SIX

Analysing data

Introduction

In general, the analysis of data may be both the most specialised and the least well understood aspect of making research. A common failing of research reports and journal articles is not to explain the process of analysing data clearly enough for readers to gauge their level of confidence in the findings or for researchers to replicate the analytic method (Odena 2013: 364).

There are many ways to analyse any given set of data. Suppose that you hold a focus group with eight first-generation immigrants from different countries of origin. You begin by having each person share some basic demographic data by way of introduction: where they have lived, how old they are, their occupation(s) before and after immigration, who and where their family members are. Then you facilitate a discussion of their experiences of emigration and immigration around themes drawn from the academic literature, including wealth and poverty, coercion and freedom, belonging, emotion, status, togetherness and separation. The resulting data would be amenable to quantitative and qualitative analysis. In quantitative terms, you could do only descriptive statistical analysis, as the size and nature of your sample would not support inferential statistics. But it might be interesting to calculate such things as the length of people's journeys; the similarity or difference of participants to the national picture of immigrants; the variance in distances between family members. In qualitative terms you could of course focus your analysis on the themes from the academic literature that you used to facilitate the discussion in the first place. But you could also:

- use a recognised analytic technique, such as interpretive phenomenological analysis (Finlay 2011: 140)
- look at the metaphors people used, to see what they might tell you (Fletcher 2013: 1555–6)
- analyse interactions between people in the focus group to find out what those add to the analysis of data content (Farnsworth and Boon 2010; Halkier 2010; Belzile and Öberg 2012)
- consider any silences, pauses or omissions in order to try to uncover what might have been left unsaid and why (Frost and Elichaoff 2010: 56)
- ask someone else to analyse the data independently to see whether or not they reach the same conclusions as you (Odena 2013: 365)

- involve your participants in the analysis, for confirmation and reciprocal learning (Nind 2011)

and so on.

It is not possible, in a single book chapter, to explain the many different approaches to analysing data in sufficient detail for readers to use them all effectively. The aim of this chapter is to give you an overview of some of the more creative approaches to data analysis, together with an understanding of where rules must be applied. This will help you to identify areas of data analysis that you would like to investigate further, and provide some signposts to enable you to do so.

Ethics in data analysis

As you analyse your data, you're responsible to a lot of people: participants, funders, commissioners, supervisors, examiners and so on. This applies whether you're analysing primary or secondary data, or both. Even analysis of historical or archival data can involve responsibilities to the descendants of participants (Seal 2012: 689–90) and can be emotionally demanding for the researcher (Einwohner 2011: 423; Seal 2012: 689).

Data analysis is difficult and the process can seem impenetrable. It is essential that you do not invent or distort your data, or misuse statistical techniques (Poon and Ainuddin 2011: 307). For example, some people who are new to the quantitative analysis software SPSS decide to run all the tests it can do. In this way you are likely to find one or more which give a statistically significant result – if you're using a 5% level of probability, then five from 100 will do so by chance. This kind of 'fishing' is highly unethical. The key is to know which statistical test or tests are appropriate to use for your data-set (Davis 2013: 17).

An interesting **panel discussion** with American academics and practitioners about data, analysis and ethics can be viewed online.

Data preparation and coding

Meticulous data preparation is essential; there is not much scope for creativity in accurate transcription or data entry. Coding data can also feel quite tedious and may be very time consuming. When it has been prepared and coded, data usually needs to be sorted into categories and sub-categories, a process that can become very complex (Mason 2002: 151).

However, even these apparently repetitive and laborious processes require some creativity. For example, take the transcription of data recorded by audio or video. There are a large variety of decisions to be made about transcription, and there is no 'best way' or 'right answer' (Hammersley 2010: 556). These decisions include such things as: should you record non-speech sounds that people make, such as laughter, coughs, sighs and so on? If so, how? Do you record pauses? If so, do you measure their length, or just note each occurrence? When transcribing video data, should you include body movements, gestures, information about the surrounding environment? How do you lay out your transcription on the page, and how do you identify the different speakers/actors in the transcript? (Hammersley 2010: 556–7). Similar decisions are required for quantitative data: is it in a spreadsheet already? If so, is the spreadsheet fit for your purposes? If not, what do you need to do to make it so? If the data is not in a spreadsheet, how can you construct one to facilitate your analysis? How do you code missing data? And so on. Gathering data online can often avoid the need for much, if any, data preparation (Stacey and Vincent 2011: 621), but, in these cases, care must be taken to set up your data gathering in such a way that the format of the data gathered will enable the necessary analysis.

Quantitative versus qualitative data analysis

Quantitative and qualitative data need to be analysed separately, using different techniques, and in research where both quantitative and qualitative data have been gathered, the datasets will be analysed separately before the analyses are integrated to produce the research findings.

Quantitative data analysis involves the use of statistical techniques to describe data, compare different groups of data and make inferences about populations from random samples. There is no room for creativity in the actual calculations, unless you are a skilled statistician with enough knowledge and experience to move the field forward by, for instance, creating a new algorithm (for example, Mulder 2011: 15). There is also scope for creativity in taking an analytic technique developed outside social science and using it within social science. For example, sequence analysis was first developed in biological research to compare DNA sequences and is now used in social research for analysing sequences that have 'a specific order of crucial importance that cannot be changed' (Brzinsky-Fay and Kohler 2010: 360), such as life histories and career trajectories. But again, considerable expertise is required to identify analytic techniques from one field that can be usefully applied in another.

