RESEARCH

Exploring practical approaches to maximising data quality in electronic healthcare records in the primary care setting and associated benefits. Report of panel-led discussion held at SAPC in July 2014

Sheena Dungey[1,2], **Simon Glew**[3], **Barbara Heyes**[4], **John Macleod**[5] and **A. Rosemary Tate**[2]

[1]CPRD, Centre of the MHRA, London, UK
[2]Department of Informatics, University of Sussex, Brighton, UK
[3]Brighton & Sussex Medical School, Mayfield House, Brighton, UK
[4]PRIMIS, University of Nottingham, Nottingham, UK
[5]MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, UK

Background: Electronic healthcare records provide information about patient care over time which not only affords the opportunity to improve patient care directly through effective monitoring and identification of care requirements but also offers a unique platform for both clinical and service-model research essential to the longer-term development of the health service. The quality of the recorded data can, however, be variable and can compromise the validity of data use both for primary and secondary purposes. **Objectives:** In order to explore the challenges and benefits of and approaches to recording high quality primary care electronic records, a Clinical Practice Research Datalink (CPRD) sponsored workshop was held at the Society of Academic Primary Care (SAPC) conference in 2014 with the aim of engaging GPs and other data users. **Methods:** The workshop was held as a structured discussion, led by an expert panel and focused around three questions: (1) What are the data quality priorities for clinicians and researchers? How do these priorities differ or overlap? (2) What challenges might GPs face in provision of good data quality both for treating their patients and for research? Do these aims conflict? (3) What tools (such as data metrics and visualisations or software components) could assist the GP in improving data quality and patient management and could this tie in with analytical processes occurring at the research stage? **Results:** The discussion highlighted both overlap and differences in the perceived data quality priorities and challenges for different user groups. Five key areas of focus were agreed upon and recommendations determined for moving forward in improving quality. **Conclusions:** The importance of good high quality electronic healthcare records has been set forth along with the need for a practical user-considered and collaborative approach to its improvement.

Key words: data quality; electronic healthcare records; primary care

Received 21 May 2015; revised 7 November 2015; accepted 20 November 2015; first published online 18 January 2016

Correspondence to: Sheena Dungey, CPRD, Centre of the MHRA, 151 Buckingham Palace Road, London SW1W 9SZ, UK. Email: sjbdungey@gmail.com

Introduction

Exploiting the information contained within electronic healthcare records (EHR) data will be

key to addressing major challenges to public health both nationally and globally, ultimately offering a means of maximising efficiency and equality in care. There are, however, significant challenges in using EHRs effectively and particularly in ensuring the quality of data recorded. Incorrect or missing data could render records as useless or indeed misleading such that conclusions drawn from the data could have a negative impact. Among other difficulties, recording data can be time consuming to the extent of conflicting with the GP's primary focus of patient consultation in an already time-constrained environment. Understanding the requirements of and the demands upon GPs must be central to addressing the issue of data quality (DQ) within EHRs.

As part of on-going work into DQ at the Clinical Practice Research Datalink (CPRD) and in collaboration with the University of Sussex (UoS), a workshop session was held at the Society for Academic Primary Care (SAPC) conference in 2014 with the aim of exploring issues of DQ in primary care EHRs from the perspective of different users of GP data and with particular focus on how and why data is recorded in the first instance. The intended outcome was a furthered understanding of both the challenges and the direct benefits to GPs of ensuring high quality data with a view to establishing a workable approach to recording data and maximising benefits to all users of EHRs.

Participation

Four panellists, with expertise spanning the use of EHRs, attended the workshop session. These were Simon Glew, a recently qualified GP and NIHR Academic Clinical Fellow in Primary Care Research at Brighton & Sussex Medical School; Barbara Heyes, Business Development Consultant for Primary Care Information Services (PRIMIS), University of Nottingham; John Macleod, GP and Professor in Clinical Epidemiology in Primary Care at the University of Bristol and Rosemary Tate, Medical Statistician at the UoS and lead researcher for the CPRD/UoS DQ project.

In addition, 25 conference delegates attended the workshop, approximately half the room consisting of GPs; the other half being epidemiologists and data and computer scientists.

The discussion

The workshop session took the form of a structured discussion focussed around three questions with participation from the panel and the audience.

What are the DQ priorities for clinicians and researchers? How do these priorities differ or overlap?

