

7. Big health data in dementia

7.1 Focus of the use case

This use case considers the opportunity to leverage big data for improved health, particularly focusing on the huge and unfortunately growing area of dementia. The questions here will relate to how existing and new forms of data can improve both policymaking and science with relation both to dementia detection and treatment and to dementia care, in addition to the potential for eventual dementia prevention once the mechanisms of dementia are more fully understood. The stakes here are enormous. According to Alzheimer's Disease International:

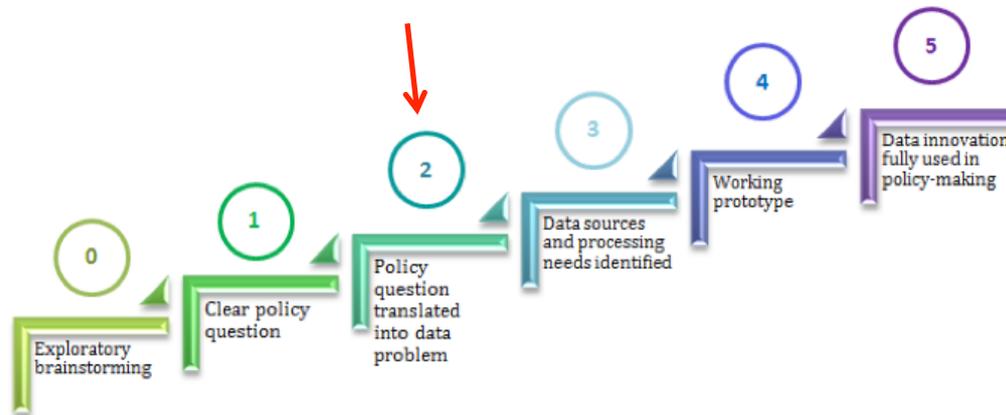
"...over 46 million people live with dementia worldwide...[and] this number is estimated to increase to 131.5 million by 2050. Dementia also has a huge economic impact. Today, the total estimated worldwide cost of dementia is US \$818 billion, and it will become a trillion dollar disease by 2018. This means that if dementia care were a country, it would be the world's 18th largest economy, more than the market values of companies such as Apple (US\$ 742 billion), Google (US\$ 368 billion) and Exxon (US\$ 357 billion)." (Alzheimer's Disease International 2015, p. III).

The policy implications of such a massive health problem are myriad. In December 2013, the then-G8 (now G7) convened the Global Dementia Summit in London to discuss how the world's leaders could address dementia. UK Prime Minister David Cameron identified four major challenges in dementia research:

1. Addressing the market failure undermining dementia research and development;
2. The problem that even when new drugs are developed, it still takes far too long to get them to patients;
3. A lack of collaboration and openness with different scientists all over the world using different data and trying different approaches but not working together enough; and
4. Need for more investment.

This use case focuses on the sorts of data that can improve policy choices both in the realm of science policy and in healthcare policy. One challenge of dementia (discussed more below) is that there are significant challenges in translating scientific research into both short-term and long-term care and treatment needs. One fundamental problem in the area of dementia is the poor understanding of pathophysiology of diseases associated with dementia. As a result, there are currently no reliable early tests for dementia and no clinical treatment. Thus, while there is considerable investment in science, the benefits of those investments are not yet tangible in the form of tests and treatments. From a policy perspective, this use case can be considered to have reached step 2: the policy question of how better to advance dementia research and then translate that into dementia care is translated into a data problem (namely that researchers are working with many sources of promising medical and non-medical data), but the data sources and processing needs and opportunities need extensive further consideration before these basic sources of data can be translated into anything that policymakers or any non-medical/non-specialist stakeholders could use effectively.

Figure 13 Use case readiness level



7.2 The rationale

The health sector accounts for significant amounts of spending, with total healthcare expenditures exceeding 10% of GDP in at least six EU member states⁴⁹, and undertaking a use case of the opportunities for data-driven policy in health would be a similarly enormous undertaking. While dementia is also a huge issue, as discussed above, limiting our focus to this one set of diseases at least somewhat constrains the big data use case under consideration.