Chapter Five mentioned the research of UK and US researchers Emma Tonkin, Heather Pfeiffer and Greg Tourte (2012), who gathered 600,000 tweets about the London riots of August 2011. Their approach to analysis was interesting. They began by identifying duplicate tweets (including retweets). Then they identified references to other participants in the social network, which enabled them to create a graph of participation. They also identified recognisable individuals and locations, repeated phrases or hashtags and URLs. The researchers then used natural language processing (that is, computerised analysis of language as it is used) to identify interesting terms and index tweets containing those terms. That enabled them to create frequency tables of those interesting terms, which meant that they could identify popular topics and show how those topics changed over time (Tonkin, Pfeiffer and Tourte 2012: 52).

What is important in quantitative data analysis is to understand the rationale behind any statistical technique that you might consider using, and so be sure that you select the right technique(s) for the question you are aiming to answer. For most people, creativity in quantitative data analysis lies in deciding which techniques to use, and how, in the context of each unique research project – and, of course, in interpreting the results (Bryman 2012: 592). However, it has to be said that even professionals such as quantitative sociologists don't always understand the purposes of the statistical tests they use (Engman 2013: 257). Also, research has shown that quantitative analysts are not always skilled in interpretation (Laux and Pont 2012: 3).

There is more scope for creativity in qualitative data analysis. There are several kinds of qualitative data analysis, including:

- content analysis – a semi-quantitative technique for counting the number of instances of each category or code (Robson 2011: 349)
- thematic analysis – identifying themes from coded data (Robson 2011: 475)
- narrative analysis – analysing stories from primary or secondary data (Bryman 2012: 582)
- conversation analysis – detailed analysis of the verbal and non-verbal content of everyday interactions (Bryman 2012: 527)
- discourse analysis – analysing patterns of speech and interaction in a detailed and sometimes semi-quantitative way, for example by measuring the length of pauses (Bryman 2012: 529)
- metaphor analysis – analysing metaphors from primary or secondary data (Fletcher 2013: 1555–6)
- phenomenological analysis – analysing participants' stories from, and descriptions of, their 'life-worlds', or individual experiences and perceptions, with a focus on meaning (Papathomas and Lavallee 2010: 357; Mayoh, Bond and Todres 2012: 28)
- life course analysis – analysis of the 'interaction between individual lives and social change' (Brittain and Green 2012: 253).

This is by no means an exhaustive list, but is intended to illustrate the range and diversity of approaches to data analysis.

UK researchers Ian Brittain and Sarah Green used life course analysis to study the rehabilitation of former soldiers after disabilities sustained in combat. The 'life course' is the sequence of different roles and situations an individual finds themselves in over time. The life course exists in a wider historical and socioeconomic context, containing systems of opportunities and constraints within which individuals can make choices and create their own life journeys

(Brittain and Green 2012: 253). The researchers collected relevant newspaper and internet articles from around the world, quoting former soldiers speaking about their rehabilitation. They read each article several times, to familiarise themselves with its content, and then carried out a more detailed analysis, noting preliminary comments, associations and summaries. These notes were used as guidance in identifying themes, which were clustered into subordinate and overarching superordinate themes before being collated and discussed individually to deepen the analysis (Brittain and Green 2012: 255–6).

Secondary data

Because there is so much that can be done with any dataset, and because data gathering can be onerous for participants, researchers' attention has turned more and more to the opportunities offered by secondary data – that is, data previously gathered for some other purpose (sometimes research, sometimes not) and that can be used again.

Greek researcher Helen Briassoulis researched tourism and development by analysing an **online petition** to stop the creation of a golf course at Cavo Sidero in Crete. The petition website contained informative text and video, plus the details of over 10,000 signatories from around 25 countries, and around 4,000 comments from signatories ranging from a single word to several hundred words. Where possible, signatories' identities (academic, tourist, golfer and so on) were deduced from their comments or from wider web searches (this was not possible in all cases because signatories could choose to remain anonymous). Also, comments were placed into categories, which were initially defined from the theoretical framework for the research and then refined during the process of analysis. This approach to analysis helped the researcher to understand 'the patterns and determinants of opposition to golf development' in local and global, specific and general terms, over time and from diverse socio-cultural contexts (Briassoulis 2010: 724).

For researchers used to gathering their own data, or for those who find the prospect appealing, the idea of working with secondary data may seem less attractive. Some may fear that it would feel too clinical or distant if they didn't have intimate personal knowledge of the context in which the data was gathered. Indeed, for some research disciplines, such as ethnography, it is essential for researchers to gather their own data (James 2013: 564). Being present for data gathering may add layers of sensory experience that wouldn't otherwise be available, but that still doesn't mean the researcher knows everything (James 2013: 567). Analysis of secondary data can be just as creative as analysis of primary data, requiring judicious use of the researcher's analytic imagination (James 2013: 570). Careful and rigorous analysis of the same dataset as secondary and primary data may lead to different insights, but that doesn't mean either set of findings is 'wrong' (James 2013: 574).

Some of the other pros and cons of secondary data are conveniently listed on a **web page**, while another web page provides access to some **online sources of secondary data** – these are UK-generated, but other countries will have close equivalents for most cases, which should be relatively easy to find online.

Analysing documentary data

Documents are not only containers of data, they can also be tools for people to use as they act in the world (Kara 2012: 126), as the following examples show.

