The benefits and risks of structuring and coding of patient histories in the electronic clinical record: protocol for a systematic review

Emma Byrne BSc MSc PhD
Research Fellow

Bernard Fernando MSc MBBS MBCS
Clinical Research Fellow
eHealth Research Group, Centre for Population Health Sciences, The University of Edinburgh, Edinburgh, Scotland, UK

Dipak Kalra FRCGP PhD FBCS
Clinical Senior Lecturer in Health Informatics, Centre for Health Informatics and Medical Education, University College London, London, UK

Aziz Sheikh BSc MBBS MSc MD DRCOG DCH FRCGP FRCP
Professor of Primary Care Research and Development, eHealth Research Group, Centre for Population Health Sciences, The University of Edinburgh, Edinburgh, Scotland, UK

ABSTRACT

Background  Data in medical records have in part been recorded in structured and coded forms for some decades. However, the patient history is as yet largely recorded in an uncoded format. There is a need to consider the optimal balance of use of free text and coded data in the patient history. This review protocol summarises our plans to identify, critically appraise and synthesise evidence relating to approaches taken to introduce structure and coding within patient histories in electronic health records, and the empirically demonstrated benefits and risks of structuring and coding of patient histories in health records.

Objectives  To determine how structured and coded data are being introduced for the recording of patient histories, the benefits observed where structuring and coding have been introduced and the risks encountered when structuring and coding are introduced.

Methods  We will search the following databases for evidence of published and unpublished material: CINAHL; EMBASE; Google Scholar; IndMED; Lilacs; MEDLINE; NIHR; Paklit and PsycINFO. We will, depending on the study designs employed, use the Cochrane EPOC, Joanna Briggs Institute (JBI) and Newcastle–Ottawa instruments to critically appraise studies. Data synthesis is likely to be undertaken using a narrative approach, although meta-analysis will also be undertaken if appropriate and if the data allow this.

Results  This protocol should represent a reproducible approach to reviewing the literature regarding structuring and coding in patient histories. We anticipate that we will be able to report results in early 2011.

Conclusion  The review should offer increased clarity and direction on the optimal balance between structuring/coding and free text recording of data relating to the patient history.

Keywords: clinical coding, electronic health records, patient history, structuring of records
Introduction

Systematic reviews

Systematic reviews are focused reviews that attempt to collect and appraise high-quality research on a particular question.1–5 In the medical literature, historically these reviews have tended to focus on identifying, appraising and synthesising evidence in relation to the effectiveness of pharmacological treatments, and in order to answer such questions the focus on randomised controlled trials (RCTs), with quantitative summaries derived using meta-analytic techniques, is appropriate.6 The basic systematic review approach – which in essence seeks to promote the use of a carefully planned and transparently executed approach to interrogating the literature – can, however, and indeed now is being used to address a much broader array of epidemiological, exploratory and health services research questions.

This protocol is concerned with assessing an internationally important question for health services, namely, what are the benefits and risks of structuring and coding of patient histories in the electronic clinical record.

Why this review is necessary

Several arguments have been advanced in favour of increasing the structure and coding of the patient history. For example, it has been argued that structured records are more complete than unstructured records7 and that there are gains to be made in using automated processing of coded data (for research and administration purposes)8 – for example in decision support systems9 or in quality of care monitoring.10 Finally, structured data support the development of standardised interfaces, facilitating the use of medical record data by healthcare professionals and patients alike.11 Large investments, both financial and organisational, are being made on the strength of these arguments.

There are also arguments against any increase in structuring and coding. These arguments stress the fact that applying codes or working within structures is effortful,12 that standards are not uniformly implemented13 and there are aspects of the clinical encounter that structure and code sets are not rich enough to capture.14 It has also been argued that advances in technology may soon make manual structuring and coding obsolete.15–17

These arguments form the context of the debate about structuring and coding. It is essential, therefore, to determine in relation to the patient history which of these arguments are supported by empirical evidence and which are more speculative. It is to this end that we are performing a systematic review of the literature on this subject.

