From Sample to Population Generalization in Qualitative Research

Florentina Scârneci-Domnișoru

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

The generalization of research results is not among the concerns of qualitative researchers. Most of them consider that this is not the aim of qualitative studies. Nevertheless, those who raise the issue of generalization believe either that generalization is impossible or that transferability would be the only accessible variant of generalization in qualitative research. Despite this, the article presents a theoretical sampling procedure that allows internal generalization (from sample to population) of the results of qualitative research that have multiple units of analysis. It describes four steps that ensure the transparentizing of the sampling process by specifying exactly the population of interest and the applied sampling criteria. The procedure does not place emphasis on the size of the sample, but on the criteria for choosing the participants and seeks to obtain the generalization of findings following data variations.

Keywords: internal generalization, inductive generalization, theoretical generalization, theoretical sampling, theoretical representativeness.

1. Introduction

There are more and more authors who believe that the split of methodology into qualitative and quantitative was not at all a good idea (see Onwuegbuzie and Leech, 2005, Ercikan and Roth, 2006 or Vogt, 2008 and Scârneci-Domnișoru, 2018). This split had all sorts of negative consequences, some of them very harmful if we look at the research process or at the features of scientific knowledge, in general.

One of the great problems thus created is the disappearance of some terms in the language of researchers, who were grouped under the name of qualitativists. The qualitative researchers have given up some concepts that characterized research in social sciences or have tried replacing some concepts with terms specific to qualitative methodology (see, for example, replacing the validity and reliability with credibility, dependability and others, or the sample with selection).

The problem is that no matter what classification we try, no matter what new type of research we invent, no matter what methodological innovation we propose, something should remain constant – that is, the features that make research what

it is. It is a mistake to give up or to redefine contextually, for instance, the rigour, objectivity, precision, or generalization.

In this article, I will refer to generalization, a term that disappeared from the qualitative research vocabulary (it is true it was replaced with a variant of generalization – transferability – that I will describe in paragraph three). Citing other authors, Maxwell (2021) shows that 'many qualitative researchers, accepting the positivist *account* of generalization, completely rejected this as a legitimate goal for qualitative research'. Maxwell also shows that the most influential supporters of this view are Yvonna Lincoln and Egon Guba who, in different publications, hold that 'the only generalization is: there is no generalization' and that 'generalizations are *not possible*' (p. 111).

I will argue in this article that generalization is possible also in qualitative research. Moreover, I support the idea of Payne and Williams (2005) that we should turn the generalization of findings into a purpose and that we should approach the generalization matter openly when presenting research results: 'research design should *plan* for anticipated generalizations, and generalization should be more *explicitly formulated* within a context of supporting evidence' (p. 295).

The problem we have is not only that generalization is disregarded by most qualitativists, but also that many of those who carry out qualitative research (especially of the applicative ones) generalize, even if they do not use explicitly this term. The even greater problem is that decisions for entire populations are often made after applying, for instance, several interviews or after the conducting of two or three focus groups (see also Onwuegbuzie and Leech (2010, p. 881) that talk about over-generalizations by 'making general recommendations for future practice and providing general policy implications based only on a few cases'). This behaviour is dangerous and can disqualify researchers professionally, not to say how much damage it can do to the prestige of qualitative research.

Therefore, I will show in this article that generalizations are possible, but only by observing certain rules. I will argue that in qualitative research even the most quantitative generalization is possible, that from sample to population, and I will describe in detail the procedure that allows this. Before that, I will present a more detailed description of the problem in the following paragraph.

2. The problem

Qualitative research is often criticised 'for its lack of generalizability' (Bloor and Wood, 2006, p. 93). The criticism is that the findings of a study cannot be applied to a wider population and this is not a minor thing, that we could easily ignore. A famous quotation of Hans Reichenbach, a leading German philosopher of science, '*the essence of knowledge is generalization*', captures the size and importance of the problem that the knowledge generated through qualitative research faces.

The capacity to generalize results should be one of the most important features of scientific research or, as Ercikan (2009) claims: 'generalizability of research findings is one of the key criterion researchers use to determine the value and usefulness of research' (p. 211). Any researcher should count among his/her preoccupations the generalization of the results of his/her research because 'no matter how interesting or insightful the research is, if it is not generalizable then it

is not considered to be evidence that can be put to use' (Bloor and Wood, 2006, p. 94).

In the methodological literature of social sciences, generalization is considered to be specific to quantitative research: 'often research that utilizes larger samples of data is associated with greater degrees of generalizability. Research that uses large-scale data bases and statistical approaches to analyses (typically referred to as "quantitative") is associated with higher levels of generalizability than research that utilizes interpretive approaches with small samples (typically referred to as "qualitative")' (Ercikan, 2009, p. 211).

Qualitativists seem reconciled with the situation: unlike quantitative research, 'qualitative research, with its reputation for small-scale, researcher-dependent and discovery-oriented inquiries, is said to be good for providing detailed descriptions, identifying relevant factors, and generating plausible hypotheses for more systematic study. It is not supposed to be good for developing propositions, models or theories that generalize' (Eisenhart, 2009, p. 52).

They decided that generalization cannot be made in their studies and that it is useless to worry about it: 'researchers, including qualitative researchers, often say that generalization from qualitative inquiry is inappropriate or unwarranted' (Eisenhart, 2009, p. 51). Therefore, the issue of generalization is not mentioned in the chapter with results of qualitative research. Payne and Williams (2005) point out that 'while this is methodologically inadequate, it has not prevented the publication of more than a dozen articles with this deficiency in *Sociology*, i.e. a leading journal, during 2003. On these grounds, avoiding the question is apparently a legitimate practice under contemporary canons of academic publishing in sociology' (p. 299).

Although they avoid openly tackling the issue of generalization or they maintain it cannot be made, as it was mentioned in the introduction, many of the results of qualitative research are generalizations. The authors mentioned above carried out an analysis of the articles published in a year in 'Sociology' and showed that 'sociologists, using a variety of qualitative methods, have in fact actively engaged in generalization, but have done so without discussing the reach of their conclusions' (p. 309). Onwuegbuzie and Leech (2010), in their turn, have analysed all the articles that presented qualitative research published over time in the 'Qualitative Report' journal and they discovered that 'a significant proportion (i.e., 29.6%) of studies involved generalizations beyond the underlying sample that were inappropriately made by the author(s)' (p. 890).

Therefore, articles and research reports do not quite specify that results are valid *only* for the 15 interviewed people, *only* for the ten people under observation, or *only* for the 20 analysed documents. Moreover, as the concept of research universe also disappeared from the vocabulary of qualitativists, the population to which the generalization might be applied is quite hard to identify and delineate. One cannot tell whether they talk, for instance, of all workers in a department or a factory, of all workers in a field, in a country or on the entire planet. It is even worse that sometimes social policies related to health, work, gender, or ethnicity are made on the basis of such illegitimate generalizations.

Thus, the problem is that we carry out too much qualitative research without the value and usefulness given by generalization, that we should try to make the results of qualitative research generalizable in effect, declared and justified, following a clear procedure. Before describing this procedure, I will list in the following paragraph a few types of

generalization defined in literature in order to understand to which I refer in this article.