- Legal judgements are used by lawyers as benchmarks to assess new cases.
- An individual may use their 'last will and testament' to benefit a charity.
- An organisation's statement of customer service standards may be used by customers in negotiations with that organisation.

The analysis of documents will benefit from taking this into account. There are three steps to analysing documents: superficial examination, or 'skimming'; thorough examination by careful reading and re-reading; and interpretation (Bowen 2009: 32). During this process, 'meaningful and relevant passages of text or other data are identified' (Bowen 2009: 32) and patterns, categories and themes can be found within the data. In a mixed-methods study it is also possible to apply pre-existing codes, for example those used with other datasets in the study such as interview transcripts, to documentary data (Bowen 2009: 32). This can be a useful technique for data integration.

John Vincent, from the US, and Jane Crossman, from Canada, used textual analysis to examine gendered narratives and nationalistic discourses in Australian newspapers' narratives about Australian tennis players Lleyton Hewitt and Alicia Molik during the centennial Australian Open Championships. The researchers collected three national daily papers with extensive sports coverage, chosen to appeal to three different types of readership, from the day before the championships began to the day after they ended. They found 108 articles focusing on Hewitt and 79 focusing on Molik. Each was read twice and narratives relating to gender and nationality were highlighted, and the articles were then transcribed as MS Word documents. Then the researchers used open and axial coding to generate 'multiple and layered elements' of analysis (Vincent and Crossman 2009: 264). Open coding was used to organise the raw data into themes and categories by searching and re-searching the transcripts for dominant narratives, contradictions and inconsistencies. Axial coding was used to link these themes and categories with each other and with individual codes. When the coding was done, the researchers used a multi-theoretical framework focusing on gender, power and nationality to focus and amplify the findings of the research, to 'uncover the textual constructions of gender and national identity permeating the dominant discourses' (Vincent and Crossman 2009: 265). Vincent and Crossman found that multiple levels of coding, combined with a well-developed theoretical framework, offered a rigorous and systematic approach to analysis that resulted in a 'dynamic and layered analytical framework that led to theoretical and data-driven insights' (Vincent and Crossman 2009: 265).

Analysis of talk

There are two central methods of analysing talk: discourse analysis and conversation analysis. While neither method is new, both are highly creative, with scope for further creativity in finding new ways to use and develop each method.

Conversation analysis (CA) is an evolving analytic method based on the idea that any verbal interaction is worth studying to find out how it was produced by the speakers (Liddicoat 2011: 69). CA requires a detailed form of transcription, capturing not only the words that are spoken but also aspects of talk such as intonation, volume of speech, pauses, non-word utterances such as 'um' and 'er', overlapping talk, interruptions and non-verbal sounds such as laughter or coughs (Groom, Cushion and Nelson 2012: 445). The aim is to facilitate a thorough analysis of people's conversation in normal everyday interactions, perhaps focusing on specific types of interaction such as greetings or leave-takings. CA has also been used to study people's talk in more artificial situations such as research interviews (Groom, Cushion and Nelson 2012: 440). Unlike discourse analysis, which focuses on talk within context and structure, CA focuses on what people actually do as they talk (Liddicoat 2011: 8). When it was first devised in the 1960s–70s, CA was used in isolation from theory, but more recently it has been situated within theoretical frameworks around topics such as power and identity (Groom, Cushion and Nelson 2012: 446). CA is a demanding and time-consuming analytic technique (Mercer 2010: 8), although it doesn't require the gathering of much data (Bryman 2012: 525). There is considerable scope for creativity in using CA with different theoretical frameworks and as part of mixed-methods investigations.

Discourse analysis (DA) is based on the concept that the way we talk about something affects the way we think about that phenomenon. 'Discourse' in this context doesn't refer solely to talk itself, it refers to talk that is constructed within the constraints of a social structure. DA can be applied to other kinds of data, such as written texts (Bryman 2012: 528) and images (Rose 2012: 195). CA and DA are not mutually exclusive; it can be helpful to integrate CA into DA for a more detailed analysis of talk or texts (Bryman 2012: 528).

Michael Corman, from the University of Calgary in Qatar, took a constructivist approach to DA in studying the way mothers talk about placements outside the home for their autistic children. Nine mothers from western Canada had taken part in semi-structured interviews in their homes for an earlier study. Corman extracted relevant sections from those interview transcripts to use as secondary data in investigating 'how their talk accounted for placement and their social reality surrounding placement' (Corman 2013: 1323). He used predetermined questions to orient himself as he analysed the data.

- What is the talk of participants accomplishing?
- Why is the subject matter being brought up now and in this way?
- How is participants' talk being used to make claims?
- How do participants make their talk persuasive?
- What discourses do participants invoke to talk about placement, and why? (Corman 2013: 1324)

This analysis allowed Corman to make visible the ways in which mothers constructed their own realities by talking about, and making meaning of, their experiences. This in turn enabled an increase in understanding of the mothers' stress factors and coping mechanisms.

Some researchers choose to gather talk from people in more natural settings.

In the year 2000, four dual-income American families with at least one child were asked to audio-record as many of their interactions as possible for one week. This led to over 450 hours of recordings, and the transcripts ran to over 1 million words. The study was designed by American researchers Deborah Tannen and Shari Kendall, and the aim was to find out how women and men talk at home and at work, and how they use language to balance work and family life. Part of the rationale for the methodology was that self-recording over such a long period would lead participants to become habituated to the recording device, and indeed the intimate nature of some of the data suggests that this did happen to some extent. However, this was not the whole story. Cynthia Gordon, a member of the research team, re-examined all the transcripts for evidence of times when talk focused on the recording device. She found that all participants focused on the recording device at times, giving it different roles such as 'burden', 'spy' or 'audience', and effectively using it as a resource within their interactions (Gordon 2013: 314).

