Applying Big Data to One of the World’s Biggest Problems: Alzheimer’s Disease

A Strategy Paper Series

Authored by:

Cynthia Duggan
Associate Director, Alzheimer's Disease and Dementia Initiative, The New York Academy of Sciences

Drew Holzapfel
Executive Director, The Global CEO Initiative on Alzheimer’s Disease (CEOi)

George Vradenburg
Convener, The Global CEO Initiative on Alzheimer’s Disease (CEOi)

Chris Young
Life Sciences & Healthcare Partner, Knowledgent
A Global Fