International Journal of Population Data Science

Journal Website: www.ijpds.org

Defining 'sensitive' health status: a systematic approach using health code terminologies.

Boyd, Andy^{1*}, Cornish, Rosie¹, Provis, Jennifer¹, Teyhan, Alison¹, and Macleod, John¹

¹ALSPAC; University of Bristol

Objectives

"Clearly, details about an individual's mental health, for example, are generally much more 'sensitive' than whether they have a broken leg." UK Information Commissioners Office

There is a perceived wisdom - based on issues such as social taboos, religious sensitivities, or financial implications linked to health status - that some health data is more sensitive than others. This distinction is present in many of the regulatory interpretations of privacy law (e.g. the UK Information Commissioners Office interpretation of the EU Data Directive, illustrated above), and is factored into the thinking of ethics and other regulatory decision-making committees. However, these particularly 'sensitive' data are defined at a regulatory level in broad terms (e.g. mental health), yet need implementing by researchers in precise terms. In 2013 our longitudinal research study was given approval by the UK Secretary of State for Health to access identifiable patient health records with the exception of those relating to mental health, sexual health or termination of pregnancy. Our objective therefore was to develop a generalisable informatics approach which enabled us to filter out sensitive records at the point of extraction.

Approach

We developed a methodology based on the Cochrane systematic review approach: firstly using internationally recognised definitions of health concepts and reference texts (e.g. British National Formulary drug manual) we identified keywords associated with sensitive health events (including symptom and diagnostic terms, drug and appliance codes, community and secondary care references); secondly, through data-mining code terminologies - using both code terms and information embedded within the structure of the schema itself - we identified code values relating to these terms; thirdly we minimised our results through filtering out spurious results via manual review; finally, the resulting

code lists were then cross-referenced with other terminologies to ensure interoperability.

Results

We produced separate definitions of mental health and sexual health events initially using Read codes. Using NHS cross-reference tables we were able to translate Read observation and diagnostic codes to the SNOMED CT vocabulary, but were unable to translate Read drug codes into the SNOMED/DM+D vocabulary.

Conclusion

We have demonstrated a systematic and partially interoperable approach to defining 'sensitive' health information. However, any such exercise is likely to include decisions which will be open to interpretation and open to change over time. As such, the application of this technique should be embedded within an appropriate governance framework which can accommodate misclassification while minimising potential patient harm.

*Corresponding Author:

Email Address: a.w.boyd@bristol.ac.uk (A. Boyd)