Jepson, M., Salisbury, C., Ridd, M., Metcalfe, C., Garside, L., & Barnes, R. (2017). The 'One in a Million' study: creating a database of UK primary care consultations. *British Journal of General Practice, 67*(658), e345-e351. https://doi.org/10.3399/bjgp17X690521
The ‘One in a Million’ study: creating a database of UK primary care consultations

INTRODUCTION

General practice accounts for over 85% of all NHS doctor–patient encounters, and in England alone 37,000 GPs1 see around 1 million patients every day across 7,800 diverse practice settings.2 Despite an increase in telephone consulting, 90% of patient consultations still happen face-to-face in GP surgeries.3 Understanding of what happens in these encounters is limited because of a relative lack of research in situ. This is often attributed to perceived difficulties in recording routine consultations, including obtaining ethical permissions, recruiting participants and organising data collection, safe transfer, and storage. Most recent consultations research (2003) has been based on patient interviews, surveys, or medical records, which provide a limited and potentially biased account of what actually occurs.4 Consequently, the skill-sets necessary for working with consultations data have had fewer opportunities to be passed on and may be in danger of declining. In contrast, some of the earliest and most influential studies in the history of general practice research were based on the study of directly observed or recorded consultations. One such study by Byrne and Long,5 published in 1976 but still widely referenced today, remains remarkable for several reasons: for demonstrating the acceptability and feasibility of collecting large datasets of recordings of ‘live’ consultations between doctors and patients in the UK; as the first real evidence base capturing the difference between doctor-centred and patient-centred care; and for capturing a snapshot of what is now our primary care heritage. Despite early critiques of their analytic approach,6 Byrne and Long remains a groundbreaking study. Furthermore, UK training for general practice has been heavily influenced by the use of videotaped consultations.7

Recently, there has been a rise of interest and investment in large datasets of primary care health care, ‘Big-data’: that is, indirect records of selected information collected around the consultation. For example, the Clinical Practice Research Datalink8 contains anonymised medical records data for over 11 million patients. By contrast, no UK datasets currently exist of primary care consultations themselves that might enable capacity building in terms of knowledge and skills to improve consultation outcomes. Although other researchers have collected recordings of UK primary care consultations, most of these have relied on audio-only and have not been collected with data sharing in mind. As such, the data have been subject to restricted ethical permissions.9

The acceptability and feasibility of collecting audiotaped GP consultations with related data have been demonstrated in small studies.10,11 Stakeholders believed that...
collecting such data would be useful for training doctors and quality improvement, but expressed concern about data quality, security, confidentiality, and governance.\(^{10}\) Videos provide a more accurate record of consultations including vocal, verbal, and visual behaviour, and a recent review has shown that most people regard video-based research in healthcare settings as acceptable and worthwhile.\(^{12}\)

Therefore, the aim of this project was to create a controlled database of high-quality videotaped GP–patient consultations, with linked practice, GP, and patient data, with consent for reuse for the purpose of future research and teaching.

**METHOD**

The design was a cross-sectional study of routine GP consultations with adult patients.

**Participant selection and recruitment**

A purposive sample of 12 practices from three clinical commissioning groups (CCGs) in the west of England, including urban, suburban, and semi-rural areas of high and low deprivation, was recruited to the study via the NIHR Clinical Research Network. Twenty-three GPs volunteered to have up to 20 routine consultations recorded over two or three half-day sessions between July 2014 and April 2015.

All adults aged ≥18 years seeing study GPs on recording days were handed a study leaflet on arrival, and given the opportunity to discuss the study with a researcher before deciding whether or not to participate and have their consultation recorded (by either video or audio-only). Participating patients

### Table 1. Archive data: measures collected, from whom and at which time point

| Data source/when completed | Patient | GP |
|----------------------------|---------|----|
| Same day as appointment    |         |    |
| Pre-visit survey           |         |    |
| Post-visit survey          |         |    |
| Follow-up survey           |         |    |
| GP checklist               |         |    |
| GP survey                  |         |    |
| Measure                    |         |    |
| Patient demographics       | n = 328 (98.2)\(^{a}\) | n = 325 (97.3) |
| Pre-visit expectations\(^{13}\) | n = 287 (85.9) | n = 23 (100) |
| Control preferences scale\(^{14}\) | n = 301 (90.1) |    |
| Perceived influence over consultation outcome\(^{13}\) | n = 176 (52.7) |    |
| Health-related quality of life measure\(^{15}\) | n = 325 (97.3) |    |
| Perceptions of consultation outcome\(^{13}\) and decision making\(^{15}\) | n = 325 (97.3) |    |
| Perception of decision making\(^{14}\) | n = 23 (100) |    |
| Depth of doctor–patient relationship\(^{17}\) | n = 23 (100) |    |
| Enablement\(^{18}\) | n = 23 (100) |    |
| Recall of and adherence to treatment recommendations\(^{19}\) | n = 23 (100) |    |
| Patient beliefs about medicines\(^{20}\) | n = 23 (100) |    |
| GP clinical certainty of treatment plan\(^{21}\) and confidence in patient following treatment plan\(^{21}\) | n = 23 (100) |    |
| GP demographics | n = 23 (100) |    |
| GP work-related burnout, stress, and job satisfaction\(^{22}\) | n = 23 (100) |    |