3. Types of generalization

Although most qualitative researchers do not consider generalization a problem of interest, there are a few who have turned it into a serious preoccupation. They place the beginning of the interest in generalization in Znaniecki, around 1934, who proposed to inductively develop and generalize causal theories from data through 'analytical induction' (an alternative to 'enumerative induction', applied when testing the predictions of an existing theory with new data, see Maxwell, 2021).

Over time, other authors have similarly talked about two main strategies for generalization in social research that they called statistical or empirical generalization and analytic or theoretical generalization (see also Maxwell). Today, most researchers concerned with generalization speak of three distinct variants: '(a) extrapolation from sample to population; (b) analytic generalization; and (c) case-to-case translation' (Ercikan, 2009, p. 212). These three types of generalization are associated with three types of research: the first, 'with research using representative random samples, analytic generalization with experimental and quasi-experimental methods, and case-to-case translation with qualitative methods' (idem.).

The most respected of these currently conveyed variants of generalization is the first one: 'the strongest argument for generalizing is usually thought to be extrapolation from a sample to a population' (Firestone, 1993, p. 16). It is associated with statistical generalization facilitated by probabilistic sampling, but although it is highly rated, there are researchers (see Gobo, 2004 or Ercikan, 2009) who succeeded to prove, with arguments, its limits.

On the other hand, analytical generalization does not refer so much to population and the sampling procedure: 'generalizing to a theory is different from generalizing to a population' (Firestone, 1993, p. 17), it refers to generalizing a particular set of results to a broader theory: 'to generalize to a theory is to provide evidence that supports (but does not definitively prove) that theory' (idem). Judging by the way it is described by Firestone, analytical generalization seems to be made at the end of research, when the end product is compared to existing theories and brought as a proof that theories are fit also for other contexts than those in which they were generated.

Regarding the third possibility (case-to-case translation), this coincides with what in the literature of qualitative methodology is called transferability. It 'involves providing logical arguments to support claims that findings in one particular research setting may apply to another setting (case)' (Ercikan, 2009, p. 213), it 'implies that you need to provide enough information about meanings, contexts, and processes operating in your study setting or population that a reader can adequately judge the likelihood that your findings would apply to a different specific setting, group, or population' (Maxwell, 2021, p. 115). We can note this type of generalization is no longer the task of the researcher but that of the public in front of whom he/she presents his/her results. That is, others decide if the results of your research can be or cannot be generalized.

For most qualitativists the subject matter of generalization supposes only the reference to the third variant –

transferability, which is not even likely to be up to them. But the subject should not close here. There are in literature at least two other classifications of generalization that allow (re) opening the subject matter of a stronger generalization in qualitative research than transferability. I refer to these classifications also to define and delineate with precision the generalization addressed in this article. The first classification is that of Maxwell (2021). He proposes to distinguish between internal and external generalization: 'internal generalization refers to generalizing *within* the setting, group, or population studied, to persons, events, and activities that are not directly represented in the data collected; […] external generalization, in contrast, refers to generalization *beyond* the person(s), setting, case, or time specifically studied, to other persons, settings, cases, or times' (p. 112).

Maxwell and Chmiel (2014) show that similar distinctions have also been made by other authors, for instance between lower-order generalizability and higher-order generalizability: 'lower-order generalizability is the generalizability of findings within the unit of analysis; higher-order generalizability is generalizability of findings across units of analysis of the same type (e.g., across similar organizations or neighbourhoods)' (p. 541).

In terms of the classification proposed by Maxwell, we can say that extrapolation from sample to population is an internal generalization (or lower-order) and analytical generalization and transferability are external generalizations (or higher-order). As the title of this article shows, I am interested here only in internal generalization, that is the possibility of extending the results of the research from the level of those interviewed or observed (sample), to the level of research universe (population), or as Maxwell (2021) pointed out, 'from the specific individuals, events, or activities on which data are collected, to the group, institution, or case that the author states is being studied', p. 112).

The same author specifies that, for internal generalization, the term sampling is essential (but he underlines that it refers to a broad meaning of sampling, not to the statistically limited one): 'for internal generalization, the concept of "sampling," in a broad sense that is not limited to statistical methods, is important for making claims that your results are generalizable to the actual setting, group, or population you are studying' (p. 115). Therefore, in this article, I will present a variant of sampling (not statistical!) in qualitative research that allows for an internal generalization of findings. Thus, I start from the idea of Maxwell that 'internal generalizations are not necessarily statistical' (p. 112).

The second important classification (for this article) of generalization is that pertaining to Smaling (2003). He distinguishes, depending on the type of applied reasoning, between inductive generalization and analogical generalization. The author lists three forms of inductive generalization: 'statistical generalization, variation-based generalization and theory-carried generalization' (p.2). Smaling indicates that each of these three types of generalization 'is concerned with drawing conclusions pertaining to (a class of) subjects, cases, situations, etc., that have not been researched, based on a limited number of persons, cases, situations, etc., that have been researched. In all those cases, a generalization is made from research results to a population or to a scope belonging to a theory. In this sense, these forms of generalization can be seen as forms of inductive generalization: methods of arriving at general or universal propositions' (p. 8). The author also describes three types of analogical generalization on which I do not insist any longer because they are not subject to this article. I refer here only to internal generalization that can be performed on the basis of inductive reasoning.

As mentioned above, three types of generalization based on an inductive reasoning are described in literature. Alongside

the statistical probabilistic generalization, there are also two types of inductive generalizations that allow the extend of research results at the level of the entire research universe (i.e. variation-based generalization and theory-carried generalization). Eisenhart (2009) appreciates that ‘many people who say that qualitative research is not generalizable seem to define the concept in probabilistic terms, i.e., as a procedure for making general claims about a population from a sample, based on statistical probabilities’ (p. 52). But Smaling (2003) argues that ‘statistical generalization is not always the best form of generalization. There are alternatives that can be better, certainly in situations in which little is known about the population’ (p. 25).

Mitchell (1983) explains these alternatives by the existence of two inferential processes – statistical inference and logical or scientific inference – that allow for the generalization of results from sample to population: the rationale of extrapolation from sample to parent universe ‘involves two very different and even unconnected inferential processes — that of statistical inference which makes a statement about the confidence we may have that the surface relationships observed in our sample will in fact occur in the parent population, and that of logical or scientific inference which makes a statement about the confidence we may have that the theoretically necessary or logical connection among the features observed in the sample pertain also to the parent population’ (p. 207).

There are authors who hold even the superiority of logical inference to statistical inference – see Znaniecki (apud Gobo, 2008) who describes the process as being ‘the true method of science and it is the superior method (because it discovers the causal relations of a phenomenon rather than only the probabilistic ones of co-occurrence)’ (p. 195).

Thus, in this article, I am interested in describing the way in which theoretical inductive generalization can be performed (variation-based and theory-carried) on the basis of logical or scientific inference. Smaling argues that variation-based generalization supposes the covering of the variation and that this can be performed when ‘it is more important to display the variation than to know how often a certain variant occurs’ (p. 25). Regarding theory-carried generalization, the author makes a distinction between subsumptive generalization and abductive generalization for the situations in which what guides generalization is an existing theory or one that is grounded (discovered) on this occasion. It is a matter of the same distinction made by Eisenhart (2009) between what she calls theoretical and grounded generalization where ‘the goal of grounded generalization is to produce new theories or explanations whereas the goal of theoretical generalization is to make existing theories more refined and incisive’ (p. 60).