Gordon's research is creative because she takes a fresh look at an aspect of the research process, recording of data, which is often taken for granted or viewed as neutral by researchers. Gordon's analysis shows that research participants have a very different perspective, which is helpful in enabling researchers to think differently about this basic tool of our trade.

Visual analytic techniques

Diagrams and maps can be particularly useful in data analysis to help you visualise your data and the ideas and relationships that develop as you work through the analytic process. Maps have been used in this way within a range of disciplines including geography, psychology, sociology, anthropology and education (Powell 2010: 539–40). Diagrams have been used in grounded theory analysis for many years (Strauss and Corbin 1998: 12) and are also relevant to other forms of data analysis. These visual techniques help the researcher to move from coding or theme identification to conceptualisation (Strauss and Corbin 1998: 218). They are also great vehicles for using, and stimulating, creativity and imagination (Strauss and Corbin 1998: 220).

UK researchers Charles Buckley and Michael Waring used diagrams at various stages of grounded theory studies of children's attitudes to physical activity (Buckley and Waring 2013). At the analytic stage, they found that creating diagrams helped them to generate, explore, record and communicate insights about their data. Drawing on the work of Clarke (2005), they also suggest that using diagrams in data analysis can help to uncover some otherwise hidden parts of the research process, and so rebut potential accusations of reductionism (Buckley and Waring 2013: 150). 'During the process of research, the use of diagrams can help the researcher make sense of relationships that may not have been previously explicit. In this way, they become an active part of the theory generation and not only support developing conceptualisation but also actively encourage clarity of thought.' (Buckley and Waring 2013: 152).

Diagrams can of course be created by hand, or using specialist diagram software such as Gliffy, or research analysis software that supports diagramming such as NVivo. Similarly, maps can be drawn by hand or using specialist mapping software such as Esri or Maptitude.

Analysing video data

Video offers a myriad of possibilities, and enormous challenges, to the data analyst. 'Video allows us to document time in a complex fashion: action presents different simultaneous layers of temporal conducts – such as talk, gestures, gaze, body movements, postures of all of the participants ... One of the challenges of video recording social action is precisely the continuous documentation of all of these layers of timed action – which is often impossible to achieve with one camera and difficult to solve with several' (Mondada 2012: 305–6). It is also impossible to transcribe everything that could possibly be relevant: all the physical movements and gestures, directions of gaze and eye contact, handling of material objects, use of technology, details of the environment and so on (Hammersley 2010: 566).

Jessica DeCuir-Gunby and her colleagues in America gathered video data from three cohorts of teachers in their longitudinal study of mathematics teachers' professional development. Mathematics lessons were video-recorded, then coded using a three-step process.

1 Lesson mapping – description of each lesson's organisation and structure, with categories based on teachers' interactions in the classroom, for context and to identify changes over time.
2 Lesson rubric coding – quantitative examination of teacher-initiated verbal communication in specific categories such as 'language matching' (teacher using child's own language) and 'illuminating thinking' (teacher drawing attention to and/or highlighting child's understanding), to capture the number of times each category was used by each teacher.
3 Transcription of verbal communication from the lesson rubric, to capture what was said within each category.

ANOVAs and Bonferroni tests were used to analyse quantitative data from this coding system and identify any significant differences between the three cohorts of teachers. Qualitative data was analysed using a five-step method for each individual lesson.

1 Use lesson mapping categories to provide structure and framework for the lesson.
2 Pair lesson rubric codings within events highlighted by lesson mapping.
3 Place comments from field notes within lesson mapping categories.
4 Match events of lesson with the teacher's statements from group interview.
5 Integrate first four data sources to create individual cases.

The quantitative analysis enabled DeCuir-Gunby and her colleagues to provide an overall view of what happened in the cohorts' classrooms, while the qualitative analysis enabled them to scrutinise each individual teacher and lesson (DeCuir-Gunby, Marshall and McCulloch 2012: 212). This mixing of methods also enabled the researchers to identify and describe several aspects of complexity within the data, such as instances of some teachers' data corroborating, complementing or contradicting other data (DeCuir-Gunby, Marshall and McCulloch 2012: 207).

The wealth of video freely available on internet sites such as YouTube means that scholars are increasingly turning to such sources for information and data (Kousha, Thelwall and Abdoli 2012: 1710). The analysis of video data enables researchers to examine aspects of social practice that it would be difficult or impossible to study in any other way. Examples include the way that architects use sketches and other drawings to help them think and communicate as they collaborate (Mondada 2012: 317–22) and the processes of informal interaction and tacit participation that enable people with different roles in emergency call centres to work together flexibly and effectively as they respond to the needs of people in crisis (Fele 2012: 281).

Mixed-methods analysis

As shown above, it is possible to use quantitative and qualitative analytic techniques in the same piece of research, and this can enrich your findings. For example, cultural consensus analysis, a quantitative method, can be combined with cultural modelling, a qualitative method (Fairweather and Rinne 2012). Cultural consensus analysis asks three key questions about sharing of culture and then assesses the patterns in the data using mathematical techniques (Fairweather and Rinne 2012: 477). Cultural modelling is based on DA, which enables researchers to understand participants' perspectives on thoughts, knowledge and the meaning of language, and is used to demonstrate and explain relationships between cultural elements in the data (Fairweather and Rinne 2012: 482). So, both analytic techniques are used to investigate the extent to which culture is shared, albeit in different ways, and using both together can give a more complete picture than using one alone (Garro 2000: 285).