Objectives

We seek to review the:

1. approaches taken to date to introduce structure and coding within patient histories in the electronic clinical record
2. empirically demonstrated benefits of structuring and coding of patient histories in the clinical record
3. empirically demonstrated risks of structuring and coding of patient histories in the clinical record.

Methods

As the purpose of this review is to summarise the evidence on the wider issues and contexts around, as well as the technical and clinical desirability of, structuring and coding, we have chosen to frame the question in an exploratory rather than hypothesis-driven way. The review will provide insight into how structuring and coding have been successfully introduced thus far into the patient history, what benefits have been observed from the introduction of structuring and coding into the patient history and why parts of the patient history are resistant to structuring and coding.

In undertaking this work, the protocol needs to address two key interrelated overarching challenges:

1. Although the search, management and analysis of the literature will be carried out in a systematic fashion, as described in the methods section below, we do not expect the systematic review to result in a simple answer, such as the confirmation/disconfirmation of a hypothesis.
2. Because of the range and the nature of the ‘interventions’ that are under investigation, there are likely to be few, if any, studies that measure the effect of structuring or coding in such a way as to permit a straightforward meta-analysis. Rather, it is likely to prove necessary to synthesise the data using narrative techniques.
Criteria for inclusion in the review

Types of studies
Quantitative and qualitative studies will be eligible for inclusion. Quantitative studies include, but are not necessarily limited to: RCTs; controlled trials; controlled before-and-after studies; interrupted time series; before-and-after designs; analytical studies and cross-sectional studies.

Evaluations of structuring and coding are also likely to have been carried out using qualitative methods. Qualitative study designs include, but again are not limited to, those studies whose data collection methods consist of interviews, case studies, focus group studies, ethnographic observation, participant observation and action research.

The benefits and risks of structuring and coding may also have been examined by applying generalisations from other fields, or by the application of theoretical arguments. Narrative and opinion-based papers, with a clear focus on structuring and/or coding of the patient history, will therefore also be included.

Types of participants
Studies will be included wherever the data object being studied is limited to, or is predominantly, a record of patient history. This will not be limited to any particular setting within (human) health care.

Types of interventions and theories
We are interested in any change to a technical or organisational system designed to alter the way in which information is structured or coded by human or machine. Similarly, we are interested in any linguistic, computational, logical and human behavioural theories that explain the benefits, limitations and risks of coding or structuring of patient histories within clinical records.

Papers will thus be included where:

- an approach to structure or coding patient histories in the clinical record is described OR
- there is discussion of benefits or risks of structuring or coding patient history in electronic clinical records AND
- the structuring and coding described may be performed synchronously or asynchronously.

Types of outcome measures
This review will report on all relevant outcome measures reported by the authors of the included papers. Of particular interest, however, are outcomes related to data quality, patient safety, the quality of healthcare delivery and cost savings.

Exclusion criteria
Papers will be excluded if:

- the setting for the study is not medical or social care
- the structuring or coding activity pertains to any other part of the clinical record
- the clinical record is not electronically stored.

Search methods for the identification of studies
We will search the following international databases for work published over the period 1999 to 2010:

- Cumulative Index to Nursing and Allied Health Literature (CINAHL)
- Excerpta Medica Database (EMBASE)
- Google Scholar – a specialised search index for peer reviewed publications and patents
- IndMED – a bibliographic index of 38 Indian medical journals
- LILACS – an index of Latin American and Caribbean health literature.
- MEDLINE – the National Library of Medicine’s bibliographic index
- NIHR – an index of research portfolio projects
- Paklit – an index of Pakistani medical journals
- PsycINFO – an index of psychology journals.

We will also search the Xplore citation index of the Institute of Electrical and Electronics Engineers (IEEE), the Institute of Engineering and Technology (IET) library through the British Computer Society (BCS) and specialist health informatics publication repositories.