In summary, internal generalization – to which this article refers – or sample to population is performed based on inductive reasoning, so it is an inductive generalization. Inductive generalization can be statistical generalization (that being performed with probabilistic statistical sampling) or theoretical generalization (that being performed with theoretical sampling – more details about the types of sampling in the following paragraph). Theoretical generalization can be guided by variation or by theory, and the latter can be guided by an existing theory (subsumptive generalization) or a new theory (abductive generalization).

Although I have found a description in literature of the possibility of performing generalizations from sample to population otherwise than statistically, although these generalizations also bear a name (variation-based generalization and theory-

carried generalization), I have not found instructions regarding the way in which they could be performed. Moreover, Maxwell and Chmiel (2014, p. 549) sustain that in qualitative research, 'generalization cannot be guaranteed by mechanically applying a particular strategy as an algorithm or procedure'. In this article, I will try to challenge this conclusion by proposing a procedure. Before presenting the procedure, because internal generalization needs sampling, I will first refer briefly to this subject matter.

4. Types of sampling

In qualitative or mixed methodological literature, there are no clear typologies (i.e. which specify exactly the criteria by which the elements subject to classification are divided by categories) related to sampling and I would like to propose one.

We can perform two types of sampling in social research: statistical sampling and theoretical sampling. In statistical sampling, we know before entering the field how many subjects there will be in the sample and what characteristics they will have. Their number is set depending on dispersion, probability, and margin of error or depending on the minimum required of subjects in order to operate various statistical methods or tests on the data collected (e.g. F, t, χ^2, z, etc.) and the selection of subjects is made according to strict sampling rules.

Statistical sampling can be probabilistic or non-probabilistic and only the first allows the generalization of results from sample to population. There are all sorts of statistical sampling techniques and all sorts of combinations between them. For instance, statistical probabilistic sampling can be simple random, systematic, stratified or multistage and non-probabilistic statistical sampling can be quota. What these statistical techniques have in common is that they relate to the population: size, heterogeneity and its socio-demographic characteristics. Representativity in these cases is statistical, it is calculated, the sampling being performed so that 'the sample is more likely to be representative of the population from which it is drawn' (Bloor and Wood, 2006, p. 93).

We apply statistical sampling especially when we have to study large, heterogeneous populations, when we have information on their number and socio-demographic characteristics, when we collect data with structured or standardized techniques (because we can apply them to a large number of subjects), when we pursue the testing of hypotheses on populations with a certain profile and so on. In this situations, we are interested in knowing how often a certain variant occurs in a sample (and in case the sampling is probabilistic, how often a certain variant occurs in a population), we are interested especially in testing theories.

Collecting data accidentally – without the control of the number and features of respondents is not a statistical sampling (not even of the non-probabilistic type). No matter how used in quantitative research haphazard or convenience techniques are, they do not generate results with too great a scientific value. The studies in which we use such sampling could be considered explorations (but not too rigorous!), and their results could be used exclusively to set objectives or formulate hypotheses to address on other, less exploratory occasions. These sampling techniques should be avoided or used in exceptional situations and the rule should be the attempt to withdraw subjectivity from the process of statistical sampling.

Regarding theoretical sampling, the selection is made within it by the relevance of cases without knowing before entering the field either how many participants will be necessary, nor all the characteristics they will have. Processing, analysing, and interpreting data will show what sort of participants need to be further selected, and saturation will show at what moment the selection of new participants can be stopped.

Theoretical sampling can be guided by variation or theory, and these two types of sampling allow just like the probabilistic sampling techniques, generalization from sample to population. Then, there are all sorts of other theoretical sampling of the purposeful type, but they do not allow for internal generalizations. What theoretical techniques have in common is that they relate to data and the way in which they vary, to theory and the way in which it is grounded in data. In these cases, representativity is theoretical. Yes, this term is a novelty; the representativity is related, in literature, exclusively to probabilistic statistical sampling but, in the approach presented here, there is representativity also without probabilistic sample (or as Gobo (2008, p. 201) puts it: 'random and representative are neither synonymous nor necessarily interrelated terms'). Thus, although it cannot be calculated, theoretical representativity can be obtained (through a process described in the following paragraph) and generalizations can be made based on it: 'representative samples are not predicted in advance but found, constructed and discovered gradually in the field' (Gobo, 2004, p. 448).

We apply theoretical sampling especially when we have smaller and more homogeneous populations, when we collect data with unstructured or non-standardized techniques (because we can apply them only to a small number of participants), when we do not collect all data at once, at the beginning of research. In this situation, we are interested especially in the data variations and not how often they occur; we are interested in completing or discovering theories.

Collecting data accidentally – without guiding after processing, analysing and interpreting data, is not theoretical sampling. No matter how common the snowball or convenience techniques are in qualitative research, they do not generate results with a strong scientific value. The studies in which these sampling techniques are applied could be considered explorations (but not too rigorous!), and the results obtained could be used exclusively to establish objectives of formulate hypotheses to address on other, less exploratory occasions. These sampling techniques should be avoided or used in exceptional situations and the rule should be the attempt to withdraw subjectivity from the theoretical sampling process. I have deliberately repeated here what I have written above for the statistical sampling in order to underline that studies that do not pursue generalization are the same from the point of view of rigour, utility, or value, regardless the type of sampling used and irrespective of whether they are quantitative or qualitative.

Social research should pay much more attention to sampling. The greatest problems of social research are given neither by the instruments with which we work, nor by their application (even if there are many problems here also), but by the sampling procedure. In most studies, it is the most hidden stage of research, it is the least justified, the worst performed, it is a sort of Achilles' heel. Most research we perform offers results valid only for those studied and we should try to apply sampling procedures that allow us to generalize findings (be it statistical probabilistic sampling, or theoretical sampling guided by variation or by theory) more.

I will describe theoretical sampling that allows for the generalization of results from sample to population in the following

paragraph. I would like to specify beforehand that I did not refer in this paragraph to the variant of studying entire populations, where it is no longer a matter of sampling. The variant is perfect where and when it can be used. There are enough research situations (especially applicative) in which this can be done and it is good to do so. For instance, when studying the workers in a department, the management staff in a company, the children in a kindergarten, the medical staff in a hospital etc. If the populations of interest are small and we can interview, observe etc. all of them, it is good to do this because, thus, we eliminate the problem of sampling and incurring errors.

Also, I have not discussed here the variant of case studies - a person, a profession, an institution etc. Choosing the cases by the typical case, extreme case principle etc. and the discussion related to when and to what serves the studying of these cases will be subject to other reflections.

In this article, I only refer to qualitative research with multiple units of analysis in which only a part of the elements of the research universe are studied (so, to which sampling is applied). The expression multiple units of analysis disappeared also from the vocabulary of qualitative researchers; even if a large amount of their research is not case studies, they avoid using the expression multiple units of analysis in order to define the research with more participants and they have not invented another one either.

