Data Recording in Primary Care Field Studies

Patient Records Enhancement Project

Lesley Axelrod
Human Centred Technology Lab
Interactive Systems Group, School of Informatics
University of Sussex, Falmer, Sussex, BN1 9QH UK
l.axelrod@sussex.ac.uk

Geraldine Fitzpatrick
Human-Computer Interaction Group
Technical University of Vienna
Karlsplatz 13, 1040 Vienna, Austria

Flis Henwood
Social Informatics Research Unit
School of Applied Social Science, University of Brighton
Falmer, Brighton, Sussex BN1 9PH UK

Jackie Cassell
Helen Smith
Amanda Nicholson
Brighton and Sussex Medical School
Mayfield House, Falmer,
Brighton, Sussex BN1 9PH UK

Greta Rait
Department of Primary Care and Population Health,
University College London Medical School, London NW3 2NF UK

Abstract—**This position paper describes the Human-Computer Interaction (HCI) field studies component of the multidisciplinary Patient Records Enhancement Project (PREP). PREP seeks to understand variability of data found in primary care electronic records, in particular the balance between coded data and doctor's 'free text' notes. HCI fieldwork will establish variables that affect recording practices. In field studies we observe and record data recording practices in general practice (GP) surgeries, interview staff, video consultations with real patients and video consultations with standardized patients (played by medical actors). By standardizing patients we can compare the impact of other variables: different doctors, in different surgeries, using different e-health systems. Our early findings suggest that variability is due to a complex web of reasons, driven by personal, contextual and organizational processes. Findings from thematic analysis will result in design implications for studies by epidemiologists and public health researchers, design of NHS training and work processes, and design of electronic health record interfaces.**

Keywords: electronic health record; primary care; field study; human-computer interaction.

I. INTRODUCTION

The electronic health record is at the centre of pervasive health care management. There is a rapid increase in use of pervasive information technologies in healthcare settings, which is set to transform the sector. HCI research has a critical and growing role to play in understanding the challenges and opportunities about how such technologies function in practice, in order to inform design of systems that are effective, not only in budgetary terms but also in provision of care. Electronic health records are well established in National Health Service (NHS) primary care settings in the UK with General Practitioners (GPs) and their surgery staff using a range of approved 'systems of choice' [3] to record patient consultations and other patient related interactions, such as incoming letters, test results, prescriptions, etc. One consultation might consist of several entries relating to different elements. Any one entry might consist of coded information or free-text notes, typed in by GPs and other surgery staff, or a mix of both coded and free text data.

Read codes (named after James Read [4]) are arranged in hierarchical families, developed to reflect medical thinking. The logic of this coding system is rather lost in the smaller branches – for example there are a plethora of possible accident codes that are rarely used. In practice, a level of coding with four digits is often used. There are codes for diagnoses, symptoms, prescriptions and tests. It is quite possible for doctors to select different levels of codes, or different branches of codes, when referring to the same thing.

Circulatory system disease	G
Ischaemic heart disease	G3
Acute myocardial infarction	G30
Anterior myocardial infarct	G301

Figure 1. Examples of hierarchical Read codes.

T476	Hit by boat water skiing
U1287	Bitten or struck by crocodile or alligator, occurrence on farm
T5408	Hit by aircraft, without accident to aircraft, member of ground crew or airline employee injured
T55z1	spacecraft accident NOS, member of ground crew injured

Figure 2. Examples of rarely used accident codes.

Doctor's notes made in free-text are 'messy', and lack 'normal' sentence structure, e.g. often lacking full stops and

---

Funded by The Wellcome Trust (086105/Z/08/Z))

PervasiveHealth 2011, May 23-26, Dublin, Republic of Ireland

DOI 10.4108/icst.pervasivehealth.2011.246073

capital letters. They are full of jargon, typos, abbreviations and acronyms, some of which are extremely idiosyncratic or can have more than one meaning. Text might include any information. Text might be entered to support a code, e.g. to say if a code for ‘fracture of tibia’ relates to left or right side of the body. A text description of a diagnosis or test might be used instead of a code. Text might relate to family history, tests ordered, advice given, or to the doctor’s plan of action. In text entries, a notable number of terms are negated during the doctors process of differential diagnosis or for medico-legal protection. Sometimes entire letters are scanned using optical character recognition (OCR) and appear in the entry as text.