Reesa Sorin and her colleagues in Australia developed an analytical procedure using three different methods to analyse children's artworks. The first method was content analysis, a quantitative technique: the researchers developed categories for the salient features of children's drawings such as animals, houses and trees. The items in each category were counted for number and frequency. The other two methods were qualitative. One was interpretive analysis, in which categories were again identified, this time based on the mood or atmosphere of each drawing and the story the child told about their drawing. The other was developmental analysis, which suggests that stages in the development of children's artworks can be correlated with their ages. The researchers conclude that this combination of analytic methods can 'provide deep insights into young children's understandings' (Sorin, Brooks and Haring 2012: 29).

Q methodology offers another way of using statistical analysis with qualitative data about people's views, attitudes, beliefs and emotions (Ellingsen, Størksen and Stephens 2010: 395). A short **video** introducing Q methodology can be viewed online.

Alternatively, research can involve more than one type of quantitative analysis, or more than one type of qualitative analysis, conducted either concurrently, or consecutively in an iterative approach.

Erica Halverson and her colleagues in America studied four youth media arts organisations across the US to investigate how video could be used to represent young people's identity. They describe video data as 'multimodal' because it contains still and moving images, colour, a range of sounds and silences, sometimes text and so on. Halverson et al originally approached video analysis by starting with dialogue, but then they encountered a film that had no dialogue, which engendered their decision to develop a multimodal approach. Their aim was to create a multimodal analytic framework, not to analyse data in different chunks, but to reflect how the interaction of different chunks of data can create new meanings. Following the work of Baldry and Thibault (2006), they divided the film into 'phases and transitions', which were units of analysis that had some kind of internal consistency, for example through a type of shot, a consistent voiceover or the same music. Then they devised a coding scheme, based on the work of Bordwell and Thompson (2004), for each unit of analysis. This involved four broad categories based on filmmakers' key cinematic techniques:

1 mise-en-scène: anything visible within the camera's frame, such as setting and characters
2 sound: anything audible, such as dialogue and music
3 editing: the filmmaker's interventions that create the film
4 cinematography: the filmmaker's techniques for altering the image from that seen through the camera's lens.

Within each category, more detailed codes were developed, such as facial expressions, clothing, sound effects, flashback, freeze frame, lighting and close-up. Halverson et al say that using this system 'to describe the phases and transitions of the films resulted in the creation of multilayered filmic transcripts that allow us to consider each mode individually, as well as how they connect to one another to help youth consider issues of identity in their films' (Halverson et al 2012: 8).

Doing mixed-methods analysis well can be resource-intensive and time consuming, particularly in international research.

Anne Shordike and her colleagues in America, Thailand and New Zealand investigated the meanings of celebratory food preparation for older women in three different cultures, to find out what commonalities might exist across different cultural contexts. They developed a research design in which researchers from each country would gather, analyse and report on the topic from their own country, before they made comparisons between the countries. Data was gathered using three focus groups in each country in 2000–01, and was transcribed and analysed. Coding data collaboratively was difficult. A face-to-face meeting of researchers from all three teams in 2002 facilitated the development of nine initial codes, but it proved impossible to involve a Thai team member in the full coding exercise, which was done over several months by one researcher from New Zealand and one from America. Then the coding was reviewed by all team members, using electronic communication. The team planned for one researcher to write a memo about each code, incorporating data from all three countries, and e-mail it to all researchers for feedback. This also proved impossible, so researchers met again for two weeks in April 2005, intending to discuss each of the memos that had been written until everyone understood it and new insights had been generated. For each code, this required each country's researchers to present their own interpretation of their own data, and then a full discussion of the data under that code for all three countries. This process continued after the meeting, using videoconferencing. The process of identifying commonalities began at a meeting in October 2005 and took about 18 months, again with use of videoconferencing and another meeting towards the end of the process in early 2007. The research, including data analysis, was an effective collaboration (Shordike et al 2010: 351) but, as this summary shows, the data coding and analysis took around five years and a great deal of effort to complete.

A **presentation** by Professor Ray Cooksey about the analysis of quantitative and qualitative data with the support of technology can be viewed online.

Data integration

The trend of combining data and findings from different datasets has increased rapidly since the turn of the century (Ivankova and Kawamura 2010: 583; Hannes and Macaitis 2012: 405). The data and findings may be from a single mixed-methods study or from different studies, and may be qualitative, quantitative or both. One of the most challenging aspects of mixed-methods data analysis is integrating the findings from different datasets and/or different analytic techniques. As so often with new research methods, a variety of terms are used to describe the process of combining data and findings, including:

- data integration – usually within a mixed-methods study
- meta-analysis – usually for quantitative studies
- data synthesis – usually for qualitative studies
- meta-synthesis – for qualitative or quantitative studies
- evidence synthesis – for qualitative or quantitative studies
- systematic review – for qualitative or quantitative studies.