5. From sample to population generalization in qualitative research – procedure and its illustration

Before presenting the procedure, I would like to come back a little to the arguments of those who consider that generalization is not possible in qualitative research. The reason, often raised, for stating that one cannot generalize is the small number of interviewees or observed persons in such studies. It is true (and this cannot be changed) that qualitative methods and techniques cannot be applied on a large number of people. But Gobo (2008) adds that result generalization in qualitative research is ‘doubted not only because they are derived from only a few cases, but also because even where a larger number is studied these are generally selected without observing the rigorous criteria of statistical sampling theory’ (p. 193). Therefore, the problem has also the dimension of the way in which we choose participants - and something could be changed here. The theoretical sampling procedure that allows for generalization in qualitative research (described below) focuses precisely on this aspect – the criteria for choosing the participants.

There are also other authors who suggest that the problem of generalization must not be reduced to the size of the sample – see Ercikan & Roth (2006) who argued that ‘generalizability of research findings cannot be judged based on sample size’. What matters is the identification of the types of participants who generate data variety and their insertion in the sample: the variance - ‘this is the only very important rule for selecting the sample’ (Gobo, 2004, p. 444).

Therefore, we are interested in the variation of the studied phenomena, regardless of how many participants generate a variation or another: ‘the aim is to observe extensively the relations between variables, not only to assess (which is always a quite problematic task) the number of persons who feature one characteristic. Therefore, generalizability is mainly a practically and contingent outcome related to the variance of the research topic’ (Gobo, 2004, p. 453). Thus, we are

interested in who brings what variations to the collected data, not necessarily how many.

Illustration: *If I want to find out what the dissatisfactions of students in a faculty are, I look for the variation of these dissatisfactions – what dissatisfactions there are, how those who have them are, how they vary depending on the characteristics of the students and not how many have one or the other.*

The author mentioned above argues that 'it is a function of the invariance (regularities) of the phenomenon, not a standard or automatic algorithm of a statistical rule' (idem.) and that this approach has already been applied to many social studies, which, although carried out on a small number of cases, have generated results that 'have always been considered generalizable' (see the vast majority of sociological theories, from Max Weber and before, to Erving Goffman and after).

Illustration: *Let us suppose a qualitative research situation. I would like to find out, in detail, what dissatisfies the students of a faculty of social sciences (say around 600, equally distributed by specializations – Sociology, Human Resources, and Social Work and by years of study – from one to three). I decided to take some interviews. Is it possible that by interviewing only some of them (say 20, 30, or 40), to generalize the results to all 600?*

Any scientific study has to specify to which population it refers; therefore, a research universe should always be established (i.e. it should be defined a population of interest). Then, if it cannot be studied entirely, a sampling procedure should be established and how many subjects will enter (or entered) the sample should be specified and also what characteristics they will have (or had).

In qualitative studies, the research universe is not very large, nobody proposes carrying out research on the entire population of a country, on all unemployed people in a geographical region and so on. Most times, a small community is studied, the members of an organization, generally local phenomena, at a micro level. This aspect is important: the research universe in qualitative studies should be as small as possible and the population to which research refers should be as homogeneous as possible (I will come back to this statement in the following paragraph where I will question this idea, attempting to open the subject matter of generalization). I make this recommendation because large and heterogeneous populations usually generate larger data variations, and following these variations would suppose, as I will show below, collecting new and new data, more than one could collect in a piece of qualitative research (the unstandardized character of data collecting techniques does not allow us to apply them to a very large number of participants).

A short aside: we could ask ourselves if the research universe has to be as small as possible, which is (how big is) the stake of generalization in qualitative research? Generalization, no matter how frail (to 100 or 1,000 people) is an important stake. It is one thing to say 'the dissatisfactions of the 30 interviewed people in the company are…' and another thing to say 'the dissatisfactions of the employees in the company are…'. When we study at a micro level, we do not need more than 'micro generalization'.

Regarding sampling, in qualitative research statistical sampling procedures are not chosen very often and, as showed in paragraph four (concerning types of sampling), theoretical sampling is different to statistical by the fact it cannot specify,

before data collection and analysis, the number and characteristics of participants in the study.

Therefore, **step one** in the theoretical sampling procedure that allows for an internal generalization is establishing a research universe as small and as homogeneous as possible. But as there is flexibility in choosing participants, there also is flexibility in establishing the research universe. It can be modified, during the research, depending on what the processing, analysis, and interpreting of the collected data reveal. Therefore, if the studied phenomenon proves to be complex, broad, with high, significant variations, then it is possible that the research universe initially established to a larger population, be adjusted and reduced to a much smaller population. Of course, adjustment is also possible the other way round. If we initially proposed a small research universe, and the studied phenomenon proves to be less complex than we thought, if we also could follow other possible data variations, we can extend the research universe to larger and larger populations. The second variant of adjustment is preferable.

Illustration: *I chose a fairly small (600 people) and homogeneous (students from the same faculty, in similar specializations) research universe. If I had chosen all students in a country as the research universe, data processing and analysis would have probably revealed high variations between students' dissatisfactions from different universities (from different regions, of different sizes etc.), from different faculties, from different specializations etc. Because I could not have followed all these variations (I should have needed to perform more interviews than possible), I would have probably been obliged to successively reduce the research universe – not all students in the country, but only those from X university, not all students from X university, but only those from the socio-human domain etc. I would have reduced it to how many data variations I could follow.*

Likewise, if I had proposed to study only students from one faculty, and the analysis revealed few, invariable dissatisfactions, it is possible, if I want, to add students from other faculties to the sample. In this case, I will enlarge the research universe, and finally, it will be made up of students of as many faculties (or even universities) I could have followed within the research.

Step two of the procedure is the selection of the first participants in the study. Who are we going to interview or observe first, who are we going to ask to produce social documents? How do we choose them? By which criteria? We should temporarily establish, for each separate piece of research, depending on research objectives, who the most appropriate first participants are. The criterion that guides us is – what persons from the research universe possess the information we need and we take care that the first chosen are as diverse as possible, as different as possible. Depending on what we study, we ask ourselves what kind of participants could bring numerous, complete, and especially diverse information, maybe even contradictory. We try to identify information-rich persons and their maximum variation. Larsson (2009) argues that 'covering more of the variation in qualitative different views will enhance the generalizability of the study. […] In order to maximize the differences, a sample should be based on qualified guesses about how to achieve this broad variation. [...] This means that the uncommon case is as important as the most common kind of case. [...] If a wise selection of persons or cases, which could be expected to be diverse, has been made, one could expect to have covered the variation relatively well' (p. 31).

If we have many initial criteria of selection to ensure the diversity of opinions, it would be good that they are found, at

once, in the same participants. Otherwise, we will have to interview from the beginning more participants than possible. For instance, if the criteria gender, age, income, education, place of birth and many others could be relevant (to ensure the diversity of the collected data), it would mean performing at least one or two interviews with representatives from each category (two women and two men, two youths, two adults, and two elderly persons etc.). The combination of criteria in the same participants would mean – a young, poor man with primary school studies, from the rural area, an old, rich, woman with higher studies, from the urban area etc. This is why it is preferable that the research universe be homogeneous. The more heterogeneous it is, the more criteria that generate probable variations there are and the more subjects we will have to select from the beginning of data collection.

