Confused About Theoretical Sampling? Engaging Theoretical Sampling in Diverse Grounded Theory Studies

Catherine Conlon¹ Virpi Timonen¹ Catherine Elliott O’Dare³ Sorcha O’Keeffe³ & Geraldine Foley¹

Accepted pre-publication version. Published in Qualitative Health Research (2020)

Catherine Conlon BA, MA, PhD, Assistant Professor in Social Policy, School of Social Work and Social Policy, Trinity College Dublin¹, Ireland

Virpi Timonen BA, MPhil, DPhil, Professor in Social Policy and Ageing, School of Social Work and Social Policy, Trinity College Dublin¹, Ireland

Catherine Elliott O’Dare BA, PhD, Teaching Fellow in Social Policy, School of Social Policy, Social Work and Social Justice, University College Dublin², Ireland

Sorcha O’Keeffe BSS. (Social Work), PhD, Medical Social Worker, Department of Medical Social Work, Beaumont Hospital Dublin³, Ireland

Geraldine Foley BSc.OT, MSc, PhD, Assistant Professor in Occupational Therapy, Discipline of Occupational Therapy, School of Medicine, Trinity College Dublin¹, Ireland (Corresponding Author)

Address: Discipline of Occupational Therapy, School of Medicine, Trinity College Dublin, Faculty of Health Sciences Building, James’s Street, Dublin 8, Ireland. Email: foleyg3@tcd.ie Tel: +353 1 896 3217
Confused About Theoretical Sampling? Engaging Theoretical Sampling in Diverse Grounded Theory Studies

Theoretical sampling is a key procedure for theory building in the grounded theory method. Confusion about how to employ theoretical sampling in grounded theory can exist among researchers who use or who want to use the grounded theory method. We illustrate how we employed theoretical sampling in diverse grounded theory studies and answer key questions about theoretical sampling in grounded theory. We show how theoretical sampling functions in grounded theory and how it differs from sampling for data generation alone. We demonstrate how induction, retroduction and abduction operate in grounded theory and how memoing drives theoretical sampling in the pursuit of theory. We explicate how theoretical sampling can contextualize data in order to build concepts and theory. Finally, we show how theoretical sampling in grounded theory operates in secondary analysis to derive theory that goes beyond the original purpose of data collection.
Introduction

In qualitative research, sampling is not about reaching numerical significance nor does the sample size underpin what researchers interpret as significant (Merriam & Tisdell, 2016; Pope & Mays, 2006). Generating data in qualitative research is about strategically selecting participants whose views and experiences can add meaning to, illuminate and in some cases, help explain the phenomenon under study (Cleary, Horsfall, & Hayter, 2014).

Sampling that enables researchers to capture similarity, difference and/or anomalies in the pursuit of new and meaningful data, is a core feature of any qualitative research. Sampling in qualitative research is usually conceptualized as focusing on the characteristics of the cases or instances identified as pertinent sources of data for the inquiry. However, grounded theory introduced another premise for sampling concerned with generating data specifically related to conceptual categories that emerge as of interest to the inquiry. This refers to theoretical sampling. Here we outline and discuss what theoretical sampling entails and illustrate theoretical sampling in action using exemplars from five research projects in the areas of health, aging and intergenerational relations.

A common way of collecting data in qualitative research is through purposive sampling – the identification and strategic selection of information-rich cases related to the topic of interest. What constitutes an information-rich case in purposive sampling is broadly defined at the outset of a study. Theoretical sampling is a form of sampling in qualitative research that is not bounded by the limits of *a priori* selection. Rather, theoretical sampling entails jointly collecting and analyzing data to decide what data to collect next and where to find them in order to develop theory (Glaser & Strauss, 1967/2012). It involves identifying
emergent concepts in data being generated which are then used to guide where, how and from whom more data should be collected, and with what focus (Butler, Copnell, & Hall, 2018). Theoretical sampling is an innovation of the grounded theory method - a systematic research approach that builds concepts and theory from data (Bryant, 2017; Charmaz, 2014; Corbin & Strauss, 2015; Glaser, 1992). The main aim of grounded theory is the generation of theory from data. Theoretical sampling is theoretically oriented because it is directed towards the development of concepts and theory and focuses data collection on building and refining theory (Glaser & Strauss, 1967/2012). Theoretical sampling originates in the early writings on the method by the founders of grounded theory, Barney Glaser and Anselm Strauss (Glaser, 1978; Glaser & Strauss, 1967/2012), and it is employed by all variants of grounded theory.

Theoretical sampling is a core principle of grounded theory, and necessary for any grounded theory study which reports on the construction and development of a conceptual framework and/or theory about the phenomenon under study (Charmaz, 2014; Corbin & Strauss, 2015; Glaser, 1992). Theoretical sampling arises out of the logic of discovery underpinning the method which requires the researcher to engage in inductive as well as abductive and retroductive reasoning. Inductive reasoning involves moving from observing particular instances or cases, to developing a general abstract characterization with reference to the particular characteristics observed. The early conceptualization of grounded theory methodology as a means of discovery that generates insights and observations without engaging with any preconceived ideas, concepts or theories (Glaser, 1992) positioned it closest to inductive reasoning. However, as the procedures of grounded theory have been written about further, the place of abductive and retroductive reasoning
in its methods of discovery has been established (e.g., Charmaz, 2014; Flick, 2018; Oliver, 2012; Timmermans & Tavory, 2012; White, 1997). Forming a series of hypotheses through observation of surprising or anomalous data in the search for novel theory and returning to empirical data to test and confirm inferences are core practices of grounded theory. In grounded theory, ‘all is data’ which means that everything the researcher encounters when studying the topic of concern is data. Although classical grounded theory envisages theory building from data without engaging with existing literature before concepts and categories have been fully theorized (Glaser, 1992), other iterations of grounded theory (e.g., Charmaz, 2014; Clarke, 2005; Corbin & Strauss, 2015; Oliver, 2012) anticipate that data may lead the researcher beyond empirical observation to engage in an open way with extant theoretical generalizations of the social world that either animate or are animated by the particular empirical observations at hand in the search for novel explanations and theory - a retroductive logic.