\(^{a}\)Percentages calculated from a total number of 334 (with the exception of the GP survey in the final column).
were also asked to give consent for the research team and/or other researchers to use data in the future, subject to ethical approval, and for their data to be used for the development of medical and research training materials. Patients could choose to consent to any or all of these activities.

For ethical reasons surrounding ongoing consent, patients were excluded if they were aged <18 years, lacked capacity to give informed consent, were consulting on behalf of a third party, or did not speak English fluently. The sex and ethnic group of patients who declined participation, and the reasons given, were recorded on screening logs.

Patients and doctors completed questionnaires before and after each consultation. The timing and data collected are listed in Table 1.

Three months after the index consultation, typed entries relating to the index consultation and any related subsequent entries (reconsulting with the same or another GP or nurse, an out-of-hours visit, or emergency department visit) were extracted from consenting patients’ medical records. Entries were defined as being related to the index consultation if the same symptom or problem was recorded. The number of other, unrelated reconsultations in the same time period was also recorded.

Consultation data

GPs were asked to videotape (or audiotape according to participant preference) all consultations with consenting adults. All recordings were transcribed verbatim by a professional transcription service, anonymised for names and place names, and the content coded for problems and issues discussed using a published coding tool based on the International Classification of Primary Care, 2nd edition.

Practice data

Geographical location, urban/rural classification, area-related deprivation score, patient list size, number of GPs at the practice, Quality and Outcomes Framework achievement, and score on doctor–patient communication items from the GP Patient Survey were extracted from practice websites. Consultation rates and number of missed appointments in the month of data collection were extracted from a search applied to practice records systems.

Study data were managed using REDCap electronic data capture tools hosted at the University of Bristol. Individual-level Index of Multiple Deprivation (IMD) scores were derived from patient postcodes. Descriptive data analyses were undertaken using Stata (version 14). All numerical data were checked for missing values and potential invalid entries.

RESULTS

Practice characteristics

Of the 12 participating practices, six were located in areas of high deprivation (IMD 24–69) and six in areas of low deprivation (IMD 0–10). Ten were training practices and list sizes ranged from 6250 to 18 350 patients (mean practice list size 11 999, standard deviation [SD] 4264).

GP characteristics

Thirteen female and 10 male GPs participated, all of whom were of white ethnic group. Seven were aged <40 years and 16 were aged ≥40 years. The participants had been qualified as GPs for a mean of 18 years (range 2–35 years), and had worked at their current practice for a mean of 10 years and 11 months (range from 6 months to 32 years 1 month).

The flow of patients through the study is shown in Figure 1.

Figure 1. Flow diagram of patient recruitment and response rates.
The characteristics of the patient sample are shown in Table 2. Three-hundred and twenty-seven (97.9%) patients agreed for their data to be retained for use by the ‘current research team’, 300 (89.8%) for future use by ‘other researchers’, and 299 (89.5%) for ‘the development of medical and research training materials’. Of the recorded consultations, 167 were with patients attending practices in areas of high deprivation (at a consent rate of 77.5%) and 160 at practices in areas of low deprivation (consent rate 81.4%).

### Consultations data
Camera error caused seven out of 334 recordings to fail. Of the 327 successful recordings, 307 (94%) were video and 20 audio-only. The mean duration of consultations was 12 minutes 2 seconds (95% confidence interval [CI] = 11.28 to 12.36; SD 0.514; range 1 minute 19 seconds to 37 minutes 54 seconds). The patient initiated 294 out of 327 consultations (89.9%), with 17 (5.2%) initiated by the doctor or the practice (in the remaining 16 [4.9%] it was not possible to tell). Most of the interactions (275, 84%) were two-party interactions featuring just the patient and a GP; while 52 (16%) featured a third party, typically the patient’s child, partner, parent, and/or carer. Nine were ‘joint’ consultations; double appointments booked for two people to be seen consecutively.

