# Conducting a large, longitudinal, multi-site qualitative study within a mixed methods evaluation of a UK national health policy: reflections from the GPED study

## Short title: Conducting a ‘big qualitative study’: reflections from the GPED study

## Key words: qualitative, longitudinal, mixed methods, health policy, qualitative analysis.

Arabella Scantlebury, York Trials Unit, ARRC Building, ARRC Building, University of York, York, YO10 5DD

Joy Adamson, York Trials Unit, ARRC Building, ARRC Building, University of York, York, YO10 5DD

Corresponding author:Joy Adamson, York Trials Unit, ARRC Building, ARRC Building, University of York, York, YO10 5DD. Joy. [adamson@york.ac.uk](mailto:adamson@york.ac.uk)

## Acknowledgements

We would like to thank the ‘GPED qualitative research team’, with special thanks to Dr Helen Anderson and Dr Heather Leggett for your work during data collection and analysis of the GPED project. We would also like to acknowledge the wider GPED study team, particularly Professor Jonathan Benger (CI) and Dr Heather Brant (GPED project manager). Lastly, we would like to acknowledge Dr Laura Sheard, for being a ‘critical friend’ whilst drafting this manuscript.

## Declaration of conflicting interests

The authors declare that there is no conflict of interest

## Funding

The authors received no direct funding for this manuscript. However, the GPED study was funded by the National Institute for Health Research (NIHR) Health Services and Delivery Research (HS&DR) Programme, (grant number 15/145/06). The funders had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; and the preparation, review, or approval of the manuscript. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

## Abstract

Over the past decade, there has been a growing trend towards the use of ‘big qualitative data’ in applied health research, particularly when used as part of mixed methods evaluations of health policy in England. These ‘big qualitative’ studies tend to be longitudinal, complex (multi-site and multi-stakeholder) and involve the use of multiple methods (interviews, observations, documents) and large numbers of participants (n=100+). Despite their growing popularity, there is no methodological guidance or methodological reflection on how to undertake such studies. Qualitative researchers are therefore faced with a series of unknowns when designing large qualitative studies, particularly in terms of knowing whether existing qualitative sampling and analysis methods are appropriate in this context.

In this paper, we use our experience of undertaking a big qualitative study, as part of a national mixed methods evaluation of a health policy in England to reflect on some of the key challenges that we faced in our qualitative study, which broadly related to: sample size, data analysis and the role of patient and public engagement. Underpinning these difficulties was the challenge of being flexible and innovative within the largely positivist research climate of applied health research and being comfortable with uncertainty relating to the three issues outlined. The reflections we present are not to be viewed as a method ‘how to’ guide, but rather as a platform to raise key issues relating to the qualitative methods that we found challenging, in order to stimulate discussion and debate amongst the qualitative community. Through this paper, we therefore hope to demystify what it is like to undertake such a study and hope to spark much needed discussion and innovation to support the future design and conduct of qualitative research at scale.

## Background:

Over the last decade, there has been a rise in the number of large, mixed-methods studies, to evaluate the implementation and impact of national health policies.1-4 These studies tend to involve quantitative analyses of large national, routine data sets, (e.g. ‘Hospital Episode Statistics’, UK) and longitudinal, multi-method (interviews, observations, documents), multi-site qualitative studies, which consist of large numbers of participants (n=100+). These are sometimes referred to as ‘big qualitative studies.’ The underlying principles of mixed method evaluation have previously been used to justify undertaking qualitative research at scale, on the basis that through “triangulation”, or “corroboration” between the two types of evidence, confidence in study findings increases. 1 3

There are established methods for conducting and analysing large quantitative datasets, with these methods playing to the strengths and underpinning philosophy of quantitative methodology. However, the idea that qualitative methods can be used at scale seems to jar with conventional thinking that is, quantitative methods are associated with - ‘breadth’ and ‘generalisability’ and qualitative methods with ‘depth’ and a focus on small scale quality.5 Despite the growing popularity of case study research and of using qualitative data at scale, there remains a degree of elusiveness surrounding how to conduct and analyse ‘big qualitative studies.’ 6

The majority of big qualitative studies that have been conducted as part of wider mixed methods evaluations, have chosen to adapt and apply existing qualitative analysis methods such as thematic analysis and constant comparison to their data sets.1-3 However, methods such as thematic analysis, were designed to tell ‘stories’ through immersive, in-depth interpretation of context bound samples and so their suitability to large, longitudinal, multi-site data sets is relatively unknown. Indeed, even framework analysis,7 a method, which is widely advocated and applied for use in large qualitative studies (n ≈50) was not designed for the complexity and scale of the qualitative studies we are now seeing. This is not to say that these methods do not have a place in analysing large qualitative data, but to highlight that this was not their original purpose.

