the student doctor to which the participant's statement refers.

[FIGURE 3]

## 3.4. Coding the total set of multi language focus groups and critical issues

In view of the future coding of the total set of focus groups two options were available: a rater from each centre either codes the set of focus group discussion in the original language version from his/her own country; or codes a set of focus groups from all three countries, based on a random selection from the total sample in the translated English version. We asked the opinion of two experts in qualitative analysis, Salmon and Young [19]. After a discussion balancing pros and cons, the first option was chosen, in so far the coding of national language transcripts, by maintaining the original speech and colloquial expressions related to each country's culture, was liable to ensure a greater validity of coding.

The forthcoming findings from this database will be reported in a subsequent paper including also the focus group data from the Belgium centre. The training of the Belgium researchers in the use of the coding system by L.V. of the Dutch centre was facilitated by sharing the same language.

## 4. DISCUSSION AND CONCLUSIONS

### 4.1. Discussion

Qualitative research is particularly appropriate in relatively new areas when the main goal of the study is to learn more about people's perspectives and experiences. Accordingly, qualitative approaches to doctor–patient communication can provide precious information about the quality of communication from a patient perspective and we can observe a growing use of focus groups for such a purpose [20], [21], [22] and [23]. This methodology "empowers" participants to become active partners in the research process and develop an authentic patient centred approach [24].

Despite the great richness of data offered by a qualitative approach, the main criticism raised is the lack of standardized and scientifically rigorous procedures [25] and [26]. Specifically, the content analysis stage presents some important limitations: (a) the risk of a strong interference of the researcher who may "interpret" (research bias); (b) the lack of reproducibility; (c) the unfeasibility to separate individuals and group opinions with the risk to level or to overestimate participants' contributions and the consequent lack in generalizability when the group is considered the only unit of analysis [27] and [28].

These issues become still more important when complex cross national comparisons of focus group discussions on doctor–patient communication are carried out, as in the GULiVer study.

The described GULiVer procedures are an example of how qualitative data from a large set of focus groups from different countries, conducted under standardized conditions, may be coded and analysed by trying to minimize the above limits and to preserve both the richness of qualitative data and the scientific rigour. In order to reach this aim, several steps need to be thoughtfully followed.

First of all, it is fundamental to choose the most adequate methodological approach: inductive versus deductive. The main aim of the GULiVer project was to construct a patient's theory on doctors' performance. Therefore, the qualitative analysis of a set of focus groups needs to avoid, or at least limit as much as possible, researchers' subjective interpretations, which asks for an inductive, rather than a deductive approach.

Secondly, to make valid and reliable inferences, qualitative content analysis needs to entail a set of systematic and transparent procedures for processing data. A clear and adequate description of qualitative studies' procedures, particularly with regard to the data analysis is essential to ensure a high standardisation and improve the scientific value of qualitative results. As emphasized in literature, many qualitative studies are generally lacking in such a detailed and transparent methodological approach [29] and [30]. In this study all procedures have been described in detail in a study protocol and were established before we started with the data analysis. Specifically, each procedural step of the GULiVer coding system has been fixed in detail in order to raise the reliability of the system and to reduce researcher bias.

Furthermore, the researchers had the opportunity to meet regularly to compare their coding, discuss the congruencies and dissimilarities, and to find consensus. The combination of plenary investigator meetings at the crucial stages of coding system development, combined with monthly telephone meetings in order to deal with remaining issues proved to be an effective way to create an indepth understanding of the evolving coding system in all researchers and to ensure standardisation of the coding procedures. Over the nine months of development of the coding system, various adjustments have been made to the coding system as well as to its manual in order to ensure a progressively more detailed and shared version as evidenced by

the improved agreement between coders. Category definitions were better specified and completed with numerous examples.

The reliability of the results was further strengthened by offering the focus groups the same videotaped consultations as source material, which means that the content of the various focus group discussions referred to the same doctor–patient communication. Such a standardized approach makes it easier to code the content of the group discussions in a reliable way and to compare the discussions over countries.