We should be very attentive to the combination of sampling criteria. The combination of several criteria in the same participants will harden the establishing of the causes for data variations. For instance, if we note significant differences between the data collected from the young, poor man, and from the old, rich woman respectively, we could not know which of these demographic characteristics generated the data difference: gender, age, income, education, place of birth? Therefore, the more the initial criteria of selection are considered relevant, the greater the number of those from whom we collect data initially should be. An initial number of participants, double compared to the selection criteria, should be sufficient for a start (three criteria – six participants, ten criteria – 20 participants).

Illustration: *In the example with the students, I asked myself on what characteristics their dissatisfactions could vary and I decided it could be gender, age, specialization, year of study and whether they pay tuition or not. This means I should have in the initial sample male as well as female students, in the first, second and third year of study, from Sociology, Human Resources as well as Social Work etc., in different combinations. Around ten students could be interviewed, initially.*

Step three is the most hard-working and it supposes attention to the processing, analysis and interpretation of data collected both initially and subsequently. This process will indicate what sort of participants should be further selected and how many. We will see whether the sampling criteria proposed initially as relevant have given data variations or not. If they gave variations, they will be preserved as sampling criteria and will be used in selecting new subjects from which data will be collected until reaching saturation (on these sampling directions). If there are sampling criteria initially supposed as relevant that did not give data variations, we give up on them. If data variations occur which are unexplained by initial sampling criteria, then these variations should be followed through with a new sampling. This means that first, we have to identify the supposed causes for variations. They can be looked for in criteria of which we have not thought initially (the easiest is to guide ourselves, in the beginning, by the classical socio-demographics) – for instance, we consider whether the participants have children or not, if they work in the private system or for the state etc.? Practically, we have to try to realize why some participants answered or behaved differently to the others. This sampling stage is easier to perform (successfully) the more familiar we are with participants. If we know them well, it is easier to understand how some are different and how others resemble each other. We can rapidly identify, in these situations of familiarity, differentiation criteria less obvious between participants, such as those related to how sociable, serious, optimistic, generous etc. they are.

Therefore, we should be flexible regarding sampling criteria, we should give up easily on criteria initially considered important and we should adopt new ones that we would never have thought of.

If we think a possible explanation for variation has been identified, it becomes a sampling criterion. As we gather data from these new subjects, we follow if the variation persists for the newly collected data and whether the supposed explanation is confirmed. If it is confirmed, we sample by this criterion until we reach saturation. If it is not confirmed, we look for another explanation that will become a new sampling criterion and we re-do the procedure until we identify an explanation of variation to be confirmed on the field.

If we cannot explain variation, we report it as such. It could be an error (generated by the data collection context, by the dishonesty of a participant, by the reactivity of the applied methods etc.) or a future study direction (in other contexts).

Step three is very important. The possibility to generalize and the volume and characteristics of the population to which we can extend generalization depends on how well we perform step three. We will have to look for and include in the sample typical participants, who bring rich information (illustration: *the conscientious student who always comes to school, the careless student who does not attend courses very often etc.*), but also possible extreme, deviant cases (illustration: *the recalcitrant student, the student who makes complaints all the time, the student who repeated the year etc.*). It is very important to ensure that we collected data from as many people from the research universe as possible who can generate data variation and to permanently compare newly collected data to that already existing, and if variations occur (new themes, new categories, contradictions etc.), we should follow these variations, we should sample on their direction and confirm them.

In this process, we can use our knowledge related to people in the research universe – who better know the phenomenon studied, are more interested in it, who is special, different, relates differently to the studied phenomenon etc.

Illustration: *I remember a situation in which participants in my study were my colleagues who I knew very well. I asked them to produce a social document and I analysed this data as they provided it. Data processing revealed interesting patterns, I formulated bold results etc. When I compared the characteristics of those from whom I had already received data, I was startled – they were my colleagues, young girls, at the beginning of their career and I was panic-stricken when I realized I will receive more data from older, male colleagues, at the end of their career. I was imagining how all the patterns already discovered would fall apart, how all temporary results would be overturned etc.*

We should refrain our natural propensity to look for data confirmations at any price and deliberately look for participants who are different, who could act and think differently, who could complete or even contradict our results. We should use what we know about subjects in order to select those who make our results richer, more complex, more secure, ‘more generalizable’.

We can also appeal in this process to people who better know the phenomenon studied and the research universe, we can ask them to help us in choosing the most different participants. We can even ask the participants to help us (a sort of snowball sampling) to identify other participants who could have more information about the phenomenon, but especially

those who could have different opinions. But the most important aspect is to follow the newly collected data, their matching and especially inconsistency in the already formed patterns, other and other pattern formation to a full description of the phenomenon, covering as many variations as possible.

Illustration: *In the example with the students, step three looks like this: after I interviewed ten very different students and I compared the data, I noted that the dissatisfactions of the girls are not different from those of the boys. In the initial sample, I had six female students and four male students and I could note that their answers did not differ significantly. Because it did not prove to be a relevant criterion, from now on I will not use it in sampling (the gender of the next interviewees does not matter). But let us suppose that important differences occur between dissatisfactions depending on the study year (students in the first years seem to have bigger and more diverse dissatisfactions) and depending on the specialization (those from Human Resources seem to have more dissatisfactions, related rather to the curriculum, those from Social Work seem to have fewer dissatisfactions, related rather to administrative aspects etc.). Because I had at least two participants in the same study year and from the same specialization, I could note those possible patterns. These criteria stay in the game, I continue sampling by them, interviewing students from different years and different specializations in order to see if the differences already noted are maintained and to describe them completely, until saturation (saturation by the criteria year of study and specialization!).*

But let us suppose that we also found a different case among those studied. All interviewees described various dissatisfactions, only one chose to say that everything is not right, that the faculty is a complete waste of time and money. In this situation, I should realize why this participant is different, through what he/she resembles (to eliminate these criteria) and how he/she is different (as possible explanations for variation) from the other students of the faculty. It is not a matter of gender because many other female students did not answer like this; it is not specific to first year because many other students in the first year answered differently. What could it be? I look for criteria that I did not initially take into account: would he/she be a brilliant student or very slow to learn? Would he/she be among those who always come to courses or among those who are absent? I look into the registers, on the attendance lists; if I do not know him/her very well, I can ask professors or his/her colleagues if there is something that makes him/her different. Let us suppose I discover that, unlike the other interviewees, he/she is the only one who works during the faculty period. This could be an explanation, this becomes a criterion for sampling, therefore I seek to interview also other students who work. If also the freshly interviewed believe that school does not help them with anything, that nothing of what school offers them is useful for them in the workplace, then it is highly probable that a new important variation of the studied phenomenon be identified – the students who work have another configuration of the dissatisfactions related to school compared to those who do not work. This sampling criterion will be used until saturation is reached, that is until I succeed in describing completely the dissatisfactions of the category of students who work while attending university. I should do the same with any new variation that occurs in the newly collected and analysed data.