> Sun25/5 12.30 at home – colapsed in toilet long history of IHD 1aMI bIHD for crem – pt inferior ref for angiogram
>
> Feeling much better ESR27 cut pred to 5mg od re-start FeS04 200mg bd && see 3 wks w FBC && ESR
>
> occasional – but much better. nil exacerbating note negative ETT in May – no evidence of ihd. thinks gastric origin

Figure 3. Examples of free text

Anonymous versions of records are submitted to large research databases such as the General Practice Research Database (GPRD) [2]. Databases with large numbers of records are a powerful tool and are revolutionizing the work of epidemiologists and public health researchers, who use them, for example, to calculate the incidence and prevalence of diseases and the subsequent need for provision of services or to model how a patient’s journey is impacted by a particular treatment or intervention.

Coded data provides valuable information and is widely used in epidemiological and public health research. We know from studies looking at large numbers of records that the codes selected for use in primary care settings are very varied, and can seem to be used inconsistently. Sometimes different codes relating to a condition are used, or sometimes only very general codes such as ‘had a chat with patient’, rather than a code that relates to a particular disease. This means that a secondary user wishing to extract records relating to a particular condition will have to draw up a complicated list of possible codes that might relate to the condition of interest, e.g. they might need to search for 60 different codes to identify individuals with one particular disease. Valid and reliable processes to develop code lists is one intended outcome of the PREP project.

Free text is rarely used by secondary users of the data, such as public health researchers or epidemiologists. Secondary users usually rely only on coded data because of the cost of ensuring anonymity of free-text and difficulties in analysis of this messy data. When free text is obtained, it must be manually checked by a trusted third party in case it contains any words that might enable identification of an individual, such as names of people or places. Anonymised portions are then replaced by cedillas for presentation to the secondary users. There is a high financial cost to this time consuming manual labour, and consequently free text is either overlooked or purposely not included and is effectively ‘hidden’ from the view of secondary users of the data. It is not yet clear if or how much difference including this data would make to results of studies. Understanding this is another aim of PREP.

II. PATIENT RECORD ENHANCEMENT PROJECT (PREP)

The Patient Record Enhancement Project is a large multi disciplinary project concerned with understanding use of codes in primary care settings in order to develop the best methods for secondary analysis. There are several strands of work:

A. *Statistical Analysis of Database Entries*

The project is held together by statistical, epidemiological work focused on two examplar diseases (rheumatoid arthritis and ovarian cancer). In order to reliably extract relevant data from databases, complex code lists are needed, and meticulous, rigorous methods are needed to draw these up. Once data is extracted based on code lists, we need to establish if we still have missing data, because it is only present in free text format. For example, preliminary results suggest that there are significant differences in estimates of the diagnosis date for ovarian cancer if free text data is included [6].

B. *Natural Language Processing*

The Natural Language Processing (NLP) and Computational Linguistics work stream seeks to establish ways to make access to free text easier. It may be possible to increase easily available and searchable coded data and to reduce costs of anonymisation by recognizing chunks of text and replacing them with pseudo-codes. Recognition of language used in free text is a real challenge because of the lack of ‘normal’ structure and messy nature of the text. First it is required to find the many variants used to describe a condition. A further step is necessary to recognize information that is uncertain or negated, e.g. ‘not rheumatoid arthritis’ or ‘nil RA?’. The certainty of the tone of an entry is another consideration. Processes for this complex work are under development within PREP.