Finally, agreement between the raters is essential in order to improve the reliability of the coded data. In this study the results of a pilot inter-rater reliability test made it necessary to discuss the content of disagreements and to refine the system, while a second inter-rater reliability assessment was needed to validate the usefulness of the revisions in improving the inter-rater agreement. In this regard, the study is a good example of the advantages of a mixed-method approach above a qualitative study, because in most qualitative studies no reliability data are presented [18]. However, the moderate inter-rater reliability scores in this study also show that new methodologies to improve the reliability of qualitative findings need to be studied and applied.

A specific strength of this study is the possibility to distinguish between individual and group level of analysis. All focus groups have been video and audio taped so that it was possible to identify each participant's interventions during the discussions. This will enable researchers to identify the quantity and content of each participant's contributions to the discussion. Such data may be linked to the participants' clinical and sociodemographic characteristics, collected at baseline together with their own communication preferences and trust in health care. The identification of individual contributions to group discussions offers also other research opportunities, for example regarding the communication dynamics within focus groups. Moreover, the additional information collected on doctors' behaviors which were generally appreciated but were criticized as poorly performed by the student doctors (e.g. plus given to the categories and minus given to the doctors) may allow additional and more complex analyses than the straightforward category counts of frequencies. For example, such an analysis may help to better understand not only what communicative behavior is expected by patients, but also how they want the doctor to perform it in practice.

Limitations were that all researchers were fluent in English, but not always in Dutch or Italian. Subtle linguistic or cultural differences might have remained concealed, as the researchers were trying to find consensus rather than cultural diversity when developing this classification system. All researchers had a psychological or psychiatric background, which might have influenced the naming of the categories. In order to test the validity of the classification system, a follow up project will be developed in which the results of this study will be fed back to lay people to ask whether or not they recognize the results as relevant to the quality of communication from a patient perspective

### 4.2. Conclusion

The criticisms often directed to qualitative approaches may lead researchers to prefer quantitative studies which ensure a greater objectivity and reproducibility of results but at the same time run the risk of losing valuable information, especially in the field of doctor–patient communication. Qualitative research offers the opportunity to explore patients' perspective limiting researcher's interference. The procedure described in this paper is an example of how qualitative data analysed in a systematic way can be able to balance the richness of data obtainable from qualitative methodologies with the scientific rigour of quantitative approaches.

This mixed method approach will get us closer to understanding the different communicative needs and expectations people may harbor when relating with their doctor. The results of these analyses will be published in future papers.

### 4.3. Practice implications

The applied procedures show how the methodological problems of international multi centre focus group research can be dealt with. They illustrate how qualitative data can be transformed into quantitative data, bridging the gap between quantitative and qualitative research. Future research may take advantage of the described methodology for multi centre focus group research on other topics.

### Disclosure

We confirm all patient/personal identifiers have been removed or disguised so the patient/person(s) described are not identifiable and cannot be identified through the details of the story.

### Conflict of interest

The authors have no conflict of interest that could have influenced the paper.

### REFERENCES

[1] H.F. Hsieh and S.E. Shannon, Three approaches to qualitative content analysis, Qual Health Res 15 (2005), pp. 1277–1288.

[2] Mayring P. Qualitative content analysis. Forum Qual Soc Res 2000;1.

[3] S. Elo and H. Kyngäs, The qualitative content analysis process, J Adv Nurs 62 (2008), pp. 107–115. [4] Moretti F. GULiVER—travelling into the heart of good doctor–patient communication from a patient perspective. An international multicentre study. PhD Thesis, University of Verona; 2010.

[5] J. Schilling, On the pragmatics of qualitative assessments. Designing the process for content analysis, Eur J Psychol Assess 22 (2006), pp. 28–37.

[6] R.P. Weber, Basic content analysis, Sage Publications, Newbury Park, CA (1990).

[7] U.H. Graneheim and B. Lundman, Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness, Nurse Educ Today 24 (2004), pp. 105–112.