Data integration in mixed-methods research can be conducted for a number of reasons, such as to address a research question from a variety of perspectives or to bring together different parts of a phenomenon or process (Mason 2002: 33). Within a research project, data integration has three main purposes: triangulation of data, the development of richer analysis and the illustration of findings (Fielding 2012: 124). The aim is to synthesise equivalent or complementary findings and make further investigation of contradictory findings (Fielding 2012: 125). The precise methods of integration will vary, depending on the nature of the datasets, but there are some basic questions that are likely to apply in any case, such as:

1. How far can each of your datasets contribute to answering your research questions?
2. To what extent can your findings be brought together to create an explanatory narrative?
3. How much do the answers to 1 and 2 above benefit your research?

These kinds of questions can be difficult to answer, but there is no point integrating data just for its own sake, so it is important to ensure that integration is rigorous and meaningful (Mason 2002: 36).

Methods of combining data and findings from different studies usually start from the researcher's strategy for searching for, and identifying, studies that fit their criteria. Although this happens at an early stage of the research, it is included here because it is essentially an analytic process. Devising a search and identification strategy is quite a creative process, with many decisions to be made, such as:

- which criteria to use for inclusion or exclusion
- which search terms to use
- which databases and/or websites to search
- how to manage studies that don't fit with your inclusion/exclusion criteria.

The idea behind having predefined inclusion and exclusion criteria is to reduce researcher bias in the selection of studies (Petticrew and Roberts 2005: 10). However, as the selection criteria are defined by researchers they may themselves include bias (Kara 2012: 21). A short **video** about researcher bias can be viewed online. An initial question is how broad or narrow to make the criteria: broader criteria increase the likelihood of generalisability, while narrower criteria are likely to yield a more homogeneous evidence base (Salanti 2012: 81). Practical considerations also come into play here, as some selection criteria can yield hundreds of thousands of studies, which may encourage researchers to use narrower criteria.

You may choose to replicate an existing search and identification strategy, either because it has proved to be effective and would fit with your own research, or because you want to compare your research with a previous study and using the same strategy would facilitate this (Hannes and Macaitis 2012: 403). Or you may prefer to devise your own strategy. When you have selected your studies, you need to extract the relevant information and then re-analyse it as a dataset of its own, using suitable analytic tools.

Australian researchers Pat Bazeley and Lynn Kemp considered the metaphors used to describe integration in the research methods literature. These included:

- bricolage, mosaics and jigsaws – aiming for completion
- sprinkling and mixing/stirring – aiming for enhancement
- triangulation and archipelago – aiming to show that the whole is greater than the sum of its parts

- blending, morphing and fusion – aiming to explore through transformation
- conversation and DNA – aiming to explore through iterative exchange.

Bazeley and Kemp used their exploration of these metaphors within the research methods literature to define the following eight principles for data integration (Bazeley and Kemp 2012: 69).

1. There are many methods and techniques of integrating data.
2. Integration can start at any point in the research process.
3. Integration should happen at the earliest possible stage, certainly during analysis, and always before conclusions are made.
4. The level of integration must be commensurate with the aims of the research.
5. The research report should provide clear evidence of the ways in which each data element or finding depends on, or is enhanced by, others.
6. Data integration aims to obtain results that could not be obtained in any other way.
7. The research is written up around the topics it investigates and its findings, not around its methods.
8. If the requirements of other forms of publication, such as journal article word limits, make it difficult to include all the components of an integrated study, care needs to be taken to make sure the whole can be represented as well as the relevant parts.

Norwegian researcher Sofia Hussain conducted research in Cambodia that aimed to help the developers of prosthetic legs to obtain a deeper understanding of their users' needs. Hussain conducted semi-structured interviews with six Cambodian children who had prosthetic legs, incorporating 'child-friendly techniques such as drawing, photography, and role play' (Hussain 2011: 1430), and several follow-up interviews with three of those children. She also conducted group interviews with six professional rehabilitation workers and seven children who had no disabilities. Five adults who had been using prosthetic legs since childhood took part first in individual interviews and then in a group interview. Hussain also interviewed five Buddhist monks and four shamans, to find out more about cultural views of disability and health. An interpreter assisted the researcher, as most of the interviewees were Khmer speakers, and all interviews were recorded. The materials for analysis included audio tapes, field notes of interviews and observations, transcripts in English and Khmer, photographs and drawings. Hussain took a cyclical approach to her analysis, 'working from parts to the whole and back again', and 'looking for information that stands out, and might lead to a deeper understanding in relation to the lived experiences of those being interviewed' (Hussain 2011: 1431). This enabled the identification of themes, which were refined through discussion with a colleague. Hussain used reflective writing as a key part of her thematic analysis, viewing her material in the context of relevant literature and reflecting on the changes in her own understanding in the process.

This approach was drawn from the work of Gadamer (1975), who 'wrote that all parts of a text should be understood in the context of the entire text, and that the entire text should be interpreted within the framework of its parts' (Hussain 2011:1431). This analytic approach enabled Hussain to identify relevant cultural and social attitudes, and so to establish ways in which the developers of prosthetic legs, and support services such as NGOs operating locally whose staff fit prostheses, could improve the lives of children with disabilities in Cambodia.

A **Huffington Post articl**e showing one way in which Sofia Hussain's research has been used is available online.

