Big Data in Health and Medicine: Issues and Challenges

Prof Louisa JORM, PhD

Director, Centre for Big Data Research in Health
The University of New South Wales, Australia

Professor Louisa Jorm is the Foundation Director of the Centre for Big Data Research in Health at The University of New South Wales, Australia. She has spent equal periods in senior leadership roles in government and academia, giving her unique opportunities for translational research impacts.

Professor Jorm is an international leader in health "big data" research and specifically in applying advanced analytic methods to large-scale routinely collected data and linked data, including hospital inpatient, mortality, perinatal and medical and pharmaceutical claims data. She has made major scientific contributions to research in the areas of health system performance, health surveillance, data linkage and Aboriginal health. Professor Jorm has published over 130 scientific papers and been awarded around $20 million in research grants. She is a high-profile advocate for more and better use of routinely collected health data for research.

Health and medical big data come from a variety of sources, including administrative databases, clinical trials, electronic health records (EHRs), patient registries, multidimensional data from genomic, and other 'omic' measurements and medical imaging. More recently, data are being integrated from social media, wearable and implantable devices, mobile applications, occupational and retail information and environmental monitoring.

These data can be classified as 'big' in **volume** because they include large numbers of individual records (e.g. administrative data) and/or a large number of variables (e.g. 'omics' data). Increasingly, EHR data are 'big' on both dimensions. Health and medical big data are also characterised by great **variety** (including both structured data and unstructured data such as free text and images) and high **velocity** (generated in or near real-time). Because of these characteristics, health and medical big data present the potential for the discovery of relationships among pieces of information in ways that would not previously have been possible. Such knowledge discovery is already gaining pace in areas including large-scale prevention and population health management, risk prediction, precision medicine using genomic information and clinical decision support through machine-learning algorithms.

However, making the best use of big data for health and medical research presents numerous challenges. Data quality is uncertain and difficult to verify. Data standards and metadata lag behind the increasing velocity of data flows and the rapid introduction of new medical technologies. Capturing, storing, managing, processing and analysing very large volumes of data requires new technological solutions including cloud and high-performance computing. New and up-scaled methods for data analytics and visualisation that draw on both statistics and computer science are needed to extract value from big data. The emerging discipline of 'data science' integrates methods from across these domains, but more individuals with both clinical and analytic knowledge are needed to make sense of health and medical big data.

Perhaps the biggest challenges posed by big data in health and medicine relate to the competing imperatives to share data and to protect privacy and maintain confidentiality. Data sharing and reuse ensures reproducibility and maximises the value that is extracted from data. However, data may be reused in ways that were inconceivable at the time that they were collected, questioning notions of privacy that rely on informed consent for the disclosure and use of data. The granularity of big data also erode anonymity, because the relationships between pieces of information can reveal the identity of individuals, even if these have been stripped of personal information.

This presentation will discuss these issues and challenges, employing examples from the presenter's research using large-scale and linked Australian health data.