

Patient Records Enhancement Project (PREP)

Evaluating Records for Free Text Content

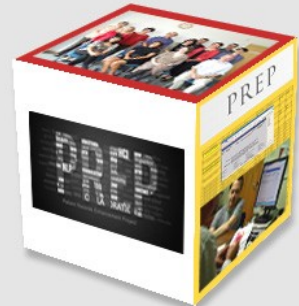
The International Implementation of Electronic Health Records Conference (26 October 2010)

FOR MORE DETAILS PLEASE SEE:

<http://www.informatics.sussex.ac.uk/research/projects/PREP>

Professor Jackie Cassell
Principal Investigator
J.Cassell@bsms.ac.uk

Or Aishath Ali
Research Assistant
A.S.Ali@bsms.ac.uk



INTRODUCTION

The PREP team is interested in electronic health records.

This project aims to make data in records more usable for researchers.

We work on electronic health records with a **particular focus on free text (unstructured data) that is not directly amenable to statistical analysis.**

AIMS

Our aim is to develop strategies for making available, for research and audit purpose, medical information that is "concealed" from researchers in the free text notes, using primary care electronic patient record (EPR) as an example.

Statistical Comparisons

We are exploring the use of natural language processing tools and visualisation tools to improve our ability to answer public health questions using data.

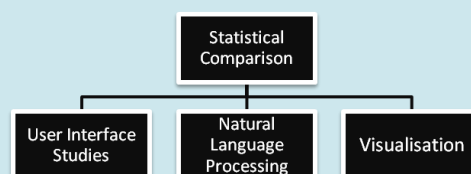
We have chosen ovarian cancer and rheumatoid arthritis as exemplar diseases, and we are making two studies of each to compare outcomes:

- using the *coded* data readily available from GPRD records.
- using the same data but after it is *enhanced* by our natural language processing with *pseudo-codes*.

We are approaching this through four overarching questions

- What determines the balance between free text and coded data in primary care?
- What determines completeness of recording in primary care?
- How does free text coding variation affect data accessibility for users of electronic patient records (particularly health researchers)?
- How can we represent NLP extracted free text to users (particularly health researchers)?

... and four methodological approaches



Natural Language Processing and Computer Science

We are exploring how we can use natural language processing to enrich parts of the data in records so that it is more usable for secondary researchers. We are developing algorithms to search through archived records for recognisable information that is currently not easily available to secondary researchers because it is not in a coded format. Our tools transform those parts of the information and so enhance the record for researchers to use with additional, data derived pseudo-codes.

User Centred Studies

Data Recording in Primary Care

We are carrying out field studies to help understand how and why data in records is assembled as it is. We have begun a series of case studies to see how different primary care staff enter data and use the electronic health record. Is it affected by the software, the type of patient or the condition?

Visualisation

Creating Visualisation Tools for Health Researchers

EPR data are hard for users to get to grips with. We are studying how researchers look through large numbers of records and we are building visualisation tools to help researchers find patterns in the data more easily and clearly. This will help new, non-expert users research the data. We will also be looking at ways to visualise free text extracts obtained from natural language processing.