Methods work that exists on ‘big qualitative data’, has focussed on applying the breadth and depth method to multiple archived datasets and so leaves a void in regards to the practicalities of undertaking data collection and analysis.5 8 The lack of guidance as to how to analyse data from primary big qualitative studies has led to some researchers being analytically innovative and the development of new analysis methods. One example of this is the pen portrait method, which was specifically designed to provide researchers with a way of concentrating large amounts of longitudinal qualitative data, from multiple data sources into a focussed account.9 The pen portrait method was published in 2019, and so there are few published examples to guide those wishing to apply the method to their data. There is also a lack of literature, which describes the practicalities of conducting large qualitative studies either within or outside of wider mixed-methods projects. This is partly because such studies are new and relatively uncommon. However, journal word count restrictions make it challenging for researchers to describe, or reflect on their methods in the confines of large mixed methods papers in sufficient detail. As a result, when embarking on such projects researchers are faced with a list of unknowns that are difficult to pre-empt at grant application stage, and which relate specifically to the practicalities of collecting and analysing large qualitative data.

In this paper, we reflect on our experiences of undertaking the qualitative components of a large mixed methods study, which evaluated the implementation and impact of General Practitioners working in or alongside Emergency Departments (GPED).10-1213 By providing a transparent and reflective account of our experiences of conducting a large qualitative study, we hope to demystify this relatively unexplored area of qualitative research and spark a wider discussion within the qualitative community across three key areas: sample size, analysis and the role of PPI in large qualitative research projects.

## A brief overview of the GPED study

GPED was a policy initiative, which started as a local solution to rising pressure on ED attendance, at a single NHS hospital and gained traction in 2017 when the government published national policy and provided £100 million of capital funding to support the introduction of GPED throughout England.13-15 The idea of GPs working in or alongside ED stemmed from research which suggested that approximately a fifth of patients who attend ED did so with primary care level acuity and so could be managed by a GP. Introducing GPs to work in or alongside the ED was therefore proposed as a solution to rising ED attendances in England as it was believed that doing so would improve patient care and reduce waiting times, unnecessary investigations, hospital admission rates and costs. To provide a robust, in-depth, whilst at the same time service-wide understanding of the impact of GPED, our study consisted of three work packages:10 1113

1. Mapping and describing current models of GPED in all EDs in England. We conducted 67 qualitative interviews. Ten were with policymakers n= 10 (e.g. NHS England, Royal College of Emergency Medicine) and 57 were with local service leaders who represented 64 NHS hospitals in England (e.g. Chief Executives, Clinical leads) to understand the background behind the policy’s development and implementation and benefits that the policy was expected to bring (work package A).
2. Retrospective analysis of routinely available data (Hospital Episode Statistics). Outcome measures included waiting times, admission rates, reattendances, mortality and the number of patient attendances. We also explored potential cost savings.17
3. Detailed case studies at ten case sites. Non-participant observations of 142 individual clinical encounters, 413 semi-structured interviews with key stakeholders (local service leaders, health Professionals (GP and ED staff); patients and carers) and a workforce survey (n=460). 11 12 13 18

## Reflections

For the remainder of this paper we discuss the key challenges that were faced during the qualitative data collection and analysis of the GPED study and how we overcame them. We focus on three key areas: i) how much data do you need; ii) analysis of large qualitative data and iii) the role of patient and public involvement in big qualitative studies. Our aim is not to provide a set of rules, or an exhaustive list of all the challenges that may be faced, but to highlight the areas which we feel warrant further consideration, exploration and discussion to inform the design and conduct of future ‘big qualitative studies.’

## How much data do you need?

The question of ‘sample size’ in qualitative research is widely debated - a quick literature search will suggest that views on what is an ‘accepted minimum’ or ‘ideal number’ varies depending on factors such as the topic, study aim and the methods being used, in addition to disciplinary differences and their associated norms and expectations.19 20 Historically, in particular within applied health research/health services research, a key focus of discussions around qualitative sampling has been around justifying the traditionally smaller samples that are associated with qualitative research and emphasising the need to achieve a varied sample to more quantitatively orientated funding panels, journal editors and colleagues. However, during GPED, our discussions were focussed on deciding when we could ‘stop’ data collection and trying to pre-empt at which point we would have too much data.