In the 327 recordings, a total of 518 problems were discussed, with an average of 1.58 per consultation (95% CI = 1.49 to 1.68; SD 0.514; median 1; range 1–5). Most problems were raised by patients (n = 441; 85.3%), with 66 (12.6%) raised by the GP, and 11 (2.1%) raised by a third party. Table 3 shows the types of problems discussed. Following Procter et al.24, the dimensions of each problem, that is, the types of ‘issues’ that were discussed with the GP to address the problem, were also classified using the nine issue types shown in Table 4.

### Survey data
The pre-consultation survey was completed by 287 patients, and 301 filled in the post-consultation survey on the day of data collection, or very soon afterwards. A follow-up survey, which was sent either by post or e-mail 10 days after the recorded consultation, was completed by 176 patients, representing a return rate of 52.7%, with rates particularly low in the more deprived neighbourhoods. All 23 GPs completed the GP surveys and 325 post-consultation checklists were completed. Details of the numbers of completed surveys are included in Table 1.

### Medical records data
Medical record entries were collected for the index consultations of 311 patients. Of these, 227 out of 311 (73.0%) patients

| Table 1. Characteristics of patient sample, N = 334 |
|---------------------------------------------------|
| **Sex**                                           |
| Female                                           | 212 | 63.5 |
| Male                                             | 122 | 36.5 |
| **Ethnic group**                                  |
| White                                            | 291 | 87.1 |
| Other                                            | 43  | 12.9 |
| **Age, years**                                    |
| 18–34                                            | 91  | 27.2 |
| 35–54                                            | 94  | 28.1 |
| 55–74                                            | 99  | 29.6 |
| ≥75                                               | 36  | 10.6 |
| Not reported                                     | 14  | 4.2  |
| **IMD quintile**                                  |
| 1st (least deprived)                             | 106 | 31.7 |
| 2nd                                              | 54  | 16.2 |
| 3rd                                              | 35  | 10.5 |
| 4th                                              | 53  | 15.9 |
| 5th (most deprived)                              | 84  | 25.1 |
| Data unavailable                                 | 2   | 0.6  |

| Table 3. Types of problems discussed (in order of frequency) |
|-------------------------------------------------------------|
| **ICPC-2 Code**                                             |
| **Problem type**                                            |
| **Frequency**                                               |
| **%**                                                        |
| **Number of consultations with this problem type**          |
| L Musculoskeletal                                           | 94  | 18.1 | 89   |
| P Psychological                                             | 63  | 12.2 | 63   |
| D Digestive                                                 | 53  | 10.2 | 52   |
| R Respiratory                                               | 48  | 9.3  | 46   |
| S Skin                                                       | 45  | 8.7  | 42   |
| K Cardiovascular                                             | 39  | 7.5  | 37   |
| A General and Unspecified                                   | 31  | 6.0  | 25   |
| U Urological                                                 | 23  | 4.4  | 23   |
| N Neurological                                               | 22  | 4.2  | 21   |
| T Endocrine/Metabolic and Nutritional                       | 22  | 4.2  | 22   |
| W Pregnancy, Childbearing, Family Planning                  | 21  | 4.1  | 21   |
| X Female Genital                                             | 20  | 3.9  | 20   |
| H Ear                                                        | 11  | 2.1  | 11   |
| Y Male Genital                                               | 11  | 2.1  | 10   |
| B Blood, Blood Forming, and Immune Mechanism                 | 9   | 1.7  | 9    |
| F Eye                                                        | 6   | 1.2  | 6    |
| **Total number of problems**                                | 518 | 100.0 |

ICPC-2 = International Classification of Primary Care, 2nd edition.
had reconsulted at least once. For reconsultations, 128 of these patients had reconsulted with the same GP and 143 with another GP in the practice for the same problem, with some patients seeing both the same doctor and a different doctor.

**DISCUSSION**

**Summary**

An archive of recordings was successfully created of over 300 routine general practice consultations with linked practice and GP data, plus pre- and post-visit data from patients and medical records, with permissions in place for reuse by other bona fide researchers. The archive is stored digitally in the University of Bristol Research Data Repository, where a more detailed description of the contents may be found. As controlled data, reuse in future studies is governed by approval by an NHS ethics committee and the University of Bristol Data Access Committee. The recorded consultations have been transcribed verbatim, anonymised for spoken names and place names, and the problems and issues discussed within each consultation have been coded. An electronic database has been created that allows the data to be searched according to numerous variables at practice, GP, patient, or visit-level, with all data points linked to the index recordings.

