

PERSPECTIVE



The big picture

Many medical images are used once then filed away. This trove of clinical data should be made available to biomedical researchers, says **Alan Moody**.

Whether it is Walmart's one million customer transactions per hour, the 220 billion photos on Facebook, or decoding the human genome's 3 billion base pairs, dealing with vast quantities of data provides an opportunity to extract and exploit previously hidden information. Such is the impact of 'big data' that the World Economic Forum considers it to be a new class of economic asset, comparable to commodities such as gold. The rush to mine this new resource is now underway.

Medical imaging data, which are accumulating in clinics in ever larger amounts, seem to be an ideal target for the 'big data' approach. This information could potentially provide significant insights into health and disease. So far, however, big data — or perhaps more appropriately, the 'big picture' — has not had much impact on this information-rich environment.

WHAT A WASTE

Clinical images tend to be used just once, by a single clinician for a single patient, before being left to gather dust. This represents a squandering of resources. Working cooperatively, whether at the local, regional, national or even international level, researchers and clinicians could use these images to identify trends and correlations, bringing scientific and clinical benefits. But first we need to understand how to harness the current clinical imaging workflow to capture these data.

The move towards using the big picture offers two immediate benefits. First, the image content is self-selecting: patients are being referred for investigation of medical problems, so their images will reflect conditions of clinical relevance and importance. Second, the image data are free of charge, or have already been purchased through current healthcare systems. We could therefore assemble an inexpensive foundation for population-based studies that would otherwise be financially unaffordable.

Consider, for example, the looming medical crisis of dementia. Ideally, the focus will be on identifying at-risk individuals so that prevention and early treatment measures can be deployed. Imaging could hold the key. But in the early stages of dementia the visual clues are likely to be subtle, and identifying them requires investigation of the small subset of the population with the potential to progress. Capturing a sufficient number of appropriate individuals for study would mean casting the imaging net very wide, at immense cost. However, patients with symptoms potentially indicating early dementia — such as vague forgetfulness — may already have been imaged in the course of their clinical visits. What if all the imaging data could be collated? The power of the combined data would allow the small signal contained within the images, denoting early sub-clinical disease, to rise above the background noise.

Research, as opposed to clinical, image networks are already being

built. Examples include the Alzheimer's Disease Neuroimaging Initiative and the Canadian Atherosclerosis Imaging Network. However, these networks are pre-defined by the disease under study; the patients recruited already have overt disease. There remains a need to explore early disease, or pre-disease, at a population level. This can best be achieved by building clinical image networks, allowing recruitment at a far larger scale — a job that will require the repurposing of existing clinical imaging data to create the big picture.

To advance towards this goal, we must change our view of image data. Currently, the development of imaging techniques, image acquisition and analysis, and qualitative interpretation of the image, is done by experts whose aim has been to make their part of the data chain as good as possible — in particular, to produce high-quality images and improve diagnosis. The next, crucial step is to create population-wide image data repositories that are available to researchers.

MAKING IT WORK

To exploit this resource, two things are needed. The first is a new breed of image-data scientist. These specialists will be data prospectors searching for a signal arising from the data, viewing combined data rather than individual studies.

The second is a user-friendly network in which to go prospecting. Helpfully, clinical images are now commonly digitized and stored on a picture archiving and communication system, where they can be examined and categorized using a variety of analytical techniques. The biomedical researchers who will use these

systems must be closely involved in the design of the processes for storing, managing and accessing the data.

It will be important to embed structured reporting within the big picture. Clinical images from cancer patients, for instance, generate information relating to the primary tumour, lymph-node involvement and metastases. Combining data on a large scale will accelerate insight into primary tumour growth characteristics and associated disease spread.

Ethical standards dictate that the use of such clinical data in a research setting requires consent from each patient, or all-encompassing institutional permission for the pooling of anonymized patient data. Addressing these and other similar requirements will be challenging, but if successful, we could have a potent new tool to accelerate not only biomarker discovery but also therapy development. Institutions that accept these challenges will be seen as pioneers, exploring this new natural resource and potentially reaping the rewards from the wealth of data it contains. ■

*Alan Moody is chair of the Department of Medical Imaging at the University of Toronto, Canada.
email: alan.moody@sunnybrook.ca*

WE COULD HAVE A POTENT
NEW TOOL TO ACCELERATE
NOT ONLY BIOMARKER
DISCOVERY BUT ALSO
THERAPY DEVELOPMENT.