We set an ambitious target of conducting approximately 10 case study sites throughout England, which would consist of: 12-15 hours of observation, 10-15 staff interviews, 10-15 patient and carer interviews and a survey of the GPED workforce at each case study site. This approach was based on the number of individuals and stakeholder groups per case site that we felt would give us maximum variation (according to e.g. professional group, gender, years of experience) and enough data for a given site to ‘stand-alone’ whilst enabling detailed cross-site comparisons. These a priori targets also needed to consider and compare longitudinally different types and/or models of GPED service. Our study was designed on the basis that whilst some sites would have a GPED model in place (established sites), the majority would not (prospective sites) and the national policy and injection of significant amounts of capital funding would facilitate the implementation of GPED throughout the UK. Based on the expertise within the GPED study team at the time of protocol development we anticipated that to allow for cross-site comparisons we would need to collect data across approximately six prospective and four established hospital sites in order to accommodate the potential for our ‘best guess’ that approximately 4-6 GPED typologies would be identified.

However, at the time our study was funded, despite being a new initiative, there were different ‘baseline levels of GPED throughout England.’ For instance, some hospitals were using capital funding to introduce a GPED service, whilst others were modifying existing GPED services and some (who did not receive capital funding) had no plans to introduce GPED services. The distinctions between GPED models were therefore less obvious, and far more context dependent than we had first envisaged -there were different interpretations of GPED locally, different levels of local ED demand and variations in local context (e.g. different patient populations and provision of community GP services).

Our solution to these challenges was to alter our approach and be responsive to each case site. Given that the concept of established and prospective sites did not really exist, our focus shifted to understanding and capturing the complexity behind the policy and the different ways that it was being implemented throughout England. For example, changes to GPED services over time were often far lower than we expected from a policy which made available £100million of capital funding to NHS organisations. We therefore adopted a more streamlined and flexible approach to data collection at our time 2 and 3 follow-up visits. This involved interviewing and observing ‘key players’, and being responsive to the needs of each site, whilst being mindful as the project progressed of what data we already had, and which was needed, new or different to add to our GPED story. For instance, in addition to the unforeseen challenges that we faced around our case sites, patient interviews were far less informative, or rich in the detail that we needed for the GPED study. The number of patients available was hugely dependent on ‘who came through the door’ and the GPED model in place at each site. We therefore placed less emphasis on patient interviews and focussed resources on observing and interviewing staff at case sites.

The challenge for the future is how to factor in this need for adaptability and on-the-ground responsiveness in the design and conduct of future big qualitative studies. Our experiences of GPED demonstrate the difficulties of anticipating how much data will be enough for big qualitative studies in advance of data collection, and the need to be comfortable with uncertainty and to work flexibly. However, GPED also taught us that even during data collection, when we had the benefit of having undertaken significant amounts of fieldwork, it was still challenging to determine how much data is enough and at what point it is ‘acceptable’ to adopt a more scaled back approach to data collection. Equally, whilst current guidance rightly urges qualitative researchers to strive for varied samples as opposed to saturation18 there is a need to balance designing a study, which proposes recruitment methods that will facilitate varied samples and account for this need for adaptability, against the requirements of funding panels and peer-reviewers who often expect a-priori sample sizes to be specified and then met. Whilst in an ideal world we would recommend being transparent about this need for flexibility and the uncertainty that comes with the territory of undertaking large qualitative data collection, the palatability of changing your approach to funding panels, peer-reviewers and quantitative colleagues in a mixed method study, is uncertain.

## Trying to see the wood for the trees – analysing big qualitative data

### Which method should I choose?

When designing our study, we suggested that, based on previous published examples of large mixed methods studies we would analyse our data thematically.2-4 21 However, during early familiarisation with our data, it became apparent that this dataset would require more than ‘simply’ applying and adapting a thematic approach. Instead, the number of researchers involved in analysis (n=6), the volume of longitudinal, multi-method qualitative data and lack of transparent examples of how previous big qualitative studies have approached their analysis required us to be analytically innovative.

For the sake of readability, in this paper we describe the key stages of our analysis through a series of steps and also provide an ‘at a glance’ overview of our entire analysis process through figure 1. However, we wish the reader to consider our analysis as an unpredictable and iterative process, which could not be constrained to a single method. Our analysis approach was a balancing act, which involved using a range of methods to provide continuous in-depth explorations of our data, whilst simultaneously reducing its volume down. Indeed, it was only through being both curious and pragmatic that we were able to convey our key messages without losing richness and complexity.