If I am the professor or the colleague of students within the research universe, I think which of them could give me more information about dissatisfactions, which of them I heard talking about this subject matter, which of them speak more openly about subjects related to school, which are more special, more critical etc. I ask other professors or students or even those who were already interviewed if they know other students who could talk to me about dissatisfactions, if there

are others who could have different opinions, thus, I try to cover as far as possible the subject matter of dissatisfactions related to the faculty they attend, by interviewing various students and by following carefully each new, different thing that they tell me.

If we have followed all data variations, we should be ready for**step four** – that is result generalization. The study of the phenomenon is completed when we have discovered data variations, when we have followed them by sampling, guiding ourselves by them until reaching saturation on each new sampling criterion. It is possible that complex phenomena could not be studied entirely. Variations of phenomena that we did not identify could exist, some that we can no longer follow could exist because they would imply too much extra work etc. What is important is to cover as much as possible of the studied phenomenon and especially, to specify exactly how much we covered, that is for whom the results are valid. We should be able to tell exactly, at the end of the research – what the universe of the studied population is and what the characteristics of those to whom the results are applied are (that is, the sampling criteria used). By applying the described procedure, it is possible to state that the results can be generalized to the entire population in the research universe, who possesses the characteristics of those who were part of the sample.

As long as we correctly apply the principle of saturation on each sampling criterion that generated data variations, generalization is possible. It is necessary to bring arguments in favour of saturation. It is a known principle, and it is used in all qualitative research that respects itself. Generalization can be made because opinions, behaviours, attitudes, feelings, and people's experiences are much more similar than we imagine.

Those who carried out qualitative research have certainly passed through the surprise (repeated in every study) of noting almost weird similarities in the data collected from people who do not know each other, who did not live in the same environment etc. Of course, we cannot rely on feelings and on a few personal researcher experiences when we decide that we can generalize, but there are also experiments that show after how many interviews saturation occurs, and especially, that show how insignificant the variation noted in the obtained data is, by carrying out further interviews, after reaching saturation (see Guest et al., 2006 or Hennink et al., 2017). Guest and the co-authors, for instance, argue that 'saturation occurred within the first twelve interviews, although basic elements for meta themes were present as early as six interviews. Variability within the data followed similar patterns' (p. 59). See similar findings regarding reaching saturation in group interview studies in Guest et al., 2017 or Hennink et al., 2019.

Therefore, in step four, depending on how complex the studied phenomenon proved to be and depending on how many of its variations we could follow, we will establish the final universe of the research and the sampling criteria that we applied and for which we can guarantee results generalization.

Illustration: *In the example with the students, I will say at the end of my study that the research universe remains as proposed initially because I did not discover in the interviews so many variations of the dissatisfactions so that I could not follow them, because the studied phenomenon was not so complex as to require the reduction of research universe. Then, I could say that dissatisfactions identified are valid for all students (600), I could guarantee this for the following applied sampling criteria: gender, age, specialization, year of study, whether they pay tuition or not and whether they work*

or not during the university period. It is possible to have reached 20, 30 or 40 interviewed students by the sampling procedure described above and I am very sure that if I would interview also the remainder (up to 600 students), I would not discover too many other important variations of dissatisfactions.

The merit the procedure proposed in this paragraph has is that sampling becomes a transparent process in which the applied sampling criteria are shown exactly, and the verification will be done easily. The applied sampling criteria reflect our trial to discover more of the variations of the studied phenomenon. The more sampling criteria we follow, the more probable it is that results are valid for all people in the research universe. The more we ensure reaching saturation by several sampling criteria, the more probable it is to ensure result generalization. And if, subsequent to the research, we discover that the results are not valid for all subjects in the research universe, these differences should be specific to categories of participants who were not included in the sample, the differences should be put on the account of other characteristics than those followed in the research – *for instance, there could be isolated cases of students coming from another culture, discriminated against, sick, with special needs etc.*

Here is the clarification of another problem: how it works when we have group interviews, and not individual interviews? In the case of applying group interviews, we proceed in the same manner, establishing the research universe and the first participants. They are grouped (depending on characteristics) in order to interview them together and then, we follow variations between and within the groups. The variations guide sampling – the realization of new groups by the sampling criteria that generate data variations and collecting new data until reaching saturation by groups.

Illustration: *For instance, if I want to find out what the dissatisfactions of the employees of a company are, I choose to perform group interviews – one with workers, one with administrative personnel and one with management. If I note differences between the groups, it means that this sampling criteria used generates variation and I will perform as many group interviews with workers as needed until I reach saturation (equally with the administrative personnel and management). If within the group of workers, I note variations in dissatisfactions, I look for possible explanations. Say it is possible that they differ depending on the department in which they work. In this case, I will perform one more group interview with workers in department A, one with workers in department B etc. to confirm the difference. If confirmed, I perform as many interviews with workers in department A (and B, and so on) until saturation. I do the same with any new variation that occurs.*

The procedure presented in this paragraph can be applied to most qualitative research because most of them are exploratory or descriptive, because most of them apply to the level of small populations. It was described a theoretical sampling, guided by variation, which facilitates obtaining a theoretical representativity of the sample and which allows a theoretical generalization of the results (from sample to population generalization, of the variation-based type). I defined and classified in the previous paragraphs both theoretical sampling and theoretical generalization and I introduced there this theoretical representativity too, for which, in order to be obtained, 'the sampling plan must be set in dialogue with field incidents, contingencies, and discoveries' (Gobo, 2008, p. 207).

This possibility of generalizing is described by Smaling (2003): 'when conducting descriptive, explorative, or purely practical research, the researcher may strive for some form of generalization of the research results by focussing on

variations in which a phenomenon occurs. For this purpose, the researcher intentionally looks for cases that deviate from those already researched and even looks for extreme cases' (p. 7). The author uses 'descriptive saturation' to name the saturation reached through this procedure in order to distinguish it from 'theoretical saturation' (more details about the implications of this distinction below). We could name what was described here, following the same distinction, 'descriptive sampling', and not theoretical sampling, as Smaling does, but we would complicate things unnecessarily (see the explanation below).

Smaling continues the description of the process like this: 'the assumption then becomes that the variation occurring in the population has been sufficiently represented by the variation in the sample. The researcher does not have the intention of counting how often something occurs. The purpose is not to cover statistical distribution but merely to cover the existing variation. The sample is not representative in a statistical manner, but rather it represents the variation. The researcher can aim at generalization based on covering the variation. […] In other words, the researcher can aim at variation-based generalization' (idem.). So, it is precisely the procedure described in this paragraph.

As I have shown in the paragraph referring to generalization, beside variation-based generalization, we can talk about theory-carried generalization. The distinction refers more to the type of research we are in and less to the implemented procedure. In simple, descriptive or exploratory research, we seek to discover variations, data patterns, to reach descriptive saturation (as Smaling says) in order to make generalizations based on variation. In more complex, explanatory research, we seek to ground (or complete, reformulate) theories, to reach theoretical saturation in order to make theory-carried generalizations. But descriptive sentences generated in simple studies are still theoretical sentences and it would become unnecessarily complicated if we went too far with these distinctions. We could talk about theoretical saturation, theoretical sampling and theoretical generalization irrespective of whether the study proposes to ground theories (descriptive and explanatory statements) or only the discovery of categories or themes (descriptive statements). The procedure that leads to generalization is the same, the one described above even if, for the grounding of theories, things are a little more complicated.