Data analysis using technology

Some researchers worry that a computer will take some of the control away from the researcher within the analytic process (Bazeley and Jackson 2013: 9, Odena 2013: 358) and make the process mechanical (Bazeley and Jackson 2013: 7). However, it can equally be argued that computer-assisted data analysis gives a researcher more control. For qualitative data, this is achieved by offering increased flexibility in coding and ensuring that researchers can rapidly retrieve every item with a given code. Also, it is still the researcher's job to assign names to codes and codes to data, and to derive meaning from the slices of data served up by the computer in response to the researcher's queries. For qualitative or quantitative data, computer-assisted analysis enables work with far bigger datasets than could be analysed by hand. We saw this earlier with the analysis of 600,000 tweets carried out by Tonkin, Pfeiffer and Tourte (2012); even more impressive is the analysis of 2.6 million tweets transmitted by 700,000 users of Twitter during the London riots that was carried out by just three researchers with the aid of computers (Procter, Vis and Voss 2013: 199). The computer is a tool to help the researcher, and just as it's usually easier to bang in a nail with a hammer than with your hand, it's easier to analyse most datasets with a computer than with a pen and paper.

Electronic data needs to be coded with meticulous care, and this can be quite a laborious process. Some computer programs offer automatic coding options so that, for example, you can assign the same code to all the text under the same heading in different documents in a single action (Bazeley and Jackson 2013: 109). This can speed up the process of coding, although it is useful only up to a point, because it is still the researcher's responsibility to define the automated codes, check which other codes may be needed and implement them, and interpret the coded data (Odena 2013: 358). Concentrating on the detail of the data while coding can help to reduce the impact of researchers' unconscious biases about broader themes in the data (Odena 2013: 365) and so can lead to interesting surprises at the analytic stage.

Once the coding is done, analysis is comparatively straightforward. However, whether analysing qualitative or quantitative data, it is essential that the researcher uses analytic techniques appropriately. You need to choose what to consider, compare or calculate, and those decisions should be based on a credible methodological rationale. It is not acceptable to run queries or calculations simply because the software enables you to do so (Cooper and Glaesser 2011: 45–6); you need to know why you're running those queries or calculations.

There are a large number of proprietary software packages to support different kinds of data analysis. Excel and SPSS (Statistical Package for Social Sciences) are probably the most common packages used for quantitative analysis, with the **open source R software** gaining popularity. For qualitative analysis, Atlas.ti and MAXQDA are popular, and there are a number of open source packages available (list on Wikipedia at the time of writing). You can also find packages for the analysis of specific types of data, such as observational or visual data. There are an increasing number of packages being adapted or designed to help with the analysis of large datasets (Cooper and Glaesser 2011: 31; Crowston, Allen and Heckman 2012: 523; Angus, Rintel and Wiles 2013: 261). And there are numerous providers of custom-built software if you can't find the functions you need in any off-the-shelf packages.

Few packages support the analysis of both qualitative and quantitative data, and at the time of writing perhaps the best-known software that does is NVivo. This enables you to import and code across a wide range of data sources, including documents you have created, documents created by others (in word processing or PDF formats), spreadsheets, text files (.txt), images, audio, video, web pages and social media (Facebook, Twitter, LinkedIn, YouTube and so on) (Bazeley and Jackson 2013: 195–209). NVivo is very effective for qualitative data analysis, and reasonably effective for quantitative data analysis (particularly surveys). And it is extremely effective for mixed-methods analysis, supporting a consistent approach to coding across all kinds of data source and so enabling coherent integrated analysis of different types of data (Bazeley and Jackson 2013: 213). Furthermore, NVivo can be useful in the building and development of theory. Grounded theory is often seen as particularly compatible with NVivo (for example, Hutchison, Johnston and Breckon 2010), but the software's modelling and other functions can also be helpful in working with other types of theory.

Transformative frameworks and data analysis

Data preparation, coding and analysis are the aspects of research that are perhaps most resistant to participation. At these stages, research work can be quite tedious, repetitive and time consuming, which puts some people off. There are many barriers to participation in research, and even within transformative frameworks, levels of participation vary considerably between one project and another (Chevalier and Buckles 2013: 174). Also, levels of participation can vary within a research project. Participation is often presented as static and binary – something people either do or don't do – but, in reality, participation fluctuates constantly alongside other demands of life (Jochum and Brodie 2013: 380). However, if people have been fully involved as co-researchers such that they feel the project is truly theirs as much as anyone else's, then they may be willing or even keen to go through the analytic process.

If you want to involve participants in data analysis, you'll need to make early decisions about gathering data and using an analytic process that can be accessible for your participants.

As we saw in Chapter Five, Mary Ann Kluge and her colleagues in America and New Zealand (Kluge et al 2010) used video in their case study of Linda, a 65-year-old woman who had minimal experience of sport and didn't like exercise, yet decided to aim for master's level as a senior athlete. They gathered many hours of video footage, from Linda's first-ever training session to her competing in, and winning, a race at the Rocky Mountain Senior Games. The video footage was viewed several times by researchers and participant together, with the aim of recognising material that was visually significant and could be used to reflect the associated narrative. This collaborative process enabled the participant to verify as significant the themes identified by the researchers.

Some participants may need more support with analysing data than others, such as children, or people with learning disabilities (Nind 2011: 375). Whether or not participants have extra support needs, the key to maximising participation in the analytic phase is to make the process as accessible as possible. It may also help, if appropriate for your project, to integrate data analysis with data gathering, at least to some extent.