In grounded theory variants that allow for engagement with empirical data and extant theoretical accounts to build theory (Charmaz, 2014; Clarke, 2005; Corbin & Strauss, 2015; Oliver, 2012), insights generated by considering empirical data in light of pre-existing theories or concepts are ‘points of departure’ from which the researcher returns to generate further, or review existing, empirical data to ‘sample’ or try out if their tentative explanation ‘fits’ – abductive reasoning (Reichertz, 2009; Timmermans & Tavory, 2012). Abductive reasoning involves the teasing out of possible causes or antecedents of what is emergent in data. Retroductive reasoning comes from a broader understanding of the context or conditions where data are being and have previously been collected. Theoretical sampling is the key data collection method that steers the logic of discovery in grounded
Theoretical sampling directs the researcher in moving back and forth between data and the emergent theory to identify empirical instances with the potential to further deepen insights into theoretical constructs that have emerged from data or (with the exception of classical grounded theory) that pre-date the inquiry. Theoretical memoing in grounded theory is the key tool that helps the researcher to record insights and directs further data collection until theoretical saturation is reached.

In grounded theory, the goal is always to reach theoretical saturation of the data (Charmaz, 2014; Corbin & Strauss, 2015; Glaser, 1992) - distinctly different from saturation signalled simply by the absence of ‘new’ data from fieldwork. Rather, theoretical saturation refers to the point when no significant new insights from data can better describe, dimensionalize or contextualize the categories that have emerged from data (Charmaz, 2014; Corbin & Strauss, 2015). In classical grounded theory, it is considered to be the point when comparisons of key incidents in data yield no additional properties of patterns emerging from data (Glaser, 1992). Theoretical saturation can only ever be pursued through theoretical sampling because the focus of grounded theory is to pursue saturation of a theoretical construct emerging from data to develop a theoretical account posited as the ‘best fit’ of a concept; that researchers choose a name with the best fit for the pattern of behavior (Glaser & Strauss, 1967/2012; Saunders et al., 2018). This contrasts with, for example, a concern to obtain sufficient spread of codes, categories or themes across the dataset to fully represent all dimensions of the phenomenon of interest (as thematic analysis would emphasize).
Despite the availability of writings about theoretical sampling in grounded theory (e.g., Breckinridge & Jones, 2009; Coyne, 1997; Emmel, 2013) including key works that have outlined grounded theory (e.g., Charmaz, 2006/2014; Corbin & Strauss, 2008/2015; Glaser, 1978; Glaser & Strauss, 1967/2012; Strauss & Corbin, 1990/1998), we believe, as do others (e.g., Butler et al., 2018; Draucker, Martsolf, Ross, & Rusk, 2007), that there is confusion about how to employ theoretical sampling in grounded theory among researchers who use or who want to use the grounded theory method. Many of the textbook writings on theoretical sampling in grounded theory deal primarily with the methodological underpinnings of theoretical sampling and a limited number of textbooks (e.g., Bryant & Charmaz, 2019) have focused specifically on how theoretical sampling is accomplished in practice in grounded theory research. Researchers who claim to use grounded theory do not always report theoretical sampling in their studies, and those who do report that they used theoretical sampling do not always describe how they used it (McCrae & Purssell, 2016). Experienced and novice qualitative researchers can have uncertainties about how to do theoretical sampling.

As grounded theory studies are initiated and progress, researchers may find themselves asking: Is theoretical sampling distinct from purposive sampling or simply an extension of purposive sampling? When should theoretical sampling begin in a grounded theory study? Why are retroduction and abduction important for theoretical sampling in (non-classical) variants of grounded theory? How can theoretical memoing drive theoretical sampling in theory generation? How does theoretical sampling for concepts in different contexts help capture and explain process? Can theoretical sampling be employed when data are not collected and analyzed in tandem, for example when an existing dataset is analyzed from a
new perspective (i.e. secondary analysis)? And when has theoretical sampling fulfilled the requirement for theoretical saturation?

With the help of five exemplars, underpinned by diverse epistemological and ontological interpretations of grounded theory (for detailed outlines on the different variants of grounded theory, please see Flick (2018) and Rieger (2019)), we strive to answer the above questions through accounts of theoretical sampling as employed by us. Our exemplars here show how theoretical sampling can be accomplished in three main ways at various stages of a research project. First, theoretical sampling can involve sampling for additional participants with a particular set of theoretical considerations in mind. Second, theoretical sampling can be progressed through a variety of means and techniques in the actual data collection process. Most commonly, this happens through interviewing, for instance steering questions in the direction of emergent theorizing. We show how the focus of data collection, including the questions asked, can change in the theoretical sampling process. Third, theoretical sampling can sometimes be done within the data in isolation from the process of data collection, that is, when the dataset is approached as secondary data (as illustrated in the last exemplar). We demonstrate the role of induction, retroduction and abduction in theoretical sampling, and how theoretical memoing in grounded theory helps to drive theoretical sampling. Through demonstrating how to do theoretical sampling in different grounded theory studies, our aim here is to make theoretical sampling more accessible for researchers who are thinking about conducting a grounded theory study or who are already engaging theoretical sampling in their research. Each study outlined below had ethical approval from a university-based research ethics committee and in some cases also from a healthcare institution research ethics committee.
Theoretical sampling in an ‘accessible’ population to build grounded theory