INSERT FIGURE 1

### Framework development – broad brush coding for thematic content

The diversity within our qualitative team was both our greatest strength and at times served as a weakness. The team consisted of individuals with different academic backgrounds (sociology, health sciences), varying levels of qualitative experience and those with and without clinical experience. This variation allowed for different interpretations and much debate, with our individual curiosities and interests contributing to a rounded and rich approach to our data collection and analysis. However, our differences also created tension and disagreements, particularly during the development of our coding framework. As a result, our initial coding framework consisted of a combination of a priori codes relating to the study research questions and inductive codes. (Supplementary file 1) However, the number of different interests and backgrounds at play meant that our initial coding framework was too big and the number of proposed thematic categories made the framework too unwieldy to apply to such a large dataset. Framework development therefore ‘grounded to a halt’ during the early stages of our analysis, as we could not agree on the steer and the focus of the story that we wanted to tell.

In order to find common ground and due to the volume of data we had collected and were still to collect, we re-focussed our efforts towards using the coding framework to ensure we could answer our main research questions; which was particularly important to meet the requirements for the NIHR HS&DR end of study report.11 However, we also ensured that our coding framework was broad enough to capture the data we would need to undertake more in-depth analyses for separate publications, which were driven by the interests of our research team and our data. Crucially, by having a broad coding framework, we avoided having to return to our transcripts and field notes, which given the volume of data we had collected would due to time constraints have reduced the number of separate, in-depth analyses we were able to undertake.

### The pen portrait method

Once agreed by the research team, our framework provided an important first step to our analysis. Through ‘siphoning off’ large portions of our data under broad thematic categories (e.g. local context, service literacy) we gained familiarity and felt less daunted by the volume of data we had to contend with. Initially, we had planned to then thematically analyse the data coded under each broad category separately. However, doing so at this stage was likely to make it difficult to compare and contrast across different time points, stakeholder groups and case sites. Additionally, we were concerned that the amount of rich data that we had collected meant we could easily lose sight of the bigger picture or more crucially become so lost in our data that we would fail to see the overarching story.

Instead, our approach consisted of thematically analysing observation and interview data that had been coded according to our framework to create reasonably short (approximately 10 sides of A4) pen portraits.9 These pen portraits were in-depth summaries of each case site, at each time point which included a diagram of the GPED model, a formal summary or thematic analysis of the key findings for each theme identified in our coding framework and exemplar quotes (figure 2, supplementary file 2). To complement our pen portraits and to avoid losing sight of our overarching messages we also produced a spreadsheet in Microsoft Excel, which documented every theme we had identified from our pen portraits, original coding framework and thematic analysis and a one to two sentence summary of the key messages per theme.

INSERT FIGURE 2

Adopting this approach provided us with a way of keeping the richness of our data, whilst at the same time breaking it down into more manageable chunks – this was crucial for enabling cross-site and later longitudinal analysis. Additionally, by presenting our data at the case site and theme level, pen portraits allowed us to take a step back and as a team decipher the key messages and identify areas that required more in-depth analyses. This ability to both retreat and dive into our data was pivotal to our ability to do justice to the data and made it possible for the team to take the lead on separate, more detailed analyses according to their own interests and skillsets.18 22

From this point on, the pen portraits became the backbone of our analysis as by including exemplar quotes and detailed synthesis of key themes we largely negated the need to go back to the data that had been coded according to our framework. However, when this was required, because our pen portraits were written using the same thematic structure as our framework, we were able to find quotes with relative ease and avoided having to sift through vast quantities of raw data.

### Transitioning from pen portraits to ‘write up’ – the role of an ‘interim-analysis’

Having completed the majority of our initial case site visits and produced pen portraits for these sites, we felt the need to take stock of our findings and consider our next steps. We were however uncertain how to transition from the pen portraits to report writing. Even when only considering the pen portraits for our time one case site visits, we still had a great volume of data – comparable to the amounts of data that is collected in most primary qualitative studies (approximately 10-15 page pen portraits for each of our 10 case sites). We were also mindful that our qualitative data was only part of the picture and that as a mixed methods study we needed to integrate our qualitative and quantitative data. Therefore, we felt that part of this process of taking stock should be to explore how our time one qualitative findings could be used as part of our mixed-method approach.