Dementia has several dimensions, two of which we will focus on here. The first is the medical dimension: dementia is actually a blanket term for a number of neurodegenerative disorders. Dementia is a syndrome of chronic or progressive nature in which cognitive functioning deteriorates beyond what might be expected from normal ageing. It affects different aspects of everyday life, and comes in a variety of forms, often as a result of other diseases or injuries that directly or indirectly affect the brain. The most common types are Alzheimer's Disease, vascular dementia, dementia with Lewy bodies, and frontotemporal dementia.

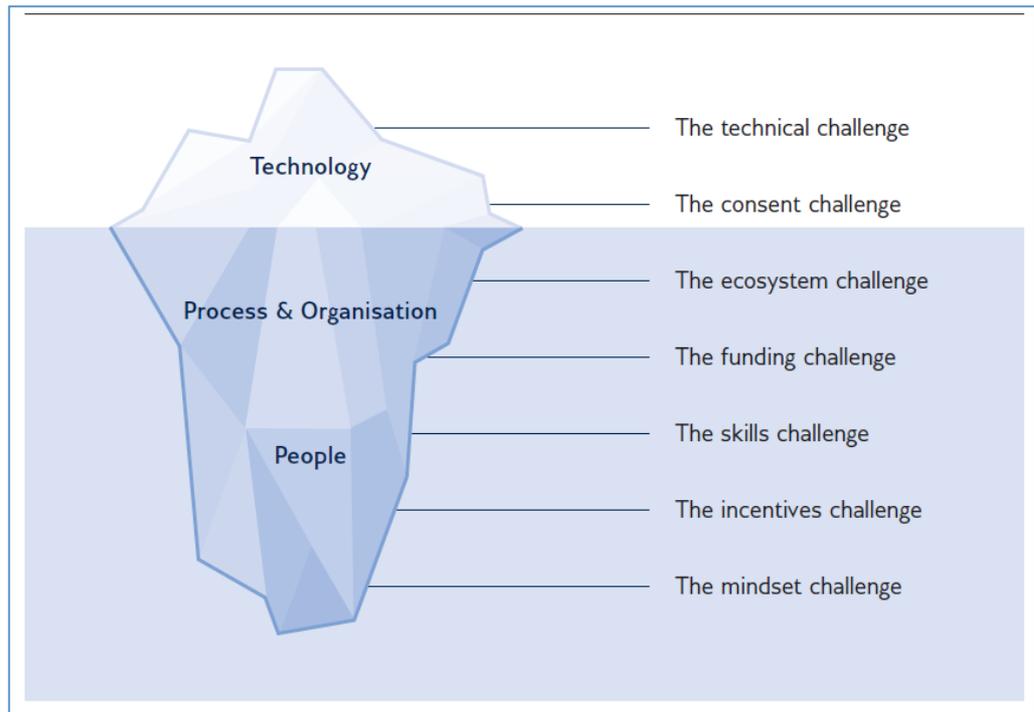
Improving prevention and diagnosis, and discovering potential ways of treatment and cure requires better understanding of the mechanisms underlying neurodegeneration. However, these mechanisms are complex, and influenced by a range of genetic and environmental influences that may have no immediately apparent connection to brain health.

The second dimension has to do with care: dementia detection requires finding more reliable biomarkers that occur before clinical symptoms occur and treatment requires advances in safe and effective treatment protocols. Many of the policy decisions surrounding dementia at the moment have to do with alleviating the burden placed on dementia patients, their caregivers, and their loved ones in the context of the growing economic and social cost associated with dementia.

Deetjen et al. (2015) identified a number of structural challenges to data sharing, shown in the 'iceberg' diagram (Figure 14 on the next page).

⁴⁹ http://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_statistics

Figure 14 Structural challenges to data sharing



Source: Deetjen et al. (2015, p. 50)

As you can see, in this diagram (which is an interpretive and simplified representation of reality) the technical challenges, while significant, are largely visible and tractable. The deeper challenges that lie under the surface are thornier: creating an ecosystem of sharing, developing data skills amongst domain experts, aligning incentives, and influencing the mindset of everyone involved in generating, processing, and analysing the data.