This exemplar is reported from a study which aimed to generate substantive theory about how people with amyotrophic lateral sclerosis (ALS) engage with healthcare services. Prior to the study that this section draws on, the healthcare experiences of people with ALS had not been substantially theorized. In this study, procedures outlined by Corbin and Strauss (2008/2015) were followed and a critical perspective (Bhaskar, 1986/2009; Clark, Lissel, & Davis, 2008) adopted, to explain how structure (i.e. conditions) and relations shaped events and interactions that were communicated from the viewpoint of participants.

Between September 2011 and August 2012, qualitative interviews were conducted with 34 people with ALS (17 men and 17 women, aged between 37 and 81 years) across the Republic of Ireland, who were sampled from the Irish ALS population-based register. The availability of the disease-specific population-based register to sample from, allowed for the capture of a diverse group of participants and for sampling as guided by data. A large majority of people on the ALS population-based register attended a national ALS clinic (in addition to availing of local and regional-based care) and we (Foley and Timonen) had access to records on potential participants at the clinic. As stated, key to theoretical sampling in grounded theory is to sample participants and/or sources based on characteristics and/or experiences that can further develop emerging categories in data (Corbin & Strauss, 2015). Access to data on service users was an important tool in the research process because it enabled (in advance) the identification of potential participants that could add further variation and meaning to concepts that were already emerging from data in the study.
Access to the population-based register also minimized delay in sourcing participants who were sampled on the basis of emerging findings i.e. it was possible to collect and analyze data in tandem.

During the early stages of fieldwork, participants were purposively sampled for heterogeneity to capture an ALS sample with variation along standard dimensions (e.g., ALS type, severity of disability, and disease stage). From grounded theory analysis of that data, categories began to emerge. Participants were then sampled for variation based on these emerging categories in the pursuit of category saturation and theory. For example, when we identified that aging and parenthood were emerging as key categories that shaped participants’ preferences for care, we continued to sample participants for variation in these contexts (e.g., people with ALS at different life stages, and those who had dependents and those who had no dependents) (Foley & Timonen, 2015). At the same time, the developing category ‘family support’ was also emerging as context that shaped participants’ preferences for and decisions about care, and so participants who had varying degrees of family support available to them were sampled. As analysis of data continued during and after fieldwork, it emerged that family context encompassed more than just participants’ reliance on family. It emerged that family context of participants’ decision-making about care was underpinned by participants’ own family roles at different life stages, and that being a parent and/or a partner at a certain life stage shaped how much support participants themselves felt obliged to provide to their family (Foley, Timonen, & Hardiman, 2014a; 2016).
We quickly progressed from ‘opening up’ our data to making connections between developing categories, and by making inferences between data, started to build relationships between categories. Existing knowledge in the field was used in order to further conceptualize the categories which had already emerged from the data. For example, in the early stages of fieldwork and analysis, it was identified that feeling reassured by healthcare professionals was important to participants. However, data coded early on under ‘feeling reassured by healthcare professionals’ reflected primarily participants’ accounts of care at the national clinic. We were cognizant of how people with ALS had already (in other research studies) expressed greater satisfaction with specialized ALS care than with non-specialized care and through retroductive reasoning integrated this into the conceptual work of generating a theoretical account of generated data. Hence, in our attempt to theoretically saturate (i.e. fully dimensionalize) the category ‘reassurance’, we (theoretically) sampled people from the register who had never engaged with services at the national clinic, and asked questions about what underpinned reassurance for them in their interactions with healthcare services. In the end, we identified that feeling reassured was for most participants fundamentally about having control over care, including end-of-life care, regardless of who provided the care (Foley, Timonen, & Hardiman, 2014b; 2014c). This example illustrates how theoretical sampling can be maximized to build theory particularly when the target population is already known and accessible.

**Retroduction and causation in theoretical sampling for grounded theory**

This exemplar is reported from a grounded theory study that explored the experiences of people with (primarily incurable) brain tumor diagnoses (glioma). Using critical realist
grounded theory which is a grounded theory method used to operationalize theory underpinned by critical realism (Oliver, 2012), O’Keeffe had sought to generate a theoretical framework to explain how people with brain tumor diagnoses adapt to and cope with their illness through the early stages of their illness trajectory (O’Keeffe, 2019).

Between February 2016 and October 2016, 16 people (10 men and six women, aged between 19 and 70 years) with a recent diagnosis of a glioma were recruited at a specialist oncology unit and interviewed about their experiences of the illness before and after radiotherapy treatment of their glioma. A treatment-planning record system detailing all patients at the specialist oncology unit who were due to begin treatment for glioma was available to O’Keeffe. Information on the treatment-planning record system was limited to diagnosis, gender, age, occupation, home address, and in some cases, marital status. Consistent with grounded theory principles, recruitment was (insofar as possible) driven by the goal to saturate emergent categories. During the early stages of fieldwork, participants were purposively sampled for heterogeneity along a number of dimensions (e.g., tumor type, life stage, family relations, and employment status) to capture an overview of the experiences of glioma patients. As analysis progressed, key contextually-driven categories emerged in the data and participants were then sampled on the basis of these emerging categories. For example, when it became apparent early on in the study that communication styles surrounding diagnostic and prognostic communication were impacting on how participants responded to their illness, O’Keeffe sampled participants who had been treated by different neurosurgical and radiation-oncology physicians and who varied in how they approached communicating diagnoses and prognoses with patients. These patients were then asked new questions about their own communication preferences.
during the disclosure of a diagnosis or a prognosis, and what style and form of communication they believed would best aid future patients to cope with their illness.