**Strengths and limitations**

The data archive created in this study appears to be the only one of its kind in the UK, allowing access for further research. It contains qualitative and quantitative data including descriptive data for each practice setting, high-quality videotapings, transcripts, and longitudinal questionnaire and medical records data. By prospectively collecting data, the precursors and immediate and more distal outcomes of consultations can be examined. Data were collected equally across areas of low deprivation and areas of high deprivation. Limitations include practices and GPs self-selecting to take part in the study, and the proportion of training practices being substantially higher than across general practice as a whole. By selecting practices within areas of high deprivation and areas of low deprivation, a wide range of patient characteristics have been included but at the expense of generalisability to practices with average deprivation. The recordings are also limited in that they provide a snapshot of 1 to 2 days of consulting for a small number of GPs, clustered within a smaller number of practices. Telephone consultations are increasingly common and these were not captured. All the practices were located across three CCGs in the west of England and consultation behaviour may be different in other areas. For all these reasons, the generalisability of the sample, and the potential to detect differences between subgroups, are limited. The aim of the study has been achieved, however, with standard operating procedures and tools developed to enable future collection of similar data.
from other areas, which can be added to the archive over time.

Comparison with existing literature

This study has taken forward the work of Rushmer et al\(^\text{10}\) and Williams et al,\(^\text{11}\) demonstrating that GPs and patients are willing to have their consultations recorded and linked data collected and archived for controlled future use. The high rates of consent to videotaping are similar to those reported by Salisbury et al.\(^\text{27}\) This study also builds on the legacy of Byrne and Long’s\(^\text{5}\) UK study of GP consulting behaviours to provide the foundations for a database that can enable not one but many future studies of primary care in situ.

Implications for research and practice

The recordings held in the database will support a wide range of future qualitative, quantitative, and mixed-methods research. This could include studies of healthcare communication including comparative studies and those assessing the relationship between consultation processes and outcomes; proof of concept studies such as the development of new communication interventions; and methodological studies such as the development of new coding tools. Furthermore, the recordings may also be of interest to others beyond medical research, such as linguists. Data sharing provides considerable added value in terms of minimising data collection costs, reduced environmental impact, and patient and practice burden. This will support low-cost studies including doctoral-level research, thus building research capacity in primary care. The archive can also be used for development of medical and research methods teaching/training materials with high face validity.

Having developed the necessary study materials, obtained relevant ethical approvals, and demonstrated the feasibility of collecting videotapings of routine GP consultations for the purpose of creating an archive for future research/teaching, the present authors welcome additional consultation data from other researchers to help grow the database. It is hoped that, as well as more consultations, future deposits using the current study protocols will include other aspects of general practice (for example, telephone consultations and home visits) and other practitioners (for example, nurse consultations).

Patient–doctor consultations are at the core of primary care. Greenhalgh et al argued that the drive for evidence-based medicine has overlooked the everyday context of clinical practice.\(^\text{28}\) Each recorded consultation in the present database represents a snapshot of the context of clinical practice for one individual in the million or so consultations that take place in England alone each day. Greater understanding of the content and conduct of these consultations, through research and teaching using the archive, will help improve the process and outcomes in primary care.

---

**Funding**

This independent study was funded by the National Institute for Health Research School for Primary Care Research (Ref: 208) and the South West GP Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. Rebecca Barnes and Chris Salisbury are partly supported by funding from the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West).

**Ethical approval**

Ethical approval for this study was given by the South West — Central Bristol Research Ethics Committee (Ref: 14/SW/0112).

**Provenance**

Freely submitted; externally peer reviewed.

**Competing interests**

The authors have declared no competing interests.

**Open access**

This article is Open Access: CC BY-NC 4.0 licence (http://creativecommons.org/licenses/by-nc/4.0/).

**Acknowledgements**

The One in a Million team would like to acknowledge all the patients and GPs who contributed to the study, Clinical Research Network — West of England, Mai Baquedano for REDCap database support, and the University of Bristol Research Data Service for archiving support. For information on how to access the One in a Million database One in a Million: Primary Care Consultations Archive, please visit http://www.bristol.ac.uk/primaryhealthcare/researchthemes/one-in-a-million/. For all enquiries and queries contact: one-in-a-million@bristol.ac.uk.

**Discuss this article**

Contribute and read comments about this article: bjgp.org/letters
REFERENCES

1. NHS England. General practice forward view. 2016. https://www.england.nhs.uk/wp-content/uploads/2016/04/gpfv.pdf (accessed 22 Mar 2017).