C. *Visualisation*

Another problem is the sheer scale of the data. A record for one patient might have hundreds of consultations, each consultation may have multiple entries associated with it and each entry can be large. Even when working with one individual patient record it can be difficult to see all the details of a lengthy history and might require a number of different screen views. For secondary users, trying to look at large numbers of patients (often tens of thousands) and wanting to ask different questions of the data, the problem is even greater. Visualisation tools are needed to assist in this task and developing these is another aim of PREP. A prototype time line view that links to a database of records is under development.

D. *Human-Computer Interaction Studies*

In the human-computer interaction work-stream we are carrying out field studies in GP surgeries in order to gain insights into data entry practices. We want to understand who enters the data, in what context, how, where and why. We are interviewing a range of staff, observing the movement of data and practice of data entry around the surgery and observing consultations with real patients, followed by walkthroughs of that data entry with staff concerned.

We are presenting 24 doctors, working in six surgeries and using four different computer systems with two 'standardised' patients (medical actors) who present to each doctor for 'dummy' consultations (48 consultations in all). By comparing records made, and from retrospective walk-throughs of data entry with the doctors, we seek understanding of variables that affect their practice of data entry, particularly those that relate to the balance of codes and free text.

III. FIELD STUDY PROCEDURES

The field studies consist of several strands:

A. *Pilot work*

We carried out informal interviews and consultations with a range of stakeholders with NHS primary care experience, drawn from the academic or private sectors, including doctors and nurses, to help draw up interview topic lists and questions. We familiarized ourselves with a typical GP system and we spent time analyzing samples of records from the GPRD, relating to different diseases, in order to gain understanding of the task of secondary users such as epidemiologists and in order to better understand the variability found in such record sets.

B. *Study design, ethics and governance*

We had to design our study in considerable detail in order to submit a proposal for NHS ethics approval, which was duly granted. This included samples of recruitment and explanatory materials, information sheets and consent forms for different stakeholders, a detailed protocol, interview schedules, plans for risk analysis and data control procedures, as well as details of funders, sponsors, indemnity etc. The university ethics committee also approved the study and our local NHS Research Governance body advised and managed aspects such as issuing NHS research passports (once Criminal Records Bureau and Occupational Health clearance was obtained). We designed a risk analysis form to be completed on surgery visits to ensure we thought through any possible issues.

The study as a whole and our information sheets in particular were carefully designed so that they were detailed and honest, but did not emphasise our interest in the balance of coded and free text, as that might have compromised our study by influencing the behaviour of the participating staff. Rather, we emphasized our interest in working practices, the consultation process and the information management processes at work, which are all likely to affect the shape of the record.

We carefully developed the roles for our standardized patients in consultation with medical colleagues, making them as contrastive as possible. The scenarios were piloted with a doctor and refined.

Our first actor played the part of an older middle aged woman presenting with rheumatoid arthritis symptoms. This is a serious chronic disease that has 'red flag' symptoms that should be responded to as a matter of some urgency. The diagnosis requires tests or referral for confirmation. Drugs given early can significantly reduce the impact and progress of the disease. We developed a back history for this patient, and created a dummy electronic health record on each GP system so that she would appear convincing to the doctors.

Our second actor played the role of a younger middle aged man presenting with a sore throat. This is a less serious condition. We gave him minimal back history so that he presented rather as a 'new' patient might.

C. *Participants and consents*

GP surgeries initially received brief 'flyers' about the project either from venues where we presented our plans, via team members, or via our local Primary Care Research Network, who facilitated recruitment. The flyers briefly explained about the field studies and invited submissions of expressions of interest. Nine surgeries returned expressions of interest. Interested GP surgeries were then contacted and given more details about the project either by phone or during a preliminary visit. The key contact person varied, sometimes it was a doctor, or sometimes a practice manager or other staff member.

There was no payment for surgeries taking part as such, but the UK Primary Care Research Network provided some financial support as an incentive for GP surgeries to take part in our approved research project. Our project budget included service support costs for staff time directly taken up by involvement in our project so that surgeries incur no loss of time for their NHS patients or costs to their NHS service from their involvement with us .