Contrary to the preceding exemplar, there were some challenges to theoretical sampling. As analysis progressed, the paucity of information available on the patient treatment-planning record system meant that it was not possible to simply return to the record to sample participants with particular characteristics that could saturate the emerging categories. Hence, patients were (theoretically) sampled via other routes within the oncology unit. On occasions, healthcare professionals based at the oncology unit recommended potential participants to O’Keeffe when she communicated to them the concepts and categories which were driving sampling. The challenges of identifying patients who corresponded with the focus of theoretical sampling moved O’Keeffe to increasingly engage in sampling for concepts within existing interviews, thereby effectively generating new data and analyzing existing data simultaneously to inform, build and explain categories. Moreover, because O’Keeffe was conducting interviews longitudinally with most participants, she was able to theoretically sample for the emerging concepts and categories from earlier interviews to guide questioning in later interviews with participants that would allow for the development and further saturation of emerging concepts and categories. Theoretical saturation was reached after analysis of data from 16 participants across 43 interviews.

Consistent with critical realist grounded theory (Oliver, 2012), emergent categories were developed through the process of retroduction. Key behaviors identified in data generated in the study were interrogated via the examination and exploration of ‘what must happen,
for this to happen’ (Archer, 1998). Rigorously asking such questions of data allowed O’Keeffe to track causal mechanisms at play and provide explanations surrounding actions and behaviors of participants already coded in data generated from participants. For example, inductive analysis alone had revealed that participants prioritized coping with their illness and identified the methods through which they sought to cope. O’Keeffe then turned to the literature in search of new insights that could deepen her understanding about what motivated people living with a glioma to prioritize coping (i.e. retroduction). She looked specifically for evidence or theory that could elaborate on why people with a brain tumor or similar prognosis had sought to cope. Returning to the literature enabled her to help identify: what is it about the experience of living with a glioma that would cause study participants’ behaviors and responses (already coded in data from participants); and what participants sought to achieve through their actions and behaviors (already coded in data from participants).

For example, following the emergence of the category ‘seeking to cope with illness’, O’Keeffe looked closely at the relationships between it and other highly-conceptualized categories. She identified that many other categories had dimensions that underpinned the need to cope and that these categories helped explain how and why participants sought to cope. The category ‘to make the most of everyday life’ was identified as a key category that underpinned the need to cope. The literature on cancer and coping at this point revealed that often people had to be able to adapt in order to cope. Further analysis of the data then identified that participants were adapting by ‘maintaining hope’ that they would survive longer than expected. Moving back and forth between data allowed O’Keeffe to identify and interrogate the relationship between coping and maintaining hope, and to hypothesize on
the causal relationship between both. O’Keeffe analyzed all data that captured how and why participants maintained hope. Analysis revealed that maintaining hope for a better outcome was in fact a necessary condition for participants to cope in order to make the most of everyday life. In the end, ‘maintaining hope in order to cope’ emerged as the core category indicated by abductive reasoning as representing a central part of the explanatory framework, superseding all other categories in explanatory importance (O’Keeffe, 2019). This example illustrates how both retroduction and abduction feature in theoretical sampling in the pursuit of theory.

**Theoretical memoing to engage theoretical sampling in grounded theory**

This exemplar for theoretical sampling is reported from a grounded theory team-based study ‘Changing Generations’ which focused on how intergenerational solidarity is constructed across generations. The study, conducted between 2011 and 2012, comprised a dataset of 100 in-depth face-to-face interviews with 52 women and 48 men, aged between 18 and 102 years. Underpinned by constructivist grounded theory (Charmaz 2006/2014) - which points to how theory is constructed between the researcher and participants - the data yielded multiple accounts of the meaning and practice of solidarity within familial relationships and of positions on solidarity at the societal level (Conlon, Timonen, Carney, & Scharf, 2014a; 2014b; Timonen, Conlon, Scharf, & Carney, 2013).

Data generation first started with convenience sampling anchored in areas selected for diversity regarding deprivation and affluence so as to allow data to be contextualized in its situational and social contexts (Charmaz, 2006/2014; Clarke, 2005; Conlon, Carney,
A unique aspect of the study was its multi-generational composition and purposive sampling was engaged to generate data on both the expectations and aspirations of those who were aging and contemplating the prospect of having support needs and the aspirations and practices of younger participants with regard to providing help, support and care for older family members. Codes initially focused on transfers i.e., the ‘give and take’ of help and support and constant comparative analysis of the dataset revealed contrasting forms of inter-reliance and independence among family generations according to socio-economic status (Timonen at al., 2013). The contrasting data yielded some conceptual insights into ‘what is going on here’ and tentative theorizing got underway.

