Using big data to advance mental health research

Anne Duffy, Maria Faurholt-Jepsen, Michael Ostacher

Over the past decade there has been an increasing awareness of the potential for data science to make important advances in brain and mental health research. This focus has coincided with the use of electronic health records in the clinic, the advent of electronic remote data capture through smart devices, publicly available de-identified large data sets, and the emergence of research consortia collaborating on the analysis of complex data sets. Interest in data science in mental health research has been further heightened by the widely acknowledged need for improved diagnostic precision and individualised risk prediction, long-term monitoring, and treatment. Using big data approaches to tackle complex mental health problems is now felt to be a key research priority moving forward. Therefore, Evidence Based Mental Health has dedicated the current Special Issue to address the question ‘How can we use data analysis to tackle mental health issues like depression, dementia diagnosis, and the monitoring of cognitive function?’.

While the shift from paper to electronic health records is on balance a positive direction towards improving clinical management of patients and facilitating big data research, challenges remain. For example, individual institutions use different electronic charting platforms and clinical notes remain in free text form, hampering the ease of systematic data extraction and coding. With the ubiquitous use of smart devices facilitating the collection of high-frequency mental health and behavioural data there are enormous possibilities for ecological sampling or real world research and refined digital phenotyping. However, in order to improve reproducibility and reach their full potential, these approaches require attention to methodological rigour in order to standardise measurement and analytical approaches.

Moreover, access to and sharing of electronic data for research purposes raises a number of new ethical concerns that, while not insurmountable, require careful thought and consensus. For example, it is imperative (and challenging) to protect individual privacy and anonymity, while facilitating ethically approved research; often involving sharing and collaboration between departments, institutions and countries with variable guidelines and regulations.

In Psychiatry, diagnosis remains a clinical construct derived from information gleaned through a comprehensive assessment of symptoms, course of illness, the psychosocial and medical context, and family history. However, symptoms and syndromes are biologically heterogeneous and the same genetic predisposition manifests somewhat differently in relatives and over generations within families. Moreover, mental disorders are genetically complex with variants that overlap to a degree between disorders and which are influenced by gene x environment interactions and gene-environment correlations.

Therefore, novel analytical techniques, such as machine learning, may be more capable than traditional statistical approaches of identifying complex patterns in large heterogeneous data, while at the same time allowing for stratification on a group or individual level. It is hoped that these approaches will leverage large available multilevel data sets that include genomics, imaging, biomarker, remote capture electronic and clinical data to advance diagnostic precision and improve prediction of illness risk and treatment response.

Nevertheless, bigger is not always better—at least in of itself. The ability of big data approaches to make breakthroughs in mental health research is something that has yet to be demonstrated and is at least partially predicated on the quality of the clinical information and characterisation of samples. Careful, prospective observation of well-characterised patients and high-risk subjects over their development and peak period of risk for illness onset together with big data approaches is likely to be a winning combination that will advance both science and clinical practice.

Contributors All authors have contributed to this Editorial and accepted the final version for submission and publication.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

References