[8] E.H. Bradley, L.A. Curry and K.J. Devers, Qualitative data analysis for health services research: developing taxonomy, themes, and theory, Health Serv Res 42 (2007), pp. 1758–1772.

[9] Y. Zhang and B.M. Wildemuth, Qualitative analysis of content. In: B. Wildemuth, Editor, Applications of Social Research Methods to Questions in Information and Library, Book News, Portland (2009).

[10] Busch C, De Maret PS, Flynn T, Kellum R, Le S, Meyers B, et al. Content analysis. Colorado State

[11] J.M. Morse and P.A. Field, Principles of data analysis. In: J.M. Morse and P.A. Field, Editors, Qualitative research methods for health professionals (2nd ed.), SAGE Publications, London (1995).

[12] Y.S. Lincoln and E.G. Guba, Naturalistic inquiry, SAGE Publications, Beverly Hills, CA (1985).

[13] D. Roter and S. Larson, The Roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions, Patient Educ Couns 46 (2002), pp. 243–251.

[14] F. Mun Oz Leiva, F.J. Montoro R'Ios and T.L. Martinez, Assessment of interjudge reliability in the open-ended questions coding process, Qual Quant 40 (2006), pp. 519–537.

[15] A. Hayes and K. Krippendorff, Answering the call for a standard reliability measure for coding data, Commun Methods Meas 1 (2007), pp. 77–89.

[16] J. Lemon, Concord: concordance and reliability. R package version 1.4.9

[17] U. Pfeil and P. Zaphiris, Applying qualitative content analysis to study online support communities, Univ Access Inf Soc 9 (2009), pp. 1–16.

[18] M. Lombard, J. Snyder-Duch and C. Campanella Bracken, Content analysis in mass communication. Assessment and reporting of intercoder reliability, Hum Commun Res 28 (2002), pp. 587–604.

[19] P. Salmon, N. Mendick and B. Young, Integrative qualitative communication analysis of consultation and patient and practitioner perspectives: towards a theory of authentic caring in clinical relationships, Patient Educ Couns (2010)

[20] T.C. Olde Hartman, L.J. Hassink-Franke, P.L. Lucassen, K.P. van Spaendonck and C. van Weel, Explanation and relations. How do general practitioners deal with patients with persistent medically unexplained symptoms: a focus group study, BMC Fam Pract 24 (2009), pp. 68–73.

[21] E. Farin, L. Gramm and D. Kosiol, Development of a questionnaire to assess communication preferences of patients with chronic illness, Patient Educ Couns 82 (2011), pp. 81–88.

[22] M. Quirk, K. Mazor, H.L. Haley, M. Philbin, M. Fischer and K. Sullivan et al., How patients perceive a doctor's caring attitude, Patient Educ Couns 72 (2008), pp. 359–366.

[23] McKinstry, Do patients wish to be involved in decision making in the consultation, Brit Med J 321 (2000), pp. 867–871.

[24] J. Bensing, S. Dulmen and H. de Haes, Communicatie tussen patiënten en hun hulpverleners. In: L. Lechner, I. Mesters and C. Bolman, Editors, Gezondheidspsychologie bij patiënten, Koninklijke van Gorcum BV/Open Universiteit, Assen, The Netherlands (2010), pp. 109–134.

[25] E.E. Johansson, G. Risberg and K. Hamberg, Is qualitative research scientific, or merely relevant? Research-interested primary care and hospital physicians' appraisal of abstracts, Scand J Prim Health Care 21 (2003), pp. 10–14. Full

[26] D. Horsburgh, Evaluation of qualitative research, J Clin Nurs 12 (2003), pp. 307–312.

[27] M. Brod, L.E. Tesler and T.L. Christensen, Qualitative research and content validity: developing best practices based on science and experience, Qual Life Res 18 (2009), pp. 1263–1278.