An unanticipated process of changing relations in exchanges of care along gender lines across different family generations began to feature in memos as data generation continued. Theoretical memos documented how older women were observing younger women’s lives, particularly the lives of their daughters, and were being prompted by this to reflect on how caring had featured across their own life course. In theoretical memos, it was identified that (some) older women were taking a critical stance towards how the gendered ordering of care roles and responsibilities shapes and constrains women’s lives. Through theoretical memos, we documented our observations about older women observing younger generations of women in their family gaining higher education and more prestigious employment than had been available to them and about older women also speculating on how the younger women’s considerably expanded involvement in diverse roles would be disrupted if they were to assume caring roles as the older women had often done. A process seemed to be at work whereby older women were having regard to the
potential for caring to curtail women’s expanding roles. We set out to theoretically sample for this in prospective data generation (by asking older women about their expectations surround receiving care from their children or providing assistance to their children), as well as retrospectively interrogating the data generated before this tentative process had emerged from the data analysis (Conlon et al., 2014a).

Interrogating interviews with older high and upper-middle socio-economic status women revealed a pattern whereby they were asserting their unavailability to care for grandchildren due to their own interests. In extensive theoretical memos, we abductively drew on sociological theories of late modernity, individualization and ‘doing gender’ to inform our understanding of this process. Theorization via memoing directed us to interrogate this process in data generated with professional women in middle years who had young children and whose own mothers were active and living close by. A process of grandmothers withholding involvement in the care of grandchildren was confirmed. Furthermore, this data revealed a process whereby - as the older women were resisting caring for grandchildren - they were also explicitly or implicitly conveying to their daughters that in so doing, they were laying down a pattern of (non-) exchange of care so that their daughters would not be expected to care for them in later years should care needs arise.

We theorized the above process of older women refraining from asking for help and support from younger women in the family and at the same time withholding providing care to grandchildren, as a shift in the place of care in intergenerational family relations (Conlon et al., 2014a). This process constituted complex forms of intergenerational solidarity at collective and private level evolving among women, so that the absence of caring for
relationships was not an indication of lack of intergenerational solidarity but rather indicated changing forms of intergenerational solidarity (Conlon et al., 2014a). This example illustrates how theoretical memoing in grounded theory can help drive theoretical sampling to interrogate data and develop theory.

**Theoretically sampling for concepts in contexts to explain process in grounded theory**

This exemplar for theoretical sampling is reported from a grounded theory study which set out to explore and understand intergenerational friendships i.e. friendships which occur between differing generations of older and younger adults. Constructivist grounded theory (Charmaz, 2006/2014) was used to generate theoretically informative data on the topic which up this point had not been well understood (Elliott O’Dare, Timonen, & Conlon, 2019a). The study focused on the perspective of the older friend. In total, 23 people aged between 66 and 95 years, were interviewed in Ireland in their homes over the course of 18 months (Elliott O’Dare, Timonen, & Conlon, 2019b; 2021).

Purposive sampling commenced via access to the database from the ‘Changing Generations’ study as detailed in the preceding section. Seven people from ‘Changing Generations’ and who fitted the recruitment criteria for the intergenerational friendship study, agreed to participate. Snowball sampling was also used whereby one participant from ‘Changing Generations’ recruited another participant who fulfilled the recruitment criteria for the intergenerational friendship study. The ‘Changing Generations’ database was a useful launch pad from which to begin the iterative grounded theory process of data gathering, coding, and memoing. Thereafter, by constantly returning to the field, concepts were
theoretically sampled to flesh out the multiple facets of intergenerational friendship and to seek theoretical explanations of ‘doing’ intergenerational friendship and ‘being’ an intergenerational friend.

To explore the process of friendship formation and maintenance for older people, concepts were theoretically sampled for in different contexts that could help to detect how and why participants were seeking friendships with younger people and how they acted as intergenerational friends. Multiple routes were taken to recruit participants including a blog posted on the website of an age-friendly university and posters distributed in recreational and meeting halls attended by older people. Early data collection and analysis had identified the nascent concept of ‘being friends in action’, which directed the study to go to sites where people typically gathered to spend time together and engage in activities of interest. Inductive analysis of data had also revealed that sharing interests, hobbies, and leisure pursuits or simply spending time with like-minded people had fostered friendship for participants regardless of age (Elliot O’Dare et al., 2021). Hence, we also sampled participants by directly engaging with social and leisure groups where people of all ages gathered.

As analysis progressed, ‘forging and maintaining’ intergenerational friendships emerged as a complex and multidimensional process. For example, it was identified that experiencing ‘status transitions’ in the form of ‘official’ retirement from work proved to be a conduit for intergenerational friendship. In order to analyze why this was the case, varied contexts for the concept ‘status transitions’ were sampled concerning employment including people who
had continued in employment in an ‘informal’ capacity after official retirement. Participants were asked why some older people have or seek intergenerational friendships during later-life status transitions. We identified subsequently that many participants were actively pursuing involvement in society beyond the traditional parameters of retirement in older age. Some participants sketched a fearful portrayal of old age (e.g., ‘old fogey’) in the absence of intergenerational activities. Analysis of the data across contexts revealed that forging and maintaining intergenerational friendships signalled a process of ‘being not only old’ through which participants both constructed and resisted negative stereotypical depictions of old age (Elliott O’Dare et al., 2019b).