Critical communicative methodology (CCM) is a particularly ethical type of participatory mixed-methods research that aims to identify and solve social problems through dialogue. A key aspect of communicative analysis is to identify a successful or 'transformative' case and interrogate it thoroughly. For example, a study of the economic crisis, aiming to find effective alternatives to the capitalist system, focused on the Mondragon Corporation, a very successful and ethical group of cooperatives. Mondragon was founded in 1955 and by 2008 had become Spain's third-largest industrial group by employees, with almost 100,000 staff. It works in industry, distribution, finance and knowledge and is successful in overcoming inequalities such as social exclusion, with a range of ways in which staff can participate in management and decision making. Mondragon was able not only to maintain but also to create employment during the economic crisis. The case of Mondragon was interrogated to find out what made it so successful and whether its success could be replicated. Statistical and interview data was gathered, coded and analysed with reference to identified barriers and enablers to social inclusion, within Habermas's 'system' and 'lifeworld' distinction, that is, the social/organisational or the personal/experiential. Identification of barriers helps to identify reasons why those barriers continue to exist, and identification of enablers helps to show what may be transferred to other contexts (Redondo et al 2011: 282).

Arts-based data analysis

The arts offer many creative ways to enrich the analytic process. Here are just three examples: screenplay writing, I-poems and a mixed-methods arts-based approach using poems, photographs and diagrammatic metaphors.

Lisbeth Berbary, an American researcher, used screenplay writing as creative analytic practice in her feminist research. She studied discourses of femininity among young women in an American sorority (college students' social club for women). Berbary gathered data through participant and informal observation, in-depth and informal interviews and artefact collections. Her analysis began with coding against her research questions, but the resulting categorisation seemed unsatisfactory because it didn't reflect the complexity of her data. So she looked for a way to deconstruct her analysis and chose to write a screenplay, albeit one that is ethnographic rather than intended for production. Berbary structured her screenplay carefully to reflect the themes from her data. She used seven different settings from the university campus, taking details from her observational field notes and making changes only where necessary to protect participants' confidentiality. Four main characters were created as composites from research participants, differing from each other in as many ways as possible while remaining true to the data. Scenes were written showing settings, characters, dialogue, action, interaction, non-verbal communication and gesture. Berbary also included 'director's comments', which enabled her to include the connections she made between perspectives, themes and data. This 'writing inquiry' enabled illumination of sorority women's experiences of gender and femininity (Berbary 2011: 186).

Given that written text is so dominant in research outputs, it is perhaps not surprising that creative analytic practice is also dominated by written art forms.

I-poems are a way of identifying how participants represent themselves in interviews, by paying attention to the first-person statements in the interview transcripts. This technique was developed by Carol Gilligan and her colleagues in the 1990s and used more recently by UK researchers Rosalind Edwards and Susie Weller in their longitudinal research investigating change and continuity in young people's senses of self over time. The interview transcripts are carefully read to identify the ways in which interviewees speak about themselves, paying particular attention to any statements using the personal pronoun 'I'. Each instance of 'I' is highlighted, together with any relevant accompanying text that might help a reader to understand the interviewee's sense of self. These highlighted phrases are then copied out of the transcript and placed in a new document, in the same sequence, each instance beginning in a new line, like the lines of a poem. I-poems can be very helpful in identifying participants' senses of self by foregrounding the voice, or voices, that they use to talk about themselves. This is an adaptable technique that can be used with participants of different ages, genders, abilities and backgrounds (Edwards and Weller 2012: 206) although working with I-poems is quite time consuming, so they're best used with a small sample or sub-sample (Edwards and Weller 2012: 215).

It would also be possible to construct we-poems, they-poems and so on, if relevant for your dataset.

Researching complex human experience suggests, to some people, that they could reflect that complexity by mixing analytic methods.

Jennifer Lapum and her colleagues in Canada used arts-based analysis in their investigation of patients' experiences of open-heart surgery. Participants took part in two interviews after their operations, the first while they were in hospital but out of intensive care and the second when they had been home from hospital for 4–6 weeks. Between the two interviews, participants kept a journal of their experiences (Lapum et al 2011: 102). The multi-disciplinary research team used an arts-based method of analysing patients' stories. They began by imagining how patients felt physically and emotionally during their experiences. Patients' stories were presented in chronological order, so this was used as an organising framework. The framework had five phases: pre-operative, post-operative, discharge from hospital, early and later recovery at home. Within this framework, key words, phrases and ideas from the patients' stories were documented and categorised. These key words, phrases and ideas were used to form free-verse poems. The team also developed concepts for photographic images that would highlight the main narrative ideas of each poem. This process yielded several poems and photographic images. Reflective discussions about the poetry and images, drawing heavily on imagination, were used to seek fuller understanding of patients' experiences. In a second phase of analysis, the images were further developed and the poetic text refined so as to 'further illuminate the complexities, ambiguities, defining features, tensions, and sensory details' of participants' stories (Lapum et al 2011: 104). This was done through 'a process of iterative dialogue, systematic inquiry, visualization, concept mapping, and metaphorical interpretation' (Lapum et al 2011: 104). The researchers discussed, wrote, and drew their findings as visual and diagrammatic metaphors. This analytic technique was explicitly used to work towards a public exhibition that would be used to disseminate the research findings and that is discussed more fully in Chapter Nine.

A **video output** from this research can be viewed online.

CONCLUSION

There are many ways to analyse primary and secondary data and to integrate different datasets, and more are being devised all the time. This means that, although parts of the process can be laborious and repetitive, there is still plenty of scope for creativity in data analysis. However, it is important to ensure that your analytic method produces findings that are firmly rooted in your data. It is equally important to ensure that your analysis and its results will be helpful in the next stages of the research process: writing, presentation, dissemination and implementation.