[28] R. Whittemore, S.K. Chase and C.L. Mandle, Validity in qualitative research, Qual Health Res 11 (2001), pp. 522–537. Full Text via CrossRef | View Record in Scopus | Cited By in Scopus (87)

[29] N. Mays and C. Pope, Rigour and qualitative research, Brit Med J 311 (1995), pp. 109–112.

[30] D.J. Cohen and B.F. Crabtree, Evaluative criteria for qualitative research in health care: controversies and recommendations, Ann Fam Med 6 (2008), pp. 331–339.
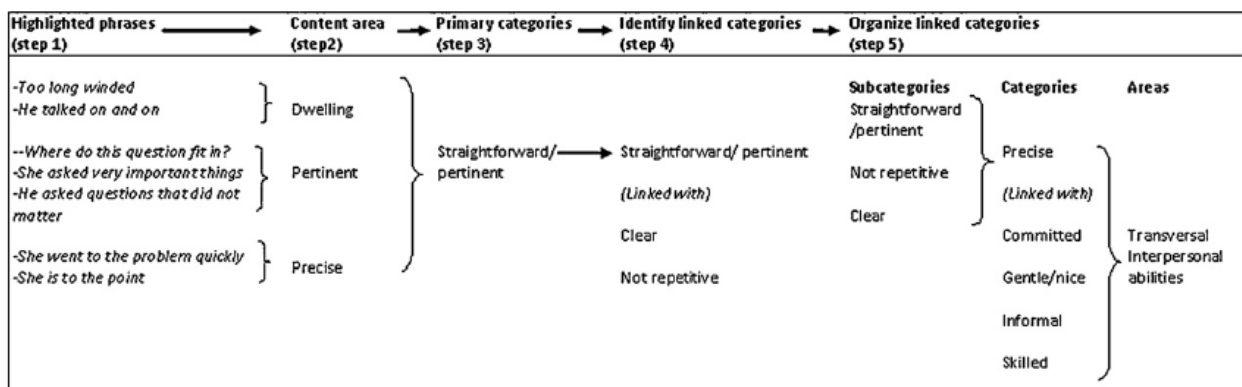
**[TABLES AND FIGURES]**



Fig. 1. The five steps for text analysis: from units to categories.

| | | Facial expression |
|---|---|---|
| Non verbal communication | Non verbal behavior | Eye contact |
| | | Touch |
| | | Others |
| | | Reading and Writing. |
| | | Laughing. |
| Process-oriented expressions | Structuring | Changing of topics and signposting |
| | | Flexibility |
| | | Time issues |
| | | Opening or closing of the interview |
| | Summarizing | Summarizing |
| | Patient-involving | Sharing plans/ideas |
| | | Asking permission |
| | | Verifying |
| | Speaking peculiarities | Repetition |
| | | Fillers |
| | | Comprehensibility |
| Task-oriented/ problem-focused expressions | (Task oriented) Attitude of the doctor | Self-confident. |
| | | Complete picture |
| | | Businesslike / Straight to the point |
| | | Other/general |
| | | Clarity of interview |
| | | Competency |
| | Collecting information | Medical |
| | | Biopsychosocial |
| | | Psychosocial |
| | Giving information | Medical |
| | | Biopsychosocial |
| | | Psychosocial |
| | Providing Solution | Providing solutions |
| Affective-oriented/ emotional expressions | (Affective oriented) Attitude of the doctor | Inviting attitude |
| | | Pleasant attitude |
| | | Showing interest/ committed |
| | | Empathic |
| | | Facilitating |
| | | Reassurance / trust |
| | | Neutral / No personal remark |
| | | Listening |
| Socio demographic characteristics | Socio demographic characteristics | Doctor's gender |
| | | Doctor's age |
| | | Doctor's ethnicity |
| Other | Other | Other |

Fig. 2. The GULiVer framework.