2. NHS Digital. Numbers of patients registered at a GP practice — April 2016. http://content.digital.nhs.uk/catalogue/PUB20480 (accessed 22 Mar 2017).

3. Hobbs FDR, Bankhead C, Mukhtar T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. Lancet 2016; 387(10035): 2323–2330.

4. DiMatteo MR, Robinson JD, Heritage J, et al. Correspondence among patients’ self-reports, chart records, and audio/visual tapes of medical visits. Health Commun 2003; 15(4): 393–413.

5. Byrne PS, Long BEL. Doctors talking to patients: a study of the verbal behaviours of doctors in the consultation. London: HMSO, 1976.

6. Mishler E. The discourse of medicine: dialectics of medical interviews. Norwood, NJ: Ablex, 1984.

7. Pendleton D, Schofield T, Tate P, et al. The new consultation: developing doctor–patient communication. Oxford: Radcliffe Medical Press, 2003.

8. Herreth E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol 2015; 44(3): 827–836.

9. Skelton J, Wearn A, Hobbs F. A concordance-based study of metaphoric expressions used by general practitioners and patients in consultation. Br J Gen Pract 2002; 52(475): 114–118.

10. Rushmer R, Themessl-Huber M, Coyle J, et al. Is the routine recording of primary care consultations possible … and desirable? Lessons for researchers from a consultation with multiple stakeholders. Patient Educ Couns 2011; 82(3): 247–253.

11. Williams B, Dowell J, Humphris G, et al. Developing a longitudinal database: a retrospective analysis of 100 million consultations in England, 2007–14. Lancet 2016; 387(10035): 2323–2330.

12. Parry R, Pino M, Faull C, Feathers L. Acceptability and design of video-based research on healthcare communication: evidence and recommendations. Patient Educ Couns 2011; 82(3): 247–253.

13. Parky R, Pino M, Faull C, Feathers L. Acceptability and design of video-based research on healthcare communication: evidence and recommendations. Patient Educ Couns 2011; 82(3): 247–253.

14. Bowling A, Rowe G, Lambert N, et al. The measurement of patients’ expectations for health care: a review and psychometric testing of a measure of patients’ expectations. Health Technol Assess 2012; 16(30): i–xii, 1–509.

15. Degner LF, Kristjanson LJ, Bowman D, et al. Information needs and decisional preferences in women with breast cancer. JAMA 1997; 277(18): 1485–1492.

16. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L) Qual Life Res 2011; 20(10): 1727–1736.

17. Janz NK, Wren PA, Copeland LA, et al. Patient-physician concordance: preferences, perceptions, and factors influencing the breast cancer surgical decision. J Clin Oncol 2004; 22(15): 3071–3078.

18. Howie JD, Heaney DJ, Maxwell M, Walker J, RJ. A comparison of a patient enablement instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations. Fam Pract 1998; 15(2): 165–171.

19. Svarstad B, Chewning BA, Sleath BL, Claesson C. The brief medication questionnaire: a tool for screening patient adherence and barriers to adherence. Patient Educ Couns 1999; 37(2): 113–124.

20. Horne R, Weinman J. Patients’ beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. J Psychosom Res 1997; 43(3): 555–567.

21. Thorn DH, Wong ST, Guzman D, et al. Physician trust in the patient: development and validation of a new measure. Ann Fam Med 2011; 9(2): 148–154.

22. Kristensen TS, Borritz M, Wilsde E, Christensen KB. The Copenhagen Burnout Inventory: a new tool for the assessment of burnout. Work Stress 2005; 19(3): 192–207.

23. Sutherland VJ, Cooper CL. Job stress, satisfaction, and mental health among general practitioners before and after introduction of new contract. BMJ 1992; 304(6841): 1545–1548.

24. Proctor S, Stewart K, Reeves D, et al. Complex consultations in primary care: a tool for assessing the range of health problems and issues addressed in general practice consultations. BMC Fam Pract 2014; 15: 165.

25. Classification Committee of the World Organization of Family Doctors. ICPC-2: international classification of primary care. Oxford: Oxford University Press, 1998.

26. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap) — a metadata-driven methodology and workflow process for providing translational research informatics support, J Biomed Inform 2009; 42(2): 377–381.

27. Salisbury C, Proctor S, Stewart K, et al. The content of general practice consultations: cross-sectional study based on video recordings. Br J Gen Pract 2013; DOI: https://doi.org/10.3399/bjgp13X674431.

28. Greenhalgh T, Howick J, Massey N, et al. Evidence based medicine: a movement in crisis? BMJ 2014; 348:g3725.