Similarly to other national mixed methods evaluations of health policy in England,23 we planned to conduct an interim analysis using our time one case site data and interviews with national service leaders (n=10). The aim of this work was to generate a set of hypotheses, some of which then could be addressed using the available HES data, in addition to informing follow-up qualitative case site data collection and analysis. To achieve this, we initially undertook a ‘deeper-dive’ of our data that had been summarised in our pen portraits under the broad category of ‘expectations of GPED.’ However, as this was a more detailed analysis of an individual theme, we also read each pen portrait in its entirety to ensure data pertaining to expectations that may have been captured elsewhere was represented. Alongside this, where more detail was required, we referred to our original data that had been coded under the broad category of ‘expectations’ in our original framework.

This interim analysis was far more complex than anticipated as there was no consensus between and across stakeholder groups and case sites regarding the purpose and anticipated impact of GPED. The ‘neat set of hypotheses’ that we had planned to produce was therefore not possible. Instead, in this analysis (including discussions within the qualitative team and with the wider GPED study group), we identified eight ‘*domains of influence’,* (figure 3)which represented the key areas that GPED was predicted to affect.12 For each domain, we detailed any positive, negative or null changes as a result of GPED and used exemplar quotes from our qualitative data to bring these early findings to life. (See figure 3 for an example of our domain explaining the potential impact of GPED on ED performance).

INSERT FIGURE 3

At this stage, we divided our team so that the majority of our researchers were conducting data collection and producing pen portraits for follow-up case sites. At the same time, a single researcher (AS) took responsibility for summarising the domains of influence and overall synthesis of the project. The benefit of this approach was that AS was in a rare position of having knowledge of the wider GPED project through her involvement in the study management group and in-depth contextual knowledge of case sites and our data through having undertaken data collection and analysis. AS was also, due to the length of the GPED project, one of the few researchers who had been involved in the study from the outset and undertaken data collection at multiple case sites. This was pivotal for ensuring our domains were data driven and balancing the requirement for them to be overarching enough to facilitate our mixed methods analysis, with the need for detail.

Integrating our data across all case sites and time points using our domains of influence was a significant undertaking. This involved completing any final thematic analysis of data from our follow-up case site visits to produce pen portraits whilst simultaneously using our domains of influence as a framework for determining which data should be included in our main analysis. When developing our domains of influence, we produced a ‘mothership’ draft (approximately 15,000 words), which consisted of a detailed write-up of all our findings, based on our time one pen portraits. This document was used to integrate our data across all case sites and time points, as the process involved reviewing our time 2 and 3 pen portraits to identify new and/or divergent information which was then added to this original draft – our mothership draft, ultimately became our final results chapter for our report to the funder.

Throughout our analysis, it was only by using all the analytical approaches outlined above together that we were able to distinguish between the areas we needed to include in our main analysis and themes that required separate, more detailed analyses for ‘spin-off’ journal publications.18 22 For example, taking a deeper dive into one of our ‘broad brush’ thematic categories, relating to the process of streaming patients at the ED front door,18 or concentrating on the perspectives of particular stakeholders across themes e.g. General Practitioners 18 and patients (in development).

### The role of theory – data mapping onto NPT

Whilst integrating our qualitative and quantitative data using our domains of influence for our main analysis, we also identified a need for a higher-level synthesis or greater ‘pulling together’ of our findings. One of the key reasons for this was to ensure that our overarching messages were not diluted and/or lost in the volume of information that was being described in our report to the funder. This was important for readability. Additionally, as the policy we were evaluating was of national interest and was accompanied by a substantial financial commitment, we also needed to ensure that our key messages (i.e. that the policy did not appear to be ‘effective’ in reducing ED pressure and there was great disparity in the way in which it had been implemented) were not lost amidst the detail of our report, or misinterpreted. To achieve this, we used Normalisation Process Theory (NPT),24 as initially outlined in the original project protocol.10

The final stage of our analysis involved ‘mapping’ our qualitative and quantitative findings to the four key constructs of NPT: coherence, cognitive participation, collective action and reflexive monitoring.24 On a practical level, this involved adapting tables published by Murray et al,25 which shows through worked examples how NPT can be used to evaluate complex interventions. To achieve this we used the results chapter from our detailed funder report11 and pen portraits to allocate our themes to each construct and sub-construct of NPT (Table 1). These tables were then refined and used as the main way of presenting our mixed methods analysis in our main results paper.13 Doing so gave us a way of succinctly, highlighting the complexity and summarised findings from our project and it was only at this stage, that we felt we had truly ‘mixed our methods’ and had a clear and distilled sense of our overall message.