**Table 1**
Coding verbal turns containing multiple information: an example.

| Phrase | Area | Category | Subcategory |
|---|---|---|---|
| *She reassured the patient in the end* | Affective or emotional expressions | Attitude of the doctor | Reassurance/trust |
| *She explained what the doctor is going to do* | Process oriented | Structuring | Changing of topics and signposting |

**Table 2**
An example of the coding manual: the area "Process oriented expressions".

Process oriented expressions
All the comments about the way a doctor is handling the process of the conversation. These are specific elements of: (a) the structure of the conversation (opening/closing, link between different part of the conversation, flexibility, time), (b) the degree in which a GP involves his patients, and (c) the structure of the doctor's speech (repetition, fillers, interruptions, jargons).
Structuring
Every statement related to any aspect of guidance of the conversation, starting from the introduction to the ending of the conversation.

| | |
|---|---|
| Changing of topics and signposting | Comments on the way the GP changed topics 'out of the blue' or signposted to the patient what will happen next ('*He had forgotten to ask something.* **Then asked something out of the blue**') ('*he* **introduced his question very well**. *He frequently said things like "now I'm going to ask about this"*') |
| Opening or closing of the interview | A specific remark about the beginning or end of the conversation ('*No doctor shook hands*') |
| Time issues | Every comment regarding the duration (time) of the conversation ('*I also felt as if she runs out things to say, she was filling time*' or '*They were hurried*') |
| Flexibility | Comments about the structure of the conversation (Doctor was flexible or used to go through a checklist) ('*Apparently they have to follow some kind of list*') |

Summarizing
Statements referring to the way in which a GP summarized the topics that were discussed in a conversation. We will only use "summarizing" when a participant is referring to a structured summary of long pieces of information (otherwise it will be coded as 'verifying'). ('*But he did say, let's sum up the story and I think that that was good*').
Patient involving
All aspects that are related to the involvement of patients in the conversation.

| | |
|---|---|
| Sharing plans/ideas | Every statement concerned with exploring a patient's explanation of the complaint, or the patient's expectation regarding the follow-up steps the GP has to take ('*Asking patient what she thinks should be done now*'). |
| Asking permission | All the statements about the GPs asking of permission or apologizing for asking questions or doing certain things ('*Asking permission to take notes. You don't go for a chat*'). |
| Verifying | Statements where the doctor asks the patient to confirm that they have understood what was said by the patient ('*To be confirmed is ok, when he says "you said so and so, did I got it right?" or "is it this you intended to say?"*'). |

Speaking peculiarities
Statements about the aspects that interrupt normal fluent speech.

| | |
|---|---|
| Repetition | Statements about the GP's repetition of the same questions, irritating repetitions ('*She repeats the questions*') |
| Fillers | Minimal comments that the GP uses, e.g. uses 'ok' all the time ('*He says okay all the time*') |
| Comprehensibility | Stopping mid-sentence, medical/technical jargon ('*She did not complete her sentences*') |

**Table 3**
An example of positive and negative value attributed by participants to a category.

| Statement | Area | Category | Subcategory | Starring phrases | +/− (category) |
|---|---|---|---|---|---|
| *I placed her last because at a certain point of the visit she has summarized [...],* **it is as she had been inattentive** | Process oriented | Summarizing | − | * | − |
| *he did say let's sum up the story and I think that was good* | Process oriented | Summarizing | − | | + |

+/−/= (category): positive, negative or irrelevant value attributed to a category;*: the participant explains why he/she does not like "summarizing".

**Table 4**
Positive and negative value attributed by participants to a student doctor's behavior.

| Statement | Area | Category | Subcategory | +/− (cat) | +/−/= (doctor) |
|---|---|---|---|---|---|
| *a doctor has to be thoughtful* | Affective or emotional expressions | Attitude of the doctor | Showing interest in patient / commitment | + | = |
| *I would not be reassured by this doctor who tells me that other people have the same problem. It does not help* | Affective or emotional expressions | Attitude of the doctor | Reassurance/trust | + | − |

+/−/= (category): positive, negative or irrelevant value attributed to a category; −/− (doctor): positive or negative value attributed to a specific student doctor's behavior; =(doctor): opinions referred to a generic doctor behavior.

