

Big data

Christine Bond

As you read this the seasonal festivities of 2017 will be behind us. Yet since this is the first IJPP issue of 2018 it behoves me once again to wish all our authors, readers and reviewers a prosperous New Year. I hope it will be a productive one for us all as we seek to secure funding to conduct rigorous research that will generate evidence to inform better patient care. Alternatively you may of course be focusing on methodological developments relevant to improving the health services research we undertake, or looking for new approaches to answer your research question. In turn of course here at IJPP we hope that this research will translate into strong papers that you will submit to IJPP for consideration! Looking forward to 2018, a topic of current relevance to researchers, policy makers and even politicians is Big Data. Can we use Big Data to answer previously unanswerable questions or to answer questions more quickly?

So what do we mean by Big Data? Data has been described by the three Vs (1) of volume (the scale and size of data), velocity (the speed at which it is generated), and variety (the varying forms of data such as both images and text). Conversely it has been said that there is in fact no universal agreement on a single definition of big data although the RCPE (2) has posited that traditionally it refers to the large volume of data generated by the electronic devices the majority of us use in daily life but it also consists of the increasing volumes of data generated when we attend for health or social care. In the UK for example, every time you visit a family doctor, receive a prescription, or attend hospital a little bit more digital information is recorded. The potential to link this data to other data about you such as your shopping habits, your travels or your lifestyle suggests infinite possibilities for understanding both the development and management of disease.

Understanding how this data can be used in research requires consideration not only of its potential but also its limitations, and of the governance arrangements needed to ensure its appropriate use. Data security is a big concern, and it needs only a quick scan of the news archives to find reports of datasets being hacked and individual details being accessed by a third party (3). These events are increasing each year as the return for the criminal hacker gets ever greater. Nonetheless because the convenience of the digital world is now a feature of most people's life, both personal and professional, there is no suggestion that our enthusiasm for taking advantage of the digital world is decreasing. What does this mean for health and health services research?

Firstly we must remember that every piece of data we deal with comes from an individual who may also be a patient and we need to treat that data with confidentiality and respect. Even anonymised data can sometimes be linked to an individual because of unusual combinations of individual data items- for example a person in a particular age group with a rare disease from an identifiable geographical area. There is much debate about who actually owns the data and who is responsible for its use. In health there is an argument that the patient themselves owns the data, some believe it is the health care professional eg the general practitioner who collected the data and some think it is the NHS/Secretary of State for Health. It is also unclear if the owner of the data is accountable for its use. So in the previous examples as the data set builds there are multiple owners at different levels. For example whilst the patient might own their own data, they don't own the whole data set whether this be the pharmacist's record of dispensed prescriptions or the GP practice

electronic patient record system. Clearly the pharmacist and GP respectively have a responsibility for ensuring these collated datasets are securely stored and access is limited to those with a justified need. Likewise once the pharmacy and GP and other records may be collated at national levels, ownership or custodianship passes on. A systematic review (4) suggested that health care professionals were in general positive towards data sharing for public health purposes. Barriers identified were concerns about costs, governance and interference with the professional-patient relationship. These were built on in subsequent empirical work which showed conditional support for data linkage for enhanced pharmacovigilance contingent on adequate anonymization of the data, ethical approval for use and adherence to legislation and professional codes of conduct (5). None of this is surprising but operationalizing it still requires thought. It is also important to understand patient views. Whilst many are happy to post intimate details on social media sites and to engage in financial transactions there is a perceived concern about similar actions with health data. Nonetheless research has confirmed that as with health care professionals patients would be willing for their health records to be shared for pharmacovigilance research as long as certain conditions were in place including again reassurance on confidentiality controlled access and appropriate public engagement to ensure full understanding by the population. (6; personal communication).

To some extent some of the above are already in place to varying degrees in different countries. For example in England there is legislation in place for the use of data (Care Act 2014) whereas in Scotland there is a more pragmatic approach (2). The Wellcome Trust lead an initiative 'Understanding Patient Data' – which aims to support better conversations about the use of patient health information, and within this there is a stream of work exploring the implications of new and emerging data-driven technologies for healthcare and research, particularly in relation to ensuring public confidence. The ABPI have published a report summarizing the challenges and opportunities of big data and emphasizing the need for increased awareness and building capacity and capability (7). However much remains for us as researchers to take forward especially thinking of medicines and pharmacy related issues. In practical terms what are the sources of pharmacy data that could be linked to other health data, and what research questions could this linked dataset address? How can pharmacy staff be engaged to ensure their data is complete and valid? Can pharmacy staff be data literate and reassure the public that their data is secure and used in accordance with both the law and current ethical standards? These and other aspects of big data will provide our research community with much food for thought in the coming months and I look forward to hearing your views and seeing your research as we work together to maximise the benefits and minimise the concerns about big data.

References

1. Laney, D., 3D data management : controlling data volume, velocity and variety 2001 <http://blogs.gartner.com/Doug-Laney/files/2012/01/>.
2. Walesby KE, Harrison JK, Ross T.C. What big data could achieve in Scotland J Roy Coll Physicians Edinburgh. 2017 47:114-9 doi: 10.4997/JRCPE 2107.201
3. World's biggest data breaches <http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/>(accessed 15.12.17)
4. Hopf YM1, Bond C, Francis J, Haughney J, Helms PJ Views of healthcare professionals to linkage of routinely collected healthcare data: a systematic literature

review. *J Am Med Inform Assoc.* 2014 Feb;21(e1):e6-10. doi: 10.1136/amiajnl-2012-001575. Epub 2013 May 28.

5. Hopf YM, Francis J, Helms PJ, et al Core requirements for successful data linkage: an example of a triangulation method *BMJ Open* 2016;6:e011879. doi: 10.1136/bmjopen-2016-011879 Christine Bond
6. Bond,C., Scobie Scott, E., Helms, P., Shaw,D., Haughney, J., Is the use of linked routinely acquired NHS data for pharmacovigilance in children acceptable to parents and young people?. *Royal Pharmaceutical Society Conference, IJPP 2013; 21: Issue supplement 2 Abstract 0174 pp 126-7 2013*
7. ABPI Big data road map 2013 <http://www.abpi.org.uk/media/1309/abpi-big-data-road-map.pdf> (accessed 14.12.17)