INSERT TABLE 1

## The role of Patient and Public Involvement in large qualitative studies

Involving patients and members of the public (PPI) in analysis is something which is being increasingly advocated in health services research. However, the extent of this involvement and role of lay contributors is variable across research projects. In some studies, PPI members essentially become part of the research team and are actively involved in the analysis i.e. through theme development and coding. Others have opted for a more ‘passive approach’ with lay contributors reviewing themes and/or final coding frameworks produced by the research team.

As we have already described our study was complex, and involved a relatively large team of qualitative researchers, multiple data sources and analytical approaches. Our concern was therefore how do we seek PPI involvement, which is not tokenistic, in a study which has over 100 patient interviews and has required the expertise of multiple qualitative researchers to collect and analyse our data? Equally, how would we handle a situation where our PPI group, which consisted of 10 members of the public, who although possessed experience of ED had not attended our case sites, disagreed with a framework we had developed over the course of 12 months? If we adopt a more ‘involved’ PPI strategy, how do we upskill our PPI group so that they are in a position to have meaningful input in such a large, complex and context specific study? As a qualitative team, we also felt strongly that given this was a mixed methods study, the level of PPI involvement should be equivalent across the quantitative and qualitative components of the GPED study.

The project had a nominated lead for all patient and public engagement (an Associate Professor of PPI), PPI activities, initially involved preparatory workshops, which introduced the qualitative and quantitative methods used in GPED. In addition to involvement in the interpretation of our quantitative study, which is described elsewhere11 we held two workshops specifically for the qualitative study. Anonymised transcripts and pen portraits of study sites were circulated to the PPI group, who were asked in advance to read and note any major themes or issues that they could identify and which related to the GPED study’s overarching research questions. The qualitative workshops were subsequently focussed on discussing the public contributor’s interpretations of our data, before discussing the framework developed by the research team and the extent that there was any overlap or differences. Lastly, we held a final mixed methods event, where quantitative and qualitative findings were discussed together with the outcomes of the workshop used as a basis for our main conclusions in our final report 11 and main results paper.13

Our efforts to involve the group in analysis were however less successful. Despite attempts to upskill the group to enable them to support thematic development during workshops we felt this exercise had limited success in terms of providing any new insights into the analysis or interpretation. As a result, we remain to be convinced about lay representatives having an active role in the analysis process, especially in projects where data collection represents the patient voice so comprehensively and is based on a great deal of complex data. Indeed, doing so during GPED created a real tension amongst the research team who felt this involvement undermined their skills as qualitative researchers and the significant amount of work it had taken to distil such a complex data set into a framework and series of tangible overarching messages.

# Conclusion

In this paper, we have reflected on our experiences of the GPED study, a large and complex mixed methods project, which involved a longitudinal, multi-site qualitative component. Our main aim was to shed some light on an area of qualitative research, which although growing in popularity remains from a methods point of view relatively undocumented and discussed – big qualitative studies. We have focussed on what we feel are the key areas of consideration for researchers designing or undertaking big qualitative studies now or in the future: i) how much data is enough, ii) how to analyse big qualitative data and iii) the role of patient and public involvement in large qualitative studies.

The challenges that we faced during GPED reflect areas which are widely debated and often a source of contention when designing and conducting qualitative studies in general. Common to all of the difficulties we faced was a need to be innovative and above all flexible in our approach to data collection and analysis. It could be argued that this uncertainty stemmed from the fact that there are few published examples1-4 and little guidance to inform decision-making for qualitative research at this scale. However, flexibility and adaptation are inherent to ‘good’ practice in qualitative research, which although has grounding principles is by definition not something that is, or can be protocol based. In a recent paper, Braun and Clarke, in describing reflexive thematic analysis touch on this very issue by emphasising that there is no ‘rule book’ or set of stages that a qualitative researcher must follow in analysis. Equally, current guidance on sampling advises against striving for saturation and pre-specifying recruitment targets in qualitative research.19 This creates an issue in practice and jars with the current research climate, which remains largely positivist in thinking and structure. For example, qualitative researchers are required to specify recruitment targets to satisfy funding panels and reviewers. Additionally, there is a practical need to state the number of participants to be interviewed or observed at the grant application stage – in order for the reviewers to be satisfied that this aspect has been given due consideration and justification and to ensure projects are appropriately resourced. As a result, a priori recruitment and sampling strategies and indeed analysis plans are often rough guestimates, which will need to be adapted during the data collection process. This conflict was amplified during GPED, which as a wider mixed methods study meant we were often asked to justify both internally and externally why we were ‘deviating from our protocol.’ For example, why had we not undertaken the exact numbers of interviews and observations that we had specified in our protocol. We were dependent on our experiences, disciplinary backgrounds and ‘gut instinct.’ Ultimately, it was a lack of convention and our constant critique and refinement of our sampling and analysis, which were pivotal to our success and it is crucial that future big qualitative studies are designed and funded with flexibility in mind.