**Table 5**

Inter-rater on categories (Krippendorff's alpha).

| Subcategories | Alpha phase 1 | Alpha phase 2 |
|---|---|---|
| Non verbal behavior | 0.58 | 0.59 |
| Structuring | 0.26 | 0.59 |
| Summarizing | _[a] | |
| Patient-involving | 0.42 | 0.41 |
| Speaking peculiarities | 0.32 | 0.45 |
| Task-oriented doctor attitude | 0.32 | 0.40 |
| Collecting information | 0.24 | 0.31 |
| Giving information | 0.18 | 0.40 |
| Providing solution | _[a] | |
| Affective doctor attitude | 0.34 | 0.54 |
| Socio-demo characteristics | 0.18 | 0.51 |
| Other | _[a] | |
| Categories | 0.30 | 0.48 |

[a] Categories with no sub categories.

| ID focus | ID partic | Turn N | Text | Area | CAT/Subcat | Cat +/-/= | Doc +/-/= | * | R | ID doc |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 18 | 33,1 | Concerning 4122 and his being professional. At the beginning he presented himself asking for data and showing the form, asking the patient if the data were correct, I did not found this professional | Task-oriented/problem-focused expressions | TASK-ORIENTED ATTITUDE OF THE DOCTOR/Competency | positive | negative | * | | 4122 |
| 3 | 18 | 33,2 | even more so when he asked questions about her private life: if she had a boyfriend, a pet, children, they did not seem to be of relevance for her problem, for this I did not chose him. | Task-oriented/problem-focused expressions | COLLECTING INFORMATION/Social | negative | negative | * | | 4122 |
| 3 | 19 | 34,0 | Asking about pets might explore pthe presence of possible allergic reactions which might cause further problems. | Task-oriented/problem-focused expressions | COLLECTING INFORMATION/Social | positive | positive | * | | 4122 |
| 3 | 17 | 35,0 | asking if she has a boyfriend I do not find pertinet to her problem | Task-oriented/problem-focused expressions | COLLECTING INFORMATION/Social | negative | negative | * | | 4122 |
| 3 | 19 | 36,0 | To understand her family and relational context might help to understand better her problems. | Task-oriented/problem-focused expressions | COLLECTING INFORMATION/Social | positive | positive | * | R | 4122 |
| 3 | 16 | 37,1 | It's true but he does not explain it, he does not say "Signora, I have to ask for these things because they might have an impact and therefore..." the other two did not do this either. | Process-oriented expressions | PATIENT-INVOLVING/Sharing plans/ideas | positive | negative | * | | 4122 |
| 3 | 16 | 37,2 | I was impressed that this doctor (4122) was the only one to whom said "I am in difficulty talking", while with the other male doctor at the end she said: I feel more relieved. | Affective or emotional expressions | AFFECTIVE ATTITUDE OF THE DOCTOR/Inviting attitude | positive | negative | | | 4122 |
| 3 | 21 | 38,0 | I do not know to what degree one can judge the technical aspects or the communication, that is the type of questions. We talk about communication, about the trusting relationship between two persons who have to resolve a problem; concerning the technical aspect I do not know how much of the talk about pets or family problems, about children, boyfriends are of value in the global evaluation | NC/citizen | | | | | | |
| 3 | 20 | 39,1 | he gave me the impression that he was taking time, there where he wrotes down some notes, trying to put the patient at ease. | Affective or emotional expressions | AFFECTIVE ATTITUDE OF THE DOCTOR/Inviting attitude | positive | positive | * | | 4122 |
| 3 | 20 | 39,2 | The patient thus enters the office being aware of this, namely it is up to the doctor to take the inizitaive of the treatment. | NC/citizen | | | | | | |
| 3 | 0 | 40,0 | Regarding this I have to clear a point to avoid confusion. We would like you to look at these interviews as interviews by real doctors. Here indeed we have a simulation. But if this would be areal encounter, would it work? | NC/facilitator | | | | | | |

Fig. 3. An example of the excel file.