BIG DATA: COULD IT EVER CURE ALZHEIMER’S DISEASE?

When, with the benefit of hindsight, people look back at the history of biomedical sciences at the turn of the 21st century, the focus will inevitably turn to dementia, and in particular Alzheimer’s disease. How did brain scientists and clinicians of that generation—our generation—deal with this global health issue? Did we make the right choices, or will we seem woefully inadequate, our judgements quaintly naïve?

One strategy that will surely come under scrutiny is the investment in big data sets. In Alzheimer’s disease there has been a surge in recent years towards ‘Big Data’ initiatives, reflecting a general drive that has galvanized other disciplines, in medicine and beyond (Cukier and Mayer-Schonberger, 2013). From business to government, many have been seduced by the possibilities that Big Data seems to offer for a whole range of problems. However, not everyone is convinced, and the debate on its merits is now in full swing in the media (Harford, 2014; Ozimek, 2014).

The same kind of reflection might also be healthy for the neurosciences. Indeed, recent concerns over the Human Brain Project, which aims to simulate the brain on supercomputers, demonstrate how contentious some Big Data projects have become. Backed by €1 billion funding from the European Commission this initiative represents a massive investment, but the size—and quality—of the potential payoff remain questionable to many (‘Open message to the European Commission concerning the Human Brain Project’, 2014).

For brain diseases, Alzheimer’s disease provides the key example to consider. New initiatives announced this year follow in the footsteps of the Alzheimer’s Disease Neuroimaging Initiative (ADNI), a major international collaboration that is now a decade old. ADNI has an impressive track record of data sharing and publications, as well as of attracting grants and investment from industry, with a total of more than $200 million backing it.

Earlier this year the OECD published its report on informatics and ‘the potential for Big Data to be a “game changer” in global efforts to accelerate innovation in neurodegenerative research’ (OECD, 2014). Then followed the announcement of the Alzheimer’s disease Big Data DREAM Challenge: a major collaborative, open science initiative that aims to identify biomarkers for the condition (‘Big Data Challenge for Alzheimer’s Disease’ 2014). These developments call not only for large-scale information repositories, but also for development of analytical tools to mine such data sets effectively.

This journal plays a role in presenting the direction of travel of the research community. We aspire to reflect in its pages some of the best science that is being performed. But Brain hopes also to provoke and stimulate debate. Although we do not have the gift of foresight, we can as a community consider what the realistic potential of Big Data is. As the man on the street might rightly ask: Could it cure Alzheimer’s disease? Is ‘big’ the answer?

That big collections might generally be useful is not the issue. In the Victorian era, for example, the energy and philanthropy of men like Augustus Pitt Rivers and Henry Wellcome provided not only fascinating insights into cultures across the globe, but also influenced the development of scientific enquiry. The utility of the vast collections they, and others, amassed is not in doubt. But what did those collections actually explain? That is the question that lies at the heart of current concerns: What is the explanatory power of Big Data?

Enthusiasts of the new approach counter that ‘Big Data is about what, not why. We don’t always need to know the cause of a phenomenon; rather, we can let the data speak for itself’ (Cukier and Mayer-Schonberger, 2013). Letting the data speak is, in many ways, to be encouraged but it risks the possibility that we won’t be able to make sense of the vast array of information we obtain. Even if we did, what would it show apart from correlation? Could it ever uncover causal mechanisms? That would seem very unlikely.

But advocates of Big Data initiatives in Alzheimer’s disease might understandably argue that first we need to find the ‘signal’ before we start to understand what it means. Indeed, they might point out that although traditional scientific endeavours have uncovered potential genetic and molecular mechanisms underlying Alzheimer’s disease, those approaches also do not provide adequate explanations for the protein manifestations of the condition, in terms of either behaviour (how different patients present with the illness) or brain network dysfunction (how different brain regions and their connections appear to be affected). Bringing together those very different levels of explanation would seem to be crucial to any proper understanding of the condition.

Perhaps the debate about Big Data is exposing some very important general issues about the way that neuroscience is drifting. The trajectories might be quite different depending upon the parochial interests and expertise of different sections of the community. So one virtue of Big Data might be that it can potentially bring together different levels of data—from genes and molecules through to imaging and cognitive function—and thereby expose the lack of integration across these. In this sense, it provides a real
opportunity to glimpse the ‘big picture’, and the holes in that canvas, rather than keeping the focus tightly on what we know best. Regardless of the pros and cons, the Big Data approach is here. Better we make the most of it and exploit it to its maximum—while understanding its limitations—than simply appealing to the argument that ‘Small is beautiful’ (Schumacher, 1973).

Masud Husain
Oxford, UK

References
Cukier K, Mayer-Schonberger V. Big data: a revolution that will transform how we live, work and think. London: John Murray; 2013.

Harford T. Big data: are we making a big mistake? Financial Times March 28, 2014, http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a 00144feabdc0.html#axzz38fqnnTBA.