These challenges are highly relevant for using more data (and more relevant data) for addressing dementia. Moreover, the improvements in the selection of indicators and in data collection are dependent on dementia's complex and poorly understood pathophysiological mechanisms.

7.3 The data process: data sharing and data linking

One of the interesting aspects of a data-driven approach to dementia is that while some data sources (discussed below) have been used for many years, others are just newly emerging and are less well developed. Both broad data (relating to the number of individuals represented in a dataset) and deep data (an indication of the number of measures and granularity of those measures related to each individual) are part of the big picture for dementia research. The types of data used will influence the types of data analysis that are feasible.

Medical data

At the core of dementia research is data obtained purposively in medical settings, such as images from Magnetic Resonance Imaging (MRI) or Positron Emission Tomography (PET); clinical, genetic, proteomic and biological data from blood or cerebrospinal fluid; and cognitive tests or surveys. Today, availability of data for dementia research is largely in place at the national level, with a number of large initiatives including two large worldwide studies: ADNI and AddNeuroMed/EMIF. ADNI (Alzheimer's Disease Neuroimaging Initiative, <http://adni.loni.usc.edu/>) is an initiative founded in the US in 2004. Its aims are to help predict and monitor the

onset and progression of Alzheimer's Disease; to establish global standards to identify and document cognitive changes; and to share data across the international research community. The North American ADNI has received more than 150 million USD of funding from 28 members of the public-private partnership, and includes imaging, genetic and clinical data from medical examinations and cognitive tests. ADNI extends beyond North America, however, and has partners in all major regions of the world. Data is shared with qualified researchers via a protected website.

AddNeuroMed was part of the EU 6th Framework Programme-supported InnoMed⁵⁰ project and contained both pre-clinical and clinical elements for dementia research, with about 700 patients from 6 centres across Europe enrolled in the clinical part between 2005 and 2008. The data from AddNeuroMed was incorporated into EMIF, the European Medical Information Framework (<http://www.emif.eu> supported via the Innovative Medicines Initiative), which enables data collection, integration, and exploration around multiple diseases, including dementia. While the EMIF website links information about various datasets together via a standardized browser, the actual data remains with the original researchers, which they provide to other researchers on a request basis.

Interoperability of these datasets, however, remains a challenge⁵¹ due to a wide variety of data collection methods across studies, as well as legal and consent-related protections designed protect institutions and individuals have an unintended side-effect of also preventing linkage. Many of the ADNI national studies allow sharing within their national borders, but due to concerns about legal liability and the non-transferability of patient consent agreements that comply with legal frameworks within the original country of research but have not necessarily been written to allow data to leave the country, sharing beyond country borders is often difficult if not impossible. However, virtual safe settings tools can in some circumstances allow analytic tools to be brought to the data being held securely rather than exporting the data to be used by analytic tools. When data cannot be shared, however, this limits the ability of these data to be combined for the purposes of medical research, but also limit the possibility that data at a more granular level than aggregated summary data can be used to inform science policy and health policy beyond the national level.

Routine administrative data such as hospital records, insurance claims, and other care records are increasingly seen as a potentially fruitful area for enhancing research and policymaking. However, these data, which are collected primarily for organizational purposes in the health sector, have until relatively recently remained of interest primarily for these operational purposes such as long-term patient record keeping, billing, and other administrative functions.

However, increasingly researchers are gaining access to these routine data for legitimate research purposes. The UK Biobank, for instance, follows individuals over a long period of time (at least 25 years), much of which is done through linkage with other forms of routine health data from electronic medical records in secondary and -- in the near future -- primary care. Likewise, in Sweden where patients have a lifelong single identifier (their personal ID or *personnummer*) linked to all the hospital registers, drug registers, and medical records, that allow both researchers and medical policymakers to create very big and comprehensive longitudinal databases.

The value of these data lies partly in the indeterminate nature of dementia diagnoses: because clear symptoms only appear once the disease is advanced, the availability of routine health data allows researchers to look backwards in time (in a manner) to the years and decades prior to the diagnosis to look for patterns in previous tests,

⁵⁰ http://ec.europa.eu/research/press/2006/pdf/case_study_innomed_en.pdf

⁵¹ Efforts to increase interoperability in Europe include the IMI EPAD project <http://www.imi.europa.eu/content/epad> which builds on existing national and regional registers of people at risk of developing Alzheimer's dementia to create a pan-European EPAD register.

diagnoses, treatments, and other information in the routine data to look for early signals that are associated with a later diagnosis of dementia.