Based on our experience we certainly would not advocate that ‘bigger is better.’ We do not dispute the fact that the amount and varied nature of our data enabled us to obtain a rich and detailed understanding of GPED nationally and that this data was far more complex and rich than any data that we had obtained before. However, there is a need to balance this against how much data is needed to obtain this level of understanding and how much data a single research team can realistically analyse in a single project, for little additional gain. In our experience, it was variation in terms of stakeholders and case sites that gave us this understanding not volume and this is consistent with current qualitative methodological guidance.19 This variation was achieved through adapting our sampling strategy as our knowledge of case sites and the subject area progressed and to a large extent streamlining our approach to one which ensured we spoke to and observed ‘key individuals’ rather than a large homogenous group.

It is also important for researchers to remain open minded when analysing large qualitative projects and to not assume that current analysis methods can be ‘scaled up’ and applied to qualitative data sets of this size. Adopting a ‘trial and error’ approach to our analysis and using multiple analytical approaches was crucial and gave us different ways of exploring our data and analytical lenses to look through. For instance, our domains of influence were crucial for facilitating our mixed methods analysis and providing a means of navigating our data, whilst Normalisation Process Theory provided a thread for transparently describing the complex GPED policy and our overarching messages. Irrespective of the stage, or method that we used, we found the need to continually distil our data down to manageable chunks, whilst not losing sight of the bigger picture and always keeping a ‘route back to our data’ crucial. The latter was particularly important, but difficult to achieve in practice as it was not always possible to know or pre-empt which avenues may warrant further exploration. Our analytical approach was therefore entirely experimental and one which drew on our experience and a range of different methods. As a result, we acknowledge the imperfect nature of our approach and that further work is needed to trial and then reflect on different methods for analysing large qualitative studies to pave an easier and more streamlined path for future researchers.

When developing the GPED PPI strategy there was an assumption that an obvious place for PPI involvement was in informing the design and analysis of the qualitative study. Although a polarising opinion, our experience is that ‘true’ PPI involvement, within the process of analysis of large qualitative studies is difficult to achieve. In these situations, meaningful involvement of a truly lay PPI group is difficult as by definition these individuals do not have the research, or analytical expertise required and assuming so arguably undermines the skill involved in the analysis and interpretation of qualitative research. This creates practical challenges for researchers that are difficult and sensitive to navigate. Using GPED as an example, we were in some ways fortunate that our patient group were in agreement surrounding our key findings as we are unclear how we could have managed a situation where a PPI group (n=10) had disagreed with the opinions of such a large and varied patient sample and months of qualitative analysis. For instance, should researchers in this situation disregard lay contributors opinions, or include them and then undermine the analysis and data that have taken years to collect and interpret? Further exploration of the most appropriate and meaningful use of PPI groups in studies such as GPED is needed to ensure that we do not use lay representatives to ‘validate’ our analysis and encourage the development of meaningful involvement.

Lastly, it is interesting to note that in such a high profile project, where there was no ‘rule book’ to follow, we received perhaps the least scrutiny from peer reviewers of any project we have undertaken. There was no critique regarding our methods and our results chapter required only small clarifications. On this basis, it could be assumed that there is an inherent trustworthiness and comfort with ‘big’ qualitative studies, speaking to our earlier notion that we are working in a largely positivist-minded research environment. There is value in collecting vast amounts of qualitative data as certainly in our case, this gave us perhaps the richest and complex data we had ever collected and with it a thorough understanding of our topic and the potential to explore other unanticipated avenues of interest through additional outputs (spin-off papers).18 22 However, this complexity is challenging and collecting data on this scale can be a hindrance not only for identifying key messages, but in creating practical problems even in a study what was as well funded and resourced as GPED. During and indeed after GPED, we often ask ourselves if we have done our data justice and all agree there are swathes of unexplored areas of interesting and important data that we will sadly never be able to analyse to their full potential.