The discussion of DQ priorities for the GP was ultimately (and unsurprisingly) focussed around whether data recording will facilitate patient care. The role of the record varies within the scope of patient care and hence so do the DQ priorities. Data may be recorded by a clinician in a consultation where the primary purpose is patient care and where secondary purposes are determined by practice and local and national clinical and reporting requirements. The definition of good DQ within these different contexts can potentially be quite different and in turn differ from the DQ priorities of the researcher whose data requirements will be related to their research question. It was agreed that the need for consistency in recording information about each patient, in terms of the set of data captured for each patient, is much less subjective for the researcher than the GP. The researcher wishes to compare the same pieces of information for many patients whereas the GP must deal with patients on an individual basis with certain pieces of information being more or less relevant leading to different patients having differently structured *data profiles*. However, it was acknowledged that this is counteracted to some extent by incentives for recording such as the UK Quality Outcomes Framework (QOF), recognised as key to shaping DQ of recording with QOF areas being much more uniformly recorded than otherwise.

An additional priority for GPs, associated with patient care, was given as consistency in recording within an individual patient's record profile from 'cradle to grave', that is in view of the fact that a patient history may be a compilation of different doctors' records. Both researchers and GPs agreed that consistently recorded patient history is of great importance in use or re-use of records for any means.

Another factor widely reflected upon was how the process of patient communication to the GP,

followed by the GP making a decision on how to record the information provided, is consequential to the quality of recording. Here it was discussed that the DQ priorities may again differ between clinician and researcher, especially when there is not initially a clear diagnosis. It was agreed that free text may be the best way for GPs to record, particularly during the early stages of consultation, as it allows them to be more expressive. However, currently researchers cannot easily use free text with organisations including CPRD being unable to collect free text amidst recent confidentiality concerns.

What challenges might GPs face in provision of good DQ both for treating their patients and for research? Do these aims conflict?

A residing concern in terms of the challenges faced by GPs in recording high quality data was that of manpower. This was expressed as a need for better recognition of the fact that it takes time to record data of high quality and that subsequently there is a shortfall in funding, support and proportionate staff numbers. This issue was felt to be heightened if data must be recorded for use in research. In addition, the point was made that high quality recording is technically challenging, particularly clinical coding, and that there is a lack of technical skills among primary care staff. It was commented that it is difficult to inspire commitment to training, especially as staff typically have many other priorities, and concluded that improving attitudes to training was a key challenge to be overcome. Along these lines, it was also noted that different staff members will have different levels of training which may lead to variable recording within practices.

Aside from the issue of resources, the point was revisited here of the doctor not always being able to determine categorically what is wrong with a patient, especially early on. The recording context can also lead to quality issues; certain stigmatised or otherwise consequential conditions are commonly minimally recorded or recorded in a manner which may result in bias at the research analysis stage. Here in particular, the aims of the GP were felt to conflict with those of the researcher.

Finally, the point was made that software systems used to record data can present challenges to recording. System utility was reported as variable and at times a hindrance to its central purpose. There was general consensus that software was an area where there was real potential to facilitate high quality recording, such as for coding, but that this was not currently being achieved.

It was discussed that using data entry templates which are pre-coded can facilitate consistent and accurate recording of data. However, they must be designed in consultation with users as clinicians will not use templates unless they support both data recording and the management of patients. Researchers should determine if templates are used for collecting the data they require as this will inform which Read codes are being used to record specific pieces of information. In some cases, erroneous coding of templates can lead to inaccurate data. For example, where a field entry for the concept 'diabetic retinopathy' is used instead of 'diabetic retinopathy screening', this leads to grossly inflated figures for patients with a diagnosis of diabetic retinopathy.

The challenges of software use were also discussed in terms of the employment of different systems throughout the United Kingdom. Inconsistent recording can result as a patient moves from practice to practice due to a lack of system interoperability or within practices due to the prevalence of locum GPs with variable experience of a given system. As well as variability in software, a lack of national standardisation, for example, in test results, was identified as a source of inconsistent recoding.

What tools (such as data metrics and visualisations or software components) could assist the GP in improving DQ and patient management and could this tie in with analytical processes occurring at the research stage?

It was widely agreed that there was scope for information being fed back to GPs – but that this requires careful consideration to identify appropriate mechanisms for provision of truly useful information. For one, it was agreed to be important that real-time/timely information is provided. It was discussed that tools to access practice statistics across diagnoses, prescriptions, referral, etc. are desirable, especially if comparison can be made with national averages, that is, giving GPs access to not just their own practice's data but to the whole database, with examples given as TrialViz (Tate *et al.*, 2014) and the Canadian

CPCSSN Primary care surveillance network[1] thus giving greater insight into what the data should *look like*.