To further expand on and explain the process of forging and maintaining intergenerational friendships, contexts were sought out that could help explain why and how older people ‘transition as an intergenerational friend’. Of note, we identified that many participants spoke of their desire to maintain their intergenerational friendships in the face of significant change in their lives. For example, participants spoke of their hope to maintain intergenerational friendships after they had transitioned to older-age community living, and this resulted in theoretically sampling more participants who had already transitioned to older-age community living and asking them about their experiences and expectations of intergenerational friendships. Analysis of participants’ accounts, together with observational memoing, revealed that people who had already transitioned to older-age community living sought continued connectedness and belonging through forming and maintaining intergenerational friendships. This example is a good illustration of sampling for concepts within multiple contexts in a grounded theory study and of how data that have been contextualized in a constructivist grounded theory study can help explain process.
Theoretical sampling during and after data collection in grounded theory

This exemplar for theoretical sampling is reported from the Irish strand of a European project SoCaTel – Co-creating for a better life (SoCaTel, n.d.) for which Timonen is PI. The project seeks to respond to the needs of aging populations in Europe by improving the accessibility and responsiveness of care services. In spring 2018, Timonen conducted focus groups with other research team members in the Dublin area. Purposive sampling was initially utilized to recruit participants who corresponded to key stakeholder groups, including older adults using or potentially using care services, policy planners, and a range of service providers. Soon after initiation of sampling, theoretical sampling was engaged to saturate key concepts that were emerging from the study’s data around the issue of care provision as that is the focus of the project. The team recruited 104 participants who took part in 21 (homogenous e.g. only older adults or social workers) focus groups and three in-depth interviews (two family caregivers with limited time, and a senior policy maker who was more informative in an interview context).

As per key tenets of grounded theory and following Corbin and Strauss (2015) grounded theory procedures, the team employed a lightly-structured interview guide and remained reflexive through the data collection, generating a memo after each interview and focus group. Each transcript was read and analyzed using grounded theory methods; this involved constant comparison between data, looking for similarities and differences between conditions and consequences surrounding key events, and identifying patterns (concepts and categories) in the data. The team were able to meet the project requirements and
wrote articles around the focal issue of care services accessibility and responsiveness (e.g., Lolich, Riccò, Deusdad, & Timonen, 2019; McDonald, Lolich, Timonen, & Warters, 2019), as the concepts around this had been theoretically saturated.

Subsequent to completing data collection and carrying out the analyses around the focal point of the project, Timonen and a researcher working with her (Lolich) identified a further topic in the data that had until then remained latent. When Timonen and Lolich examined data from a sub-sample of 31 participants who are various health and social care professionals (such as social workers, community nurses and care managers), we detected distinctive talk around what we termed social constructions of the care worker. Care workers – unlike the professionals involved in formal care provision to older adults in Ireland - do not have professional or higher education qualifications. While some basic requirements have been introduced in Ireland for those who wish to gain employment as care workers, the sector is not professionalized and the requirements are not consistently enforced as the sector remains unregulated. Hence, it was judged possible and meaningful, for the purposes of additional theoretical sampling, to distinguish between the professionals in elder care and the care workers, and to focus the analysis on how the former socially construct the latter.

We were struck by how, in this later reading of the data, professionals in the sub-sample of 31 recognized the difficult conditions and pressures of care work and expressed empathy and admiration towards carers - as reflected in expressions such as ‘the poor carer’ and ‘undervalued’ – yet in parallel they described instances where a care worker was incompetent, neglecting the client, or otherwise inadequate. We coded the professionals’
talk as reflecting contradictory archetypes, such as ‘the saintly carer’ versus ‘the morally questionable carer’; the simultaneous existence of these contradictory depictions led us to identify the category of ambivalence. We memoed extensively our theorizing about ambivalence and then pursued theoretical saturation for these ambivalent descriptions in the data, focusing on the contradictions emerging from participants’ talk, and mapped these tensions onto structural factors in the Irish care regime (Timonen & Lolich, 2019).

Employing abductive reasoning, it was concluded that the ambivalent portrayals of carers as both ‘good’ and ‘bad’ are indicative of the deep contradictions in the expectations that contemporary care systems direct at paid caregivers. Ambivalence arises from the commodified and dispensable status of care workers, and fundamental transformations in their training, working conditions and pay are required to move away from this ambivalence and towards equal status with other care professionals (Timonen & Lolich, 2019). This example illustrates how an existing dataset can be used to sample theoretically for concepts, to achieve theoretical saturation, and to derive additional theory that is related to, but goes ‘over and above’ the original purpose of data collection in a grounded theory study.

**Discussion and conclusions**

The above exemplars have illustrated how theoretical sampling can be accomplished through different steps and techniques at various stages of a research project. First, it can be done through sampling for participants with a particular set of theoretical considerations in mind (not merely sampling for additional cases or cases with characteristics that serve
purposive sampling alone). For example, in the intergenerational friendship study additional participants were sampled for based on key transitions in older age, and in the ALS study on the basis of life stage and availability of family support. Theoretical sampling can also be progressed through a variety of means and techniques in the actual data collection process. Most commonly, this is done through interviewing, for example steering questions in the direction of the emergent theorizing (as demonstrated through the first four exemplars). Theoretical sampling can also be done within data after the process of data collection, that is, when the dataset is approached as secondary data (as demonstrated through the fifth exemplar). We now turn to answering the questions posed at the start of this article.