Thus, in the same way, in the case of theory-carried generalization, 'the criterion for selecting cases from which one will generalize is not random or representative sampling but the extent to which the cases selected are likely to establish, refine, or refute a theory' (Eisenhart, 2009, p. 60). Generalization is based on the confidence we may have that the theoretically necessary or logical connection among the features observed in the sample pertain also to the parent population and the validity of the extrapolation depends not only on the typicality or representativeness (as in the inductive statistical generalization), but 'upon the cogency of the theoretical reasoning' (Mitchell, 1983, p. 207).

The generalization procedure in the theory-carried case is more complicated in step three because we talk not only of descriptive sentences, but also of explanatory, because we will have to confirm and look for falsification not only for patterns, but also for connections between phenomena, we will have to saturate not only categories, themes, but also theories, connections between categories or themes. In fact, step three in this situation is what we find described in literature as analytical induction in the grounded theory studies (between brackets, there are more and more authors who, rightfully, state that here it would be in fact a matter of abductive reasoning, not induction, but this is an issue that will not

be addressed here).

Therefore, I do not exemplify the theory-carried generalization process too, because this is similar to that already described as variation-based generalization, and the differences between the two processes (that appear in step three) are described in detail in literature. I chose to describe generalization based on variation because most qualitative studies are descriptive, and the few who perform studies of the grounded theory type have sufficient details regarding the procedure, in the materials devoted to this approach.

6. Short discussion

I recommended in the procedure described in the previous paragraph that the research universe in qualitative studies be as small as possible, and the population to which the research refers be as homogeneous as possible and I argued why. I would like to introduce now in the discussion an opening of the generalization theme. In the case of some phenomena not precisely local, in the case of populations not precisely homogeneous, can we generalize? 'Up to which point, how much' can we generalize in qualitative research? The answer is related to the characteristics of the studied phenomena – how complex it is, how much it varies, if it depends more on individual or societal traits, if it is modelled by personal perspectives or rather by habitus (Pierre Bourdieu)?

I looked back, at part of the research I have undertaken and I asked myself how much I could have generalized then. For instance, I tried to study the experience of illness, to find out what it means to live with a terminal disease by interviewing cancer patients looked after by an NGO in Brasov (Romania). There were only 40 participants, but I am sure that their experience (their feelings, their needs, their support, their wishes etc.) is identical to that of other cancer patients in Romania. And I think I could even raise the stakes: a few months ago, I wrote a review for a journal of an article about cancer patients, written by someone in Australia and I had the feeling that it was my study – the same words of the interviewees, the same feelings etc. I believe that terminal disease is experienced in the same way, and this subject matter would allow us to generalize more than we would believe possible in qualitative research.

Things are the same also when we talk about other research themes. For instance, what those who leave Romania in order to work abroad live, even if there are various experiences – of different work, of different places etc., they have their constants: the same emotions of separation, of alienation, of the unknown, of solitude etc. And reading the articles on the same theme of the researchers abroad, I could say that these constants could be specific 'to the east working in the west' – so not only to Romanians, but also to Bulgarians, Ukrainians, Polish etc.

I give another example from my studies: I tried to discover the significances that the elderly people give to volunteering. I interviewed 22 participants and I analysed the answers of 478 subjects to open questions in a survey. They were all persons above 60 years old from Brasov County (Romania), and their answers were very similar, the conception of the elderly about volunteering being modelled significantly by their experience, as young people, with 'patriotic work' specific to the communist era. I do not believe that interviewing other elderly persons in other counties in Romania would have revealed important variations, I could have generalized the results at the level of all the elderly in Romania. Moreover,

studying European statistics showed me east-west patterns regarding the perception of volunteering, the former communist countries having extremely similar distributions of the studied variables. It is very possible that the elderly in the former communist countries operate with similar significances of volunteering.

I suggest, of course, not that I should have made generalizations on the basis of feelings, but that, if I had had time and a team to work with, we could have checked the existence of some variations of the studied phenomena on larger populations. In the situation of studies like those listed above, due to the characteristics of the themes studied, it is very possible that generalizations be made, for example, on people who suffer from a terminal disease, on the population in Eastern Europe that migrates to the West or on the elderly in the former communist countries.

I have proposed so far reflections related to common themes, specific to some large categories of people – experiencing an illness, experiencing migration, the perception of volunteering, but what happens when the research themes are very personal, specific to a single person? A few years ago, I carried out 20 interviews of the life story type, with managers in Brasov County (Romania). Although participants had no connection between them, their life stories revealed similar experiences, for instance: they lived in authoritative family environments, they were assigned various responsibilities from an early age, they earned money from when they were little and they managed it by themselves etc. The surprise was even bigger when I added to the analysis other 25 interviews, carried out by journalists with top managers in Romania. Their life stories were so similar to the first that I ask myself if we could not talk, in general, of life experiences specific to those who are managers.

I have presented so far studies in which the population was fairly homogeneous – cancer patients, elderly persons, managers, but what happens when the population is heterogeneous? Can we generalize? Not long ago, I tried to find out what the resources perceived as being unequally distributed are and what the acceptable sources of inequality in Romania are; I asked the participants (41 women and men, aged between 11 and 70 years old, with primary, high school and university studies) to draw the social classes in today's Romania and in an ideal Romania. Their drawings were very similar, none of the sampling criteria mentioned generated important variations in perceiving inequality (see Scârneci-Domnișoru, 2015). It is true I did not succeed in following the existence of variations by other criteria that could generate variation, such as the income of the respondents or their residence area (urban/rural). Nevertheless, studying status-beliefs is one of the situations in which we could make generalizations on heterogeneous populations. It is a phenomenon that is related rather to habitus, therefore I believe the perceptions of Romanians regarding inequality are rather similar, irrespective of gender, age, occupation etc. I collected, in the same research, a few drawings from Italians, representing inequality in Italy and the differences between what Romanians drew, Italians respectively, were extremely high. The comparisons between the drawings of Romanians and Italians have led me to formulate some 'very' general hypotheses (that could have become theories if I had succeeded in making an exemplary theoretical sampling) – for instance, that inequality sources are perceived by people hierarchically (as in the pyramid of needs, first basic, economic inequalities, then those related to status or cultural) or that the perceived level of inequality determines its acceptability (the higher the perceived level of an inequality is, the more acceptable it becomes).

Finally, I would like to underline that if we want to try the generalization for a larger and less homogeneous population, it

would be good that the procedure is from few and/or homogeneous (i.e. a small and/or homogeneous research universe) to many and/or heterogeneous (i.e. a large and/or heterogenous research universe) and not the other way round. I have also mentioned in the previous paragraph that this type of adjustment to the research universe is preferable and I explain here why. It is very difficult to know before studying it, how complex and variable a phenomenon is. It is good to suppose that it has many variations, to try to discover them, and step by step, to enlarge the research universe as much as we can. Data variations will be easier to identify and manage if we extend the research universe than if we apply the procedure in the other direction.