The concepts identified above will be used as the basis for a set of searches (see Appendix A – search strategies). These search strategies will then be applied to the databases above.

Study identification, critical appraisal, data extraction and synthesis

Study identification
Two authors will independently screen all titles identified by the search strategy. Citations will be classified as being either relevant or irrelevant to the review.

Where a decision cannot be made based on the title, the authors will defer their decision until the abstract has been read. Where the authors do not agree on the classification, a third author will arbitrate. After title screening, the abstracts of those studies where no decision could be reached will be retrieved and assessed by the same two authors, again with arbitration where necessary. Where the abstract does not make it clear
whether inclusion or exclusion criteria apply, we will obtain the full document.

**Quality appraisal**

Full text versions of relevant publications will be obtained and screened for quality. Two authors will independently assess the methodological quality for the inclusion of studies. Quantitative work will be quality appraised using the Newcastle–Ottawa Scale (NOS) for analytical and descriptive studies, and using the Cochrane EPOC approach for experimental designs. Joanna Briggs Institute (JBI) appraisal tools will be used for qualitative studies and economic evaluations. Disagreements will be resolved by discussion and, if necessary, with the involvement of a third author.

**Data extraction**

Extraction of data from the full text versions of included reports will be carried out using appropriately adapted JBI data extraction forms. Data from each paper will be independently extracted by two authors. Differences in extraction or interpretation will be resolved by discussion between these authors, involving a third author if necessary.

**Data synthesis**

Where quantitative data allow, we will consider pooling data using fixed effect or random effects modelling, depending on the degree of heterogeneity between studies. We also expect that we will find qualitative studies. We expect, therefore, to summarise the findings in narrative form. We anticipate undertaking both a thematic synthesis, using the guidelines for narrative synthesis laid out by Popay et al., and a more theory-driven approach to interpreting findings.

**Discussion**

Data in medical records have been recorded in structured and coded forms for some decades. However, the patient history is as yet not widely coded. This review, once complete, will present an evidence-informed critique of the literature. It will review the demonstrated benefits and risks of, and the arguments commonly advanced for, structuring and coding of patient histories. This will advance our understanding of why structuring and coding are relatively rarely used in patient histories, whether more structuring and coding is desirable and, if so, how this might be achieved. The systematic review should be ready for reporting by early 2011.

**ACKNOWLEDGEMENTS**

We would like to thank the members of the Independent Project Steering Committee overseeing this project: Simon de Lusignan (Chair), Nick Booth, Steven Kay, Peter Short and Lee Priest.

**REFERENCES**


FUNDING
This report is independent research commissioned by the NHS CFH Evaluation Programme (NHS CFHEP 009). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NHS CFH Evaluation Programme or the Department of Health.

CONFLICTS OF INTEREST
None.

ADDRESS FOR CORRESPONDENCE
Aziz Sheikh
eHealth Research Group
Centre for Population Health Sciences: GP Section
University of Edinburgh
Medical School
Doorway 3
Teviot Place
Edinburgh
EH8 9AG
UK
Email: aziz.sheikh@ed.ac.uk

Accepted November 2010
Appendix A: Search Strategies

The following database searches were trialled on 1st December 2010. The number of items returned is given in brackets.

PubMed/MEDLINE

Searching for terms in all fields:

1) Benefits or Risks: (4,591,987):
   benefit* OR advantage* OR gain* OR assist* OR help* OR improve* OR eas* OR easy OR desire* OR intend* OR risk OR cost OR barrier* OR upheaval* OR obstacle* OR obstruction* OR difficult* OR confus* OR disrupt* OR hazard* OR threat* OR problem* OR *danger* OR disadvantage* OR avoid* OR undesir* OR unwanted