Patients willing to allow their consultations to be recorded were selected as suitable and approached by their usual NHS surgery staff, who gave them preliminary information. Those giving an expression of interest were given more detailed information sheets a week before their appointments and if still interested in taking part, were asked to attend ten to fifteen minutes early for their appointment in order for us to check their understanding of the project information, answer any questions and gain properly informed signed consent. Exit consent, after the recorded consultation was also sought, in case of anything unforeseen that occurred during the consultation. In cases where it was not possible to give a week's notice of the information beforehand, the patients were followed up via a phone call 24 hours after the appointment to check they were still happy for their data to be used.

D. *Procedure*

In order to properly inform and consent all of the staff that we might encounter on visits to a surgery, we offered to present our proposal at a staff meeting, invited questions and comments and handed out information leaflets. Our key contact in each surgery helped organize a schedule of visits during which individuals could be informed and consented, and data collection, including observations, interviews and filming could be carried out.

E. *Data*

Our data collection involves multiple medias and should provide a very rich picture of our topic. It includes:

- Sketches of the layout of the surgery and positions of workstations, supported by digital images where possible;

- Observations (recorded as field notes) on the work practices and movement of data observed, supported by occasional questions for clarification and supported by digital images where possible;
- Digital images, screenshots or video of work practices (where dummy patient details could be used for illustration);
- Video of real patient consultations with nurses, doctors or other health professionals
- Video of simulated consultations of GPs with medical actors playing standardised patients.
- Walkthroughs of data entry for videoed consultations;
- Interview data recorded as audio files and from notes taken, with a representative range of staff from each practice.

F. *Analysis*

Using a thematic analysis approach, we are carrying out a qualitative analysis of this data by entering images, video and text into NVivo qualitative analysis software package [1]. All data is overviewed or 'gisted' to give preliminary insights, and each item is then coded in detail. As families of keywords and themes emerge from the codes they are grouped and divided as appropriate to build an overall structure that represents our findings. Using the software we can check for consistency by using multiple coders, or by over-viewing all examples of a code.

IV. RESULTS

This is a work in progress and we do not yet have definitive results. However a rich and complex picture is emerging. It is clear that there are many factors that affect the variability in the balance between coded data and free text. Different organizational practices have an impact on what is entered in the record. For example, the NHS Quality Outcomes Framework (QOF) requires doctors to show how many patients have particular codes attached to them, so that provides a motivational driver for recording those particular codes. In some local areas there are directives from consultants about what tests should be carried out in primary care settings, prior to referral to secondary care, so that will impact on practice and affect what test codes appear. In the context of the GP surgery, role titles of staff vary between surgeries and the same role might have different names, or a named role might carry different responsibilities and degree of freedom to record data in different settings. Different computer systems have different affordances. At an individual level, training, experience and habit all play a role.

V. DISCUSSION

A. *Multidisciplinary work*

This is a highly multidisciplinary project and there has been an element of emotional labour in incorporating and understanding divergent viewpoints and ways of working within our wider project team. We have all worked hard at this and gained insights from it. For example, epidemiologists have taken the opportunity to attend or review the HCI field studies which has provided them with real insights into reliability of coded data. Epidemiologists have provided the team with insights into very highly structured quantitative analysis. Computational linguists and natural language processing members have provided insights into linguistic classification and analysis methods and into issues around anonymisation and how linkage of data can impact on this. Computer scientists and designers have worked together to develop visualization tools for the whole team's benefit.

B. *Working with medical actors*

Working with medical actors has been a very interesting and useful experience. This is rather like having live 'personas' in HCI terminology. Developing scripts was challenging and required close collaboration with medical colleagues and careful piloting to get it right. The actors do much more than simply play their own role and help the doctors to keep in role too.

C. *The importance of context*

The work of GP surgeries is extremely varied and complex and staff develop their capabilities and work practices over time with the aim of best supporting their patients. The requirements placed on surgeries are ever changing and staff have to respond quickly to changes in requirements. A lot of their knowledge of how to manage their daily tasks is tacit and it would not occur to them to explain these, for example, in an interview. It is only through detailed observation of working staff that it is possible to as the 'right' questions to understand what is going on.