It was agreed that information must be presented in such a way that the impact to patient care and practice management can be understood. This is as pioneered with the PRIMIS data quality indicators toolkit,[2] again enabling benchmarking against other practices and with an emphasis on facilitating understanding of the root cause of a DQ issue and providing systems to correct data, as well as being able to recognise issues at a glance from quality indicator summaries. Identifying which statistics should be provided and methods of summarising complex DQ information into a manageable but meaningful summary set of indicators, was recognised as a key challenge of providing DQ feedback.

It was suggested that provision of DQ information could provide a means of addressing staff training shortfalls, facilitating understanding of the data and its recording and could also help identify areas where training is needed. However, it was acknowledged, referring back to the discussion of question one, that what constitutes good quality is highly subjective and guidance on interpretation of DQ measures may also need to be provided – the crux of this discussion being that if not managed efficiently the introduction of feedback systems may actually lead to an increase staff training requirements or to incorrect information been derived from the data.

It was agreed that better tools for visualisation of patient information, that is, patient profiles, are needed to be able to assess consistency in a patient's history. This tied in with a continuation of the discussion of the shortfall of software for recording. It was agreed that DQ information derived at the research stage should be being fed back into the software development process to provide insight into areas where assistance with data entry could be beneficial, for example, with coding. On a somewhat different but still highly pertinent line, GPs expressed a need to be assured of trust around such use and sharing of data with third parties.

As a final point, the development of the role of the patient in data recording was reflected upon as a key component in the advancement of the use of EHRs in general. In the context of the discussion here, it was posed that making data more accessible to patients and more interactive with the patient being able to access data in the form of tailored medical information and advice, based on the information they share with the GP, would result in an impetus for the patient to ensure data was of high quality. This could be a means of addressing the problem of inaccurate recording early on in the diagnosis process and more broadly could be a way of maximising data benefits.

Way forward

Recommendations for improving DQ in EHRs

From the discussion, key points for action are summarised as:

a. *Providing practices with DQ statistics*
   - Practicalities of providing GPs with data from a national primary care database (anonymised and aggregated) in a timely fashion need to be explored.
   - Research into the most appropriate indicators and visualisation methods should be on-going to reflect the dynamic nature of care recording and these should be matched to training needs of staff.
b. *Monitoring patient DQ over time*
   - Assessment of patient recording consistency at the research stage could be fed back to GPs/software providers.
   - Greater understanding is required of system interoperability and of levels of standardisation in recording across different data entities such as tests and prescriptions. There needs to be a drive towards both of these being improved.
c. *Feeding back to the software development stage*
   - Further investigation is needed into the relationship between software system and recording characteristics. A comparison of recording characteristics for different systems in use is required.
   - A mechanism for providing software developers with DQ information from the research stage needs to be established.

[1] Canadian Primary Care Sentinel Surveillance Network (CPCSSN), available at http://cpcssn.ca

[2] PRIMIS, available at http://www.nottingham.ac.uk/primis/index.aspx

d. *Patient involvement*
   • Both researchers and GPs need to be aware of the changing role of the patient in primary care and beyond and the growing importance of available data in directing patient behaviour.
e. *Assuring data protection*
   • There is pronounced concern around patient confidentiality, as has become especially apparent following the recent care data initiative. Efforts must be taken to promote and make transparent the data protection procedures in place at the research stage and to ensure that these are effective.

Current on-going activity

The scope of DQ in EHRs is distinctly wide-ranging, multi-faceted and dependent on use. Ultimately, the development of an approach for systematically assessing and improving DQ needs to incorporate the expertise of the broad range of EHR users and contributors. In this vein, a network has been established to act as an international forum for discussion, aiming through participation from different user groups, to develop a comprehensive, robust, integrated and widely used approach to measuring and delivering DQ across all aspects of EHRs (https://lists.sussex. ac.uk/mailman/listinfo/dq-network).

Acknowledgements

Reference

Tate, A.R., Beloff, N., Al-Radwan, B., Wickson, J., Puri, S., Williams, T., Van Staa, T. and Bleach, A. 2014: Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface. *Journal of the American Medical Informatics Association* 21, 292–98.