2) Code or Structure (2,138,602)
   code* OR encode* OR read code* OR diagnosis related group* OR international classification of diseases OR medical subject headings OR icd OR snomed OR hrg OR drg OR mesh OR language* OR ontolog* OR systematized nomenclature OR controlled vocab* OR structur* OR metadata OR template* OR form

3) Electronic records: (95,075)
   electronic record* OR health record* OR patient record* OR care record* OR medical record* OR ehr OR scr OR ecr

4) History taking (4,451,621)
   history taking OR clerking OR note capture OR note taking OR patient interview* OR reason for encounter OR clinical documentation OR structured documentation OR clinical noting OR Kardex OR interface terminology OR symptom* OR presenting complaint* OR concern* OR presentation* OR patient histor**

Final Search:

1 AND 2 AND 3 AND 4 (5,511 results)
(benefit* OR advantage* OR gain* OR assist* OR help* OR improve* OR eas* OR easy OR desire* OR intend* OR risk OR cost OR barrier* OR upheaval* OR obstacle* OR obstruction* OR difficult* OR confus* OR disrupt* OR hazard* OR threat* OR problem* OR *danger* OR disadvantage* OR avoid* OR undesir* OR unwanted) AND (code* OR encode* OR read cod* OR diagnosis related group* OR international classification of diseases OR medical subject headings OR icd OR snomed OR hrg OR drg OR mesh OR language* OR ontolog* OR systematized nomenclature OR controlled vocab* OR structur* OR metadata OR template* OR form*) AND (electronic record* OR health record* OR patient record* OR care record* OR medical record* OR ehr OR scr OR ecr) AND (history taking OR clerking OR note capture OR note taking OR patient interview* OR reason for encounter OR clinical documentation OR structured documentation OR clinical noting OR kardex OR interface terminology OR symptom* OR presenting complaint* OR concern* OR presentation OR patient histor**)}

Ovid/EMBASE:

Replication of PubMed Search with ‘mesh’ removed (2,743)
(benefit* OR advantage* OR gain* OR assist* OR help* OR improve* OR eas* OR easy OR desire* OR intend* OR risk OR cost OR barrier* OR upheaval* OR obstacle* OR obstruction* OR difficult* OR confus* OR disrupt* OR hazard* OR threat* OR problem* OR *danger* OR disadvantage* OR avoid* OR undesir* OR unwanted) AND (code* OR encode* OR read cod* OR diagnosis related group* OR international classification of diseases OR medical subject headings OR icd OR snomed OR hrg OR drg OR language* OR ontolog* OR systematized nomenclature OR controlled vocab* OR structur* OR metadata OR template* OR form*) AND (electronic record* OR health record* OR patient record* OR care record* OR medical record* OR ehr OR scr OR ecr) AND (history taking OR clerking OR note capture OR note taking OR patient interview* OR reason for encounter OR clinical documentation OR structured documentation OR clinical noting OR kardex OR interface terminology OR symptom* OR presenting complaint* OR concern* OR presentation OR patient histor**)
The benefits and risks of structuring and coding of patient histories in the electronic clinical record

Google Scholar
Duplicate of Ovid search (87,200 items returned but not of interest based on initial scanning of first 10 pages)

Alternate Search on Google Scholar
"patient history**" AND structure* - returned 4 papers, only 1 paper got through title filter.
"patient histor**" AND cod* - returned 2 papers, only 1 paper got through title filter – it was a duplicate with previous search.

CINAHL
Duplicate of Ovid search (630 results)

IndMED
"patient histor**" AND structure* - 0 returned.
"patient history**" AND cod* - 0 returned.

LILACS
"patient histor**" AND structure* - 0 returned.
"patient history**" AND cod* - 0 returned.

NIHR
"patient histor**" AND structure* - 0 returned.
"patient history**" AND cod* - 0 returned.

PAKLIT
"patient histor**" AND structure* - 0 returned.
"patient history**" AND cod* - 0 returned.

PsycINFO
Duplicate of Ovid search (219 results)