D. *Sociotechnical nature of system*

By looking at data entry in context, we see an emerging picture of a truly socio technical system. If we had used less rich data collection methods this might not have been so apparent. This socio technical system has a number of levels:

- national requirements e.g. the Quality Outcomes Framework (QOF) that drives doctors to code certain items accurately in order to receive payment;
- regional directives from consultants about referral practice;
- local decisions about roles of staff and choices about computer systems and protocols about use of templates;
- individual decisions about what to enter and how to do so.

VI. CONCLUSIONS AND FUTURE WORK

In these PREP field studies we are working closely with a multidisciplinary team and with GP surgery staff to explore the technical, social, medical and cognitive contexts of the electronic record that is at the heart of pervasive health care management and the socio-technical healthcare system. This project is examining how medical record applications are used in primary care settings in the UK, especially with respect to what doctors choose to enter as free text and what they record via structured codes.

The primary purposes of PREP are to inform the design of research of secondary users of health data such as public health researchers, and to inform designers and users of GP systems about issues arising. PREP is investigating several issues that affect secondary users of health records and have important implications for the design of e-health record systems, implications for training of staff and of design of work processes in primary care settings in order to promote acceptance and use of systems. We plan to complete our analysis and publish full results.

The e-health record is becoming increasingly pervasive as it spreads beyond the doctor's consulting room, into the wider surgery, into clusters of surgeries that work together and reaches into the hospital, home, leisure settings, social care settings and workplace. It is important to understand the needs of all the different stakeholders and user groups and to empower them to contribute to the overall system design.

A number of issues have already arisen that suggest future directions for our work. The shape of the record is changing and will continue to change, as health care becomes ever more pervasive in different settings and links to different initiatives. We need to consider what this means for the different stakeholders and design the whole socio-technical system in response to that. Secondary users play a very important role in the system and in health care management but are often overlooked as stakeholders when systems are designed.

Factors that will affect the changed shape of the record include:

- Government initiatives are pushing the health system towards a self care model, in response to global changes such as economic capability and ageing of the population resulting in pervasive systems reaching into homes;
- Healthcare is becoming more feminised as increasing numbers of GPs are women;
- Assistive living initiatives, smart homes and the uptake of health and fitness technologies (also government driven) mean that different types of data and continuous data streams will find their way into the record;
- Increased integration of records e.g. from primary (GP) and secondary (hospital) settings or from health and social care or from NHS and private care settings are likely;
- Increased patient access to records might alter doctors' decisions about what to put in the record and patients might add their own data;
- Increased linkage to information that is presented in such a way as to support health and fitness behaviours;
- Changes in the coding system used will occur – both due to constraints from system developers, in response to incentives and as we move to a more globally compatible interoperable system of coding [5].

ACKNOWLEDGMENTS

We thank all our participating GP surgeries from Surrey and Sussex. We acknowledge the contributions of our colleagues and collaborators from University of Sussex, Brighton and Sussex Medical School, Brighton University, University College, London and The General Practice Research Database (GPRD).

REFERENCES

[1] Computer Assisted Qualitative Data Anaysis http://caqdas.soc.surrey.ac.uk/Support/Choosingsoftware/softwareoptions.html accessed 1st March 2011

[2] General Practice Research Database http://www.gprd.com/home/ accessed 1st March 2011

[3] GP Systems of Choice http://www.connectingforhealth.nhs.uk/systemsandservices/gpsupport/gpsoc accessed 1st March 2011

[4] Read codes http://www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes accessed 1st March 2011

[5] Snomed http://www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/snomed accessed 1st March 2011

[6] A. R. Tate, G. R. Alexander Martin, T. Murray-Thomas, S. R. Anderson, and J. A. Cassell. Determining the date of diagnosis - is it a simple matter? The impact of different approaches to dating diagnosis on estimates of delayed care for ovarian cancer in UK primary care. BMC Medical Research Methodology, 9, June 2009.