*When should theoretical sampling begin in a grounded theory study?*

Morse (2007) depicts sampling in a grounded theory project beginning with convenience sampling to select first (one or more) participants that are accessible, moving on to purposive sampling to select participants indicated by initial analysis of interviews and shifting to theoretical sampling when emerging theories or categories have developed and the focus has turned to generating data that addresses the need of filling out those categories. Sampling in grounded theory shifts from purposive sampling to theoretical sampling when sampling becomes focused on the construction and/or development of theory (Morse & Clark, 2019). Glaser (1978) states that theoretical sampling should begin after the first codes begin to emerge but there is no definitive agreement in other grounded theory versions on when theoretical sampling should be employed in a study. However, the point where the researcher is honing in on developing patterns and processes in the data i.e. asking ‘what is going on here’ (for example in our studies - why patients with an incurable brain tumor sought to cope or why being reassured was important to people with
ALS) and seeking to actively pursue key concepts and contexts in data generation, is a signal that the study is shifting to theoretical sampling (Charmaz 2014; Corbin & Strauss, 2015). It is also a sign that a grounded theory researcher is exercising theoretical sensitivity (Glaser, 1978).

In the exemplar on intergenerational friendship, once a key process of ‘being friends in action’ was theorized as a process in generated data, the premise for sampling shifted from being to do with participants’ (and/or their friends’) characteristics to instead focusing on sampling for the activities the friendships were happening through. The grounded theory method can invite early tentative theorizing and so a researcher who tunes in to ‘what is going on’ in the data and who accepts the freedom and creativity permitted by the grounded theory method for theorizing about central or meaningful processes early on can discern potential theoretical categories quite quickly (Timonen, Foley, & Conlon, 2018).

An interesting feature of the exemplar on intergenerational friendship is that it demonstrates that shifting to theoretical sampling is not necessarily a discrete stage and in fact theoretical sampling and other sampling methods may intersperse across the dataset. For example, ‘status transitions’ featured as a process of interest that emerged mid-way through the study. When it did emerge, the focus for sampling reverted back to purposive to sample for varied employment related to status transitions. This was necessary to generate the comparative data that could allow a theoretical category capable of explaining this process to emerge. The theoretical category that emerged was ‘being not only old’ and this became a category to be fully conceptualized through the analysis of both existing data
generated prior to the concept emerging, and prospective data generated. Theoretical sampling can be used in parallel with other participant-oriented sampling methods as conceptual categories emerge at different times. The above exemplar also demonstrates - along with other accounts of how and when we theoretically sampled in our grounded theory studies - that theoretical sampling does not only occur in the later stages of a grounded theory study. Rather, theorizing can begin early, using theoretical sampling once a conceptual category of interest is identified.

**How is theoretical sampling different from purposive sampling?**

Flick (2018) details five dimensions to purposive sampling involving purposively selecting i) extreme or deviant cases ii) typical cases iii) cases for maximal variation iv) for intensity of interesting features, processes or experiences and v) critical cases which exemplify relations to be studied. These illustrate how the point of reference for purposive sampling is to saturate the *breadth* of cases or instances. However, theoretical sampling is from its origination guided by a different logic whereby *emergent* categories generated from data can be interrogated and produce theoretical categories capable of explaining the process of interest (Glaser & Strauss, 1967/2012). Theoretical sampling is different from purposive sampling because theoretical sampling is rooted in data collected, the questions that data have given rise to, and the dimensions and contexts that the researcher has pursued to gain a fuller understanding of the process under study. The researcher’s hypothesizing about what directions of inquiry might yield rich comparative data to juxtapose with what has been observed to date, and offer up surprising or anomalous concepts, guides theoretical sampling to fill out the dimensions of these categories. The point of reference becomes the
emergent insights that fill out the dimensions of the category, not coverage of all dimensions of the characteristics of the study population.

For example, in the first exemplar, sampling from the ALS population-based register shifted from purposive sampling based on dimensions of the condition to theoretical sampling based on family context when ‘family support’ emerged as a theoretical construct shaping decisions about care. Theoretical sampling here shifted to being concerned with family context and away from concerns to do with the condition of ALS itself. Here a ‘flash of insight’ from broader existing scholarship prompted us to go conceptually beyond the empirical data to consider the role of context to generate data that allows for this insight to be tested against the empirical data. This reflects a concern to fill out a theoretical category that has the potential to explain why the phenomenon under inquiry presents as it does (rather than a concern to understand the breadth of the phenomenon under inquiry). We see here (as we do in the intergenerational friendship study) how theoretical sampling for contexts can capture and explain processes in data.

Why are abduction and retroduction important for theoretical sampling?

Forming explanatory hypotheses through observation of surprising or anomalous data in the search for novel theory combined with returning to empirical data to test and confirm inferences are central research practices in grounded theory (Charmaz, 2014; Corbin & Strauss, 2015; Glaser & Strauss, 1967/2012; Oliver, 2012). The grounded theory method allows for combining concepts and hypotheses that emerge from data with some existing concepts once all concepts have fit and relevance to emerging data (Glaser & Strauss,
1967/2012). According to Charmaz (2014), grounded theory can expedite theory building but to do so the researcher needs to move quickly through bulky, messy data. Abductive and in some cases retroductive reasoning are the vehicles for this, and theoretical sampling the engine driving the process (Bruscaglioni, 2015; Bunt, 2016).

Abduction helps us to puzzle about the possible antecedents or causes of what we observe in data and orients us to build theory. Retroduction emphasises the broader understanding of the context or conditions where data are being and have previously been collected. We saw within the exemplars presented earlier how insight and thinking about data (e.g. in relation to ‘reassurance’ in the ALS study or ‘needing to cope’ in the experience of glioma study or ‘ambivalence’ in how care workers are construed in the study on care services) involved making creative leaps from the data to generate theoretical categories to think with. Retroductive and abductive reasoning was engaged with in order to move away from the strictures of theoretically naïve inductive reasoning. Theory building from our data referred out to extant concepts and theories and new categories were generated and/or new dimensions of existing categories were developed out of insights emerging from our data and interrogated with other data.