In conclusion, it is possible to be more daring in qualitative research, less modest in aspirations; we can work in larger teams, extend or continue other people's research; we could go beyond the local level, conceptualize, produce new theories that are outside the box. I preserve the recommendation that the initial population be small and homogeneous to ensure we have high chances to generalize in qualitative studies, but because the procedure presented in this article focuses on the variation of the studied phenomenon, but what really matters (when it comes to generalization and its volume) is this aspect – how complex the phenomenon is, how much it varies upon the characteristics of the population.

7. Conclusion

The sampling that allows for result generalization (the probabilistic statistical one and the theoretical one) are difficult to perform; for this reason, it is no wonder that they are not prevalent in social research. Nevertheless, I argued in this article that generalizations are extremely necessary and that they should be made by following a procedure.

I tried to convince that generalization of research results should be found among the basic preoccupations of researchers, to show that sampling is not a characteristic specific only to statistical sampling, that internal generalizations are not necessarily statistical and that generalization, even in the form sample to population, can be performed also in qualitative research. I tried to argue that generalization should not be judged by the size of the sample, but by the criteria for selecting the participants and that there is representativity also without probability, that it can be obtained if sampling is guided by the discoveries in data.

I described a sampling procedure in four steps, that can lead to generalization of findings in qualitative studies with several units of analysis. I insisted on the idea that saturation has to be reached for each sampling criterion generated by the variation noted in data (or by the requirements for establishing theories) and on the idea that deviant cases and the falsification of partial results should be looked for and followed.

The procedure presented in this article can be applied to any study in which we have several units of analysis (not to case studies) and in which we begin from a research universe that is not too large and homogenous that cannot be interviewed, observed etc. in its entirety. The procedure can be used when we have flexibility of research methods, techniques and instruments (when we collect data by using unstructured or semi structured techniques – not by using a questionnaire!; when we process data by coding or hybrid variants – not when we perform a quantitative content analysis! etc.) and when we have flexibility in choosing the participants (when we have not collected all data from the beginning).

The procedure ensures the transparentizing of sampling and the proof that can be brought to support result generalization (justified generalization); it supposes specifying exactly the population to which the research refers and the applied sampling criteria for which we can guarantee result generalization. The implementation of a clearer procedure is necessary in the sampling process in qualitative research because most times, the sampling is performed carelessly, and it is necessary to seriously ask ourselves with regard to the utility and value of qualitative research that we perform without proposing the generalization of results.

Statements and Declarations

Acknowledgements

Your encouragements, Ralu, your questions, Dana, and your doubts, Carmen, made me write! I thank my friend Raluca Filip and my colleagues, Daniela Sorea and Carmen Buzea, who also undertook a review of an earlier version of this article.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical statement

Ethical approval and informed consent are not applicable for this article.

References

- Bloor, M. & Wood, F. (2006). Generalization. In *Keywords in qualitative methods* (pp. 94-95). SAGE Publications Ltd, https://dx.doi.org/10.4135/9781849209403
- Eisenhart, M. (2009). Generalization from qualitative inquiry. In K. Ercikan and W-M. Roth (Eds),*Generalizing from educational research* (pp. 51-66). New York: Routledge.
- Ercikan, K. & Roth, W.M. (2006). What Good Is Polarizing Research Into Qualitative and Quantitative?.*Educational Researcher*, 35(5), 14-23, doi: 10.3102/0013189X035005014
- Ercikan, K. (2009). Limitations in sample-to-population generalizing. In K. Ercikan and W-M. Roth (Eds),*Generalizing*

from educational research (pp. 211-234). New York: Routledge.

- Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher 22*, 16–23.
- Gobo, G. (2004). Sampling, representativeness, and generalizability. In Seale, C., Gobo, G., Gubrium, J. F., Silverman, D. (eds.), *Qualitative research practice* (pp. 435 – 456). London: Sage.
- Gobo, G. (2008). Re-conceptualizing generalization: Old issues in a new frame. In Pertti Alasuutari et al. (eds.),*The SAGE Handbook of Social Research Methods* (pp. 193-213). Thousand Oaks, CA: SAGE Publications.
- Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, 18(1), 59–82. https://doi.org/10.1177/1525822X05279903
- Guest, G., Namey, E., & McKenna, K. (2017). How Many Focus Groups Are Enough? Building an Evidence Base for Nonprobability Sample Sizes. *Field Methods*, 29(1), 3–22. https://doi.org/10.1177/1525822X16639015
- Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough? *Qualitative Health Research*, 27(4), 591–608. https://doi.org/10.1177/1049732316665344
- Hennink, M. M., Kaiser, B. N., & Weber, M. B. (2019). What Influences Saturation? Estimating Sample Sizes in Focus Group Research. *Qualitative Health Research*, 29(10), 1483–1496. https://doi.org/10.1177/1049732318821692
- Larsson, S. (2009). A pluralist view of generalization in qualitative research.*International Journal of Research & Method in Education*, 32(1), 25-38, doi: 10.1080/17437270902759931
- Maxwell, J. A. (2021). Why qualitative methods are necessary for generalization.*Qualitative Psychology, 8*(1), 111–118. https://doi.org/10.1037/qup0000173
- Maxwell, J. A., & Chmiel, M. (2014). Generalization in and from qualitative analysis. In Uwe Flick (Ed.),*SAGE Handbook of Qualitative Data Analysis* (pp. 540-553). London, UK: SAGE Publications.
- Mitchell, J. C. (1983). Case and Situation Analysis. *The Sociological Review*, *31*(2), 187–211. https://doi.org/10.1111/j.1467-954X.1983.tb00387.x
- Onwuegbuzie, A. & Leech, N. (2005). Taking the “Q” Out of Research: Teaching Research Methodology Courses Without the Divide Between Quantitative and Qualitative Paradigms. *Quality & Quantity*, 39(3), 267–295, doi: 0.1007/s11135-004-1670-0
- Onwuegbuzie, A. & Leech, N. (2010). Generalization practices in qualitative research: a mixed methods case study. *Quality & Quantity,* 44, 881–892, https://doi.org/10.1007/s11135-009-9241-z
- Payne, G., & Williams, M. (2005). Generalization in Qualitative Research. *Sociology*, 39(2), 295–314. https://doi.org/10.1177/0038038505050540
- Scârneci-Domnișoru, F. (2015). Acceptable and Unacceptable Sources of Inequality in Romania. A Visual Study. *TRAMES - A Journal of the Humanities and Social Sciences*, 19(4), 367-402.
- Scârneci-Domnișoru, F. (2018). Reflections on Social Research Methodology. *Bulletin of the ‘Transilvania’ University of Braşov – Series VII: Social Sciences and Law*, 11(2), 91-102.
- Smaling, A. (2003). Inductive, analogical, and communicative generalization. *International Journal of Qualitative Methods, 2*(1). Article 5. Retrieved from
- http://www.ualberta.ca/~iiqm/backissues/2_1/html/smaling.html

- Vogt, P. (2008). Quantitative versus Qualitative Is a Distraction: Variations on a Theme by Brewer and Hunter (2006). *Methodological Innovations Online*, 3 (1), 18-24. 10.4256/mio.2008.0007