Measuring and Comparing Data Quality in Electronic Health Care Records

A. Rosemary TATE a,1, Robert A. VERHEIJ b, Natasha BELOFF c, Tim Williams c and Dipak Kalra d.

a Dept. of Informatics, University of Sussex, Brighton, UK.
b NIVEL, Netherlands Institute for Health Services Research, Utrecht, The Netherlands.
c MHRA, London, UK.
d Centre for Health Informatics and Multiprofessional Education, University College London, London, UK.

Abstract. The quality of information depends on the quality of data from which it is derived, but data are of high quality only if they are fit for their intended use. Although electronic health records are collected mainly for patient care and audits, their use in research can also greatly benefit the quality of life of patients. The aim of this workshop is to discuss the challenges and issues involved with measuring data quality in electronic health records for epidemiological and clinical research. Although there has been a vast amount of literature on data quality in general, there exists no common framework or methodology for defining and comparing data quality in medical databases used for research. In this workshop we shall share our experiences of assessing data quality in primary care databases and discuss and develop a suggested framework that can help ensure compatibility of data quality measures for different European primary care databases. We shall also discuss how best to use this framework and the results of our investigations to help data contributors improve the quality of data at source.

Keywords. Quality framework, linked health records, health services research.

Introduction

Electronic patient records are increasingly being used for research purposes by academics, governments and the pharmaceutical industry. The quality of information obtained from these records, has potential public health and safety implications so it is crucial that the quality of the data collected is high. However, there is as yet, no established framework or terminology for measuring data quality of electronic patient records for use in research.

1. Aim of the discussion

The main aim of this discussion is to bring primary care data researchers and clinicians together to exchange ideas and to develop a methodology for measuring and comparing
data quality in different European databases. We shall also discuss how we can feedback the results of our investigations on data quality to the people who provide the data (general practitioners, nurses etc.) in order to encourage them to improve their quality for research.

2. Investigating data quality in the Clinical Practice Research Datalink (ART)

The Clinical Practice Research Datalink (CPRD, formerly GPRD) collects patient records from approximately 8% of the UK population and is used internationally for research by academics, governments and the pharmaceutical industry. In partnership with the CPRD the team at the University of Sussex is developing practice-based data quality scores for use in a) a tool which will allow pharmaceutical companies carrying out randomized controlled trials to select patients and practices for these trials (TrialViz [1]) and b) developing data quality markers for comparing different European databases (TRANSFORM [2]). We shall show some recent results of this work.

3. Investigating data quality in the Netherlands Primary Care Research Network (RAV)

The Netherlands Primary Care Research Network holds information of about 1.5 million patients listed in general practices (10% of total population) and is used in more than 500 studies and is hosted by NIVEL. Data quality indicators have been developed a) to be able to select practices for specific research and b) to guide practices in improving the quality of the data they record. Results of NIVEL’s EMR-scan [3] will be shown, including the Webtool for GPs to benchmark their data quality.

4. Comparison of data quality measures in the UK and Dutch database (NB)

Data quality markers, currently developed in TrialViz and TRANSFORM projects, will provide a foundation for a systematic quality assessment framework applicable across patient record databases, assisting in meaningful linkage of such data sources. Our recent findings, recommendations and current challenges for this unified approach to data quality will be discussed.

5. Data quality issues for linked databases (TW)

Data from disparate sources has recently been linked successfully to the background primary care data within CPRD. Linked data sets include Hospital Episode Statistics, Death registration records, and registry data for neoplastic disease and cardiovascular disease, and its utility has been demonstrated [4]. The nature of the context and processes involved in recording of particular data sets is such that no data sets can be seen as a Gold Standard data set either in general or within narrow clinical areas. The concept of what constitutes data quality in relation to linked data sets will be discussed. Analysis relating to data concordance and consistency of results involving linked data
sets will be presented. Discussion will focus on the most appropriate way in which linked data can contribute to overall data quality of a total electronic health care record.

6. Challenges (DK)

Electronic health records are being used for an increasing number of purposes. Each of these has a need and expectation of the quality of the data in the EHR [5]. We may consider data quality as an indication of fitness for purpose of the EHR, which in turn invites recognition that EHRs might be more fit for some purposes than others. The strongest drivers for EHR data quality have come from the research community, most notably epidemiology, for which completeness and consistency are perhaps the most valued qualities. However, for person centered care, faithfulness (to the situation of the patient) might be a more important quality than consistency. This last session of the workshop will reflect on the preceding presentations on approaches to assessing quality, and explore with the audience whether the different purposes of use of EHRs gives rise to competing values on what is good documentation practice, and therefore good quality data. The session will seek to identify if there is a core set of quality expectations that might be true across all purposes and, if so, how these qualities might best be promoted amongst record authors (most notably clinicians). If time permits, the potential use of DQ metrics for benchmarking EHR data to indicate its suitability for different purposes will also be explored.

7. Expected outcome

The main outcome will be a proposed framework for measuring data quality, given its intended use. Secondary outcomes could include suggestions for feedback to General Practices and methods for engaging them (in e.g. the research processes or RCTs) in order to provide incentives for improving the quality of the records for research.

References