Harnessing big data for health

Kirsten Patrick MB BCh DA

Almost all technological memory has been digital since the turn of the century, and it's led to an ever-expanding mass of "big data" — countless data points captured each day through mobile devices, cameras, microphones, wireless sensing technologies, radio-frequency identification and software logs. Everybody's talking about them, but the sheer scale of the data that exist and continue to amass is nigh unimaginable. They're exciting but dirty, and integrating salvage from this multiplatform dump of information in any kind of analysis that can answer important questions is tantamount to informatics chicanery. But how useful can big data be for improving health?

You might argue that big data are already well exploited in health care research. You might be surprised that we at *CMAJ* think that we have published only one study that we would label "big data" research.[1] The authors of this study used linked health services data and Google Trends data to try to ascertain whether nocturnal leg cramps are a seasonal complaint — strengthening their observations by showing a trend in both hemispheres.

Yes, epidemiologists have been using data from large registers and cohorts for years to answer clinical questions that can only be addressed by sophisticated data linkage. More recently, we've been mining data from electronic medical records and, even more recently, realizing the potential of aggregated data from hundreds of medical practices or multihospital networks to define patterns of morbidity, create risk scores and monitor adherence to best medical practice, cleverly developing new analytical methods as we go. But how well are we tapping the potential of real "big data" … those that amass organically on multiple platforms? I'm talking about the data that Google, Amazon, health insurers, supermarkets, and just about anyone who reels us in to use loyalty schemes, uses to monitor our lifestyle and habits in ways that health researchers can only dream of. *CMAJ* editors would like to see more proof of concept studies that can help to show if big data can be useful in answering meaningful health questions.

There are potentially many important research questions that could be answered using big data — beyond what we can uncover with routinely collected data. For example, multiplatform data could help us discover the most important "upstream" and environmental determinants of high-burden disease, to guide policy-making beyond health that could achieve health wins.[2] Big data could also help us redefine antibiotic stewardship. Just as big data are used by corporations to personalize and target advertising, so they can be used to identify the best first choice of antibiotic for individuals based on their disease, hospital admission, medication history data, combined with locoregional data on antimicrobial resistance, travel, migration, work, contact and even nutritional histories. Similarly, potential exists for developing sophisticated systems to identify adverse drug events early.

You might ask if, by encouraging health research using big data, we editors have altered our traditionally dim views of studies that arise out of "data mining" and no prespecified research question. We have not. Health researchers must continue to embrace a responsibility to do the very best quality research — to produce useful science of integrity and not to churn out questionable associations to feed the media animal or personal career progression. Careful planning of research — with clear protocols and outcome measures defined at the start — remains key. The special value of the health researcher lies in the fact that human beings, unlike cleverly coded machines, are uniquely good at recognizing and interpreting patterns. Humans know how to use self-doubt to advantage, to check and recheck the accuracy of a finding, consider bias and take care to report findings transparently.

It takes big money to buy, clean and use big data successfully, to acquire the expertise of big data specialists and to invest in the technological and analytical infrastructure required. But it *could* also lead to savings measured in both lives and health care dollars. Multisectoral collaboration is critically important. The health sector should not be left to its own scrabbling to find funds to support big data research for public health benefit while industry is investing heavily in using big data for market research. With the right proof-of-concept studies, funders of health research will be able to see the potential in supporting big data research.

References

1. Garrison SR, Dormuth CR, Morrow RL, et al. Seasonal effects on the occurrence of nocturnal leg cramps: a prospective cohort study. *CMAJ* 2015;187:248-53.
2. Big data analytics in health: white paper full report. Toronto, Montréal, Halifax, Vancouver: Canada Health Infoway; 2013. Available: www.infoway-inforoute.ca/en/component/edocman/1246-big-data-analytics-in-health-white-paper-full-report/view-document (accessed 2016 Apr. 5).