Micro-economic data

One potentially promising area that is currently largely untapped is the inclusion of non-medical data that nevertheless has the potential to shed light onto dementia. Clive Humby is one of the originators of the Tesco Clubcard in the UK, which is a customer loyalty card programme as part of one of the UK's largest supermarket chains:

“I understand loyalty card and consumer data very well, so the thing that really rang home for me was that before you're diagnosed [with dementia] and perhaps have early signs, you are starting to develop coping strategies. One of the coping strategies in the early stages is to start showing more regular or habitual behaviour: So people start doing the same thing over and over again, as a way of having a routine. And it struck me that that's where loyalty card data could be hugely valuable.” (Clive Humby, quoted in Deetjen et al. 2015, p. 9)

This idea of mining consumer behaviour data for signs of developing dementia is very new, and as yet, not scientifically tested. Furthermore, even if such signals are able to be detected in commercial data, sharing these results with medical personnel, researchers, and medical policy teams (not to mention the identified individuals) is fraught with potential difficulties in terms of personal privacy concerns, issues of consent for data reuse, and the technical question of how to link such data with medical data reliably, safely, and ethically. First and foremost of these issues is determining how cases of individuals thought to be at risk for developing dementia would be handled, both ethically and medically.

Personal physical activity data

In a similar vein, the potential for online data and crowdsourced data is also very real, but as yet largely unrealized in the area of dementia. With the increasing popularity of personal data devices such as Fitbit that allow highly granular personal health data to be stored online, there is huge potential for linking these data with other sources of medical data to be able to analyse individual health and activity data at an unprecedented level of detail. **Mobile devices are particularly interesting as they can act both as a data sensor ('receptor') but also as an 'actuator' and tool of intervention via things such as health apps or even SMS-based interventions.** The question that remains, however, is how the incentives of individuals which are aligned with their own personal health concerns can be best maximized while also allowing such data to be safely and securely shared. One can imagine the potential value of highly detailed activity and movement data linked to individuals at various stages of dealing with dementia, in terms of better informing policy about caregivers, community support, and the role of both institutional and non-institutional settings in supporting those dealing with dementia.

7.4 Reflections on challenges and next steps

There is a growing global awareness of the challenge we are all facing from dementia, whether as one of the growing number of patients afflicted with the disease, or the even larger number of people affected by it through their interactions with dementia sufferers. Because of this global awareness, dementia will continue to see concerted levels of research, funding, and policy discussions in the coming years and decades.

The challenge in terms of data for policy, however, is whether all of these efforts in various countries around the world can be effectively and safely linked up so that medical research and health policy can start to massively increase the scale at which it has operated in the past. Sharing and linking medical, administrative, consumer, and activity data is extremely complex; the worst possible outcome is that the barriers to doing so will mean that efforts to deal with dementia remain fractured, ineffectual, and lacking in impact.

The next steps in this domain involve the ongoing efforts of national and international research initiatives such as those mentioned here, the many private foundations and support organizations dedicated to fighting dementia, and a commitment from governments to support data sharing efforts that are ethically sound but also push the boundaries of what might legitimately be considered useful health-related data.

Further reading

European Directorate-General Health and Food Safety information on Dementia: http://ec.europa.eu/health/major_chronic_diseases/diseases/dementia/index_en.htm

References cited

Alzheimer's Disease International (2015), "World Alzheimer Report 2015, The Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost and Trends", <http://www.alz.co.uk/research/world-report-2015>

Deetjen, U., E. T. Meyer and R. Schroeder (2015), "Big Data for Advancing Dementia Research: An Evaluation of Data Sharing Practices in Research on Age-related Neurodegenerative Diseases", *OECD Digital Economy Papers No. 246*, OECD Publishing. <http://dx.doi.org/10.1787/5js4sbddf7jk-en>