The development of new dimensions for existing categories out of insights emerging from data and being interrogated with data, is clearly seen in the exemplar on intergenerational solidarity. The pattern observed in the data of higher SES older (grand)mothers withholding care labor when asked to help with grandchild care by adult daughters and instead spending time pursuing personal interests and hobbies was anomalous to existing conceptualizations
of solidarity. We drew on insights from the data as well as from extant theory, and abductive reasoning helped us to identify that solidarity as currently conceptualized was inadequate to capture key processes. Theoretical sampling was engaged to explore the effects of not giving and not receiving and how this could be conceptualized. Using theoretical sampling to seek out existing and to generate additional data where withholding of care featured allowed this inquiry to add a novel dimension to the well-established conceptual category of solidarity. Practices previously categorised as not being a dimension of this conceptual category got re-conceptualized by the data to be a dimension of solidarity under the new context of undoing gender norms around family care obligations.

How can theoretical memoing drive theoretical sampling in theory generation?

The work of theoretical sampling is hardly distinguishable from the work of theory construction: seeing connections between concepts, data and categories; and testing, confirming and disconfirming posited explanations. The premises on which each theoretical sampling decision is taken need to be captured in a grounded theory study as they are the building blocks for the emergent theory (Birks & Mills, 2015; Glaser, 2014). Theoretical or conceptual memos are the repositories for the elements of a grounded theory study that work with data. In other words, memos are the designated space in a grounded theory project for the analytical and creative work that the method requires to strive towards novel theory building. Memoing is where the cycling back and forth between data, extant theories and emergent concepts can take place and where hypothesizing about what might deepen, confirm, disrupt or close out the conceptual category or theory building underway is articulated.
The intergenerational solidarity study and the care services for an aging population study are good examples of how memoing in grounded theory drives theoretical sampling in theory generation. In the intergenerational solidarity study, each theoretical sampling move, whether aimed at generating more data or interrogating data to hand, was memoed and the memos captured the direction of our thinking. Memos illustrated when theoretical saturation was reached and how categories that cohered into more fully conceptualized theoretical constructs pertaining to intergenerational solidarity were formed. Importantly, theoretical memos in both studies articulated contradictions in the data that allowed for the emergence of new categories (i.e., ‘withholding expectations of care’, ‘ambivalence’) that could be further conceptualized in order to build theory.

*Can theoretical sampling occur when existing data are analyzed from a new perspective?*

We saw in the exemplar on care services for an aging population how theoretical sampling can be used to good effect when the dataset transpires to have the scope to inform a question and concept not anticipated at the outset of the study. Theoretical sampling *can* be used in secondary analysis, that is, when data are *not* collected and analyzed in tandem. This involves a focus for inquiry emerging from the dataset and in this instance, theoretical sampling can be employed for an exclusively retrospective analysis of the dataset provided that the dataset is sufficiently ‘rich’ to inform the development of new concepts. The inherent creative potential of abductive analysis in grounded theory means that data can yield new, unexpected conceptual insights which theoretical sampling can support the researcher in mining the dataset for. Theoretical sampling in grounded theory secondary
analysis is shown not to be about a procedure that directs the researcher on building a
dataset but a procedure for generating conceptually focused data and for directing a
researcher through a dataset with a conceptual focus (Whiteside, Mills, & McCalman, 2012).

*When can theoretical sampling cease in grounded theory?*

The last sentence which tells us when we have saturated a theoretical construct may seem
like the point at which theoretical sampling finishes - and so it is in relation to that particular
construct. Now a complete rendering can be written up and the researcher has a sense of
having interrogated fully a conceptual dimension of the inquiry. Morse & Clark (2019) state
that theoretical sampling ideally ceases when well-saturated concepts and theory are
comprehensive, have explanatory power, and have theoretical transference (i.e.
applicability) to contexts beyond the study.

Illustrations of how theoretical sampling was conducted across our exemplars show us that
theoretical sampling can taper when no significant new insights from data better describe
and dimensionalize the categories that explain a theoretical construct (Charmaz, 2015). Our
accounts here also demonstrate how theoretical sampling can resurface within a grounded
theory study or dataset at any time. We discussed how theoretical sampling is not
necessarily a discrete stage in a grounded theory inquiry and can be interspersed with other
sampling procedures during the study which yield new theoretical directions which
theoretical sampling will fill out. In this way theoretical sampling can “keep on giving” in
terms of conceptual output – the most welcome yield for any researcher - without having to
“keep on getting” in terms of new participants and data. However, theoretical sampling also
expedites analysis by supporting focused theoretical saturation of analytical concepts to bring data generation, analysis and theory building to an end. Being clear about the central emphasis in theoretical sampling on sampling for concepts over and above sampling for data, supports the realization of a grounded theory project that is manageable, focused, and can be concluded.

**Declaration of Conflicting Interests**

The authors declare that there is no conflict of interest.

**Funding**

The research studies conducted by the authors which they describe here were funded by the following: Health Research Board, Ireland; Irish Research Council; Atlantic Philanthropies; and Horizon 2020 – European Union’s 8th Framework Programme for Research and Innovation.
References


SoCaTel. (n.d.). SoCaTel – Co-creating for a better life. Retrieved October 1, 2019, from https://www.socatel.eu/