We therefore urge funding panels, researchers and reviewers in the future to disregard the inherent trustworthiness that seems to accompany large qualitative studies and the ‘bigger is better ethos.’ Based on our experience, we do not feel that our overall messages would have been different with less data, or less trustworthy, but that we would merely have fewer examples to ‘back up’ our conclusions. In the future, we urge the research community, particularly at a time of austerity, to fund more streamlined, targeted data collections, and/or have confidence in studies which propose flexible, adaptable data collection and analysis plans that are responsive to the unique challenges of their project and to encourage methodological innovation, particularly, to support the analysis of large, complex datasets.

## References

1. Greenhalgh T, Stramer K, Bratan T, et al. Adoption and non-adoption of a shared electronic summary record in England: a mixed-method case study. *Bmj* 2010;340

2. Pope C, Turnbull J, Jones J, et al. Has the NHS 111 urgent care telephone service been a success? Case study and secondary data analysis in England. *BMJ open* 2017;7(5):e014815.

3. Dixon-Woods M, McNicol S, Martin G. Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. *BMJ quality & safety* 2012;21(10):876-84.

4. Sheikh A, Cornford T, Barber N, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in “early adopter” hospitals. *Bmj* 2011;343

5. Davidson E, Edwards R, Jamieson L, et al. Big data, qualitative style: a breadth-and-depth method for working with large amounts of secondary qualitative data. *Quality & Quantity* 2019;53(1):363-76.

6. Eisenhardt KM. Building theories from case study research. *Academy of management review* 1989;14(4):532-50.

7. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. *The qualitative researcher’s companion* 2002;573(2002):305-29.

8. National Centre for Research Methods and Economic and Social Research Council. Big Qual Analysis Resource Hub [Available from: <http://bigqlr.ncrm.ac.uk/> accessed 21/04/2022 2022.

9. Sheard L, Marsh C. How to analyse longitudinal data from multiple sources in qualitative health research: the pen portrait analytic technique. *BMC medical research methodology* 2019;19(1):1-10.

10. Morton K, Voss S, Adamson J, et al. General practitioners and emergency departments (GPED)—efficient models of care: a mixed-methods study protocol. *BMJ open* 2018;8(10):e024012.

11. Benger JB, H; Scantlebury, A; Anderson, H; Baxter, H; Bloor, K; Brandling, J; Cowlishaw, S; Doran, T; Gaughan, J; Gibson, A; Gutacker, N; Leggett, H; Liu, D; Morton, K; Purdy, S; Salisbury, C; Vaittinen, A; Voss, S; Watson, R; Adamson, J;. General Practitioners working in or alongside the emergency department: the GPED mixed-methods study. *Health Services and Delivery Research Program 10 (30)*

12. Scantlebury A, Brant H, Anderson H, et al. Potential impacts of general practitioners working in or alongside emergency departments in England: initial qualitative findings from a national mixed-methods evaluation. *BMJ open* 2021;11(5):e045453.

13. Scantlebury A, Adamson J, Salisbury C, et al. Do general practitioners working in or alongside the emergency department improve clinical outcomes or experience? A mixed-methods study. BMJ Open. 2022; 12

14. HM Treasury. Spring Budget. London, 2017.

15. Department of Health and Social Care. Funds for next winter will ease pressure on accident and emergency (A&E) departments in England. London, 2017.

16. Iacobucci G. All emergency departments must have GP led triage by October: British Medical Journal Publishing Group, 2017.

17. Gaughan J, Liu D, Gutacker N, et al. Does the presence of general practitioners in emergency departments affect quality and safety in English NHS hospitals? A retrospective observational study. *BMJ Open* 2022;12(2):e055976. doi: 10.1136/bmjopen-2021-055976

18. Anderson H, Scantlebury A, Leggett H, et al. Factors influencing streaming to General Practitioners in emergency departments: A qualitative study. *International Journal of Nursing Studies* 2021;120:103980.

19. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative research in sport, exercise and health* 2021;13(2):201-16.

20. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. *Field methods* 2006;18(1):59-82.

21. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology* 2020:1-25.

22. Anderson H, Sacntlebury A, Leggett H, et al. Perspectives of GPs working in or alongside Emergency Departments in England: Qualitative findings from the General Practitioners and Emergency Departments [GPED] Study. *British Journal of General Practice* 2022:BJGP.2021.0713. doi: 10.3399/bjgp.2021.0713

23. Robertson A, Cresswell K, Takian A, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation. *Bmj* 2010;341

24. May CR, Mair F, Finch T, et al. Development of a theory of implementation and integration: Normalization Process Theory. *Implementation Science* 2009;4(1):1-9.

25. Murray E, Treweek S, Pope C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. *BMC medicine* 2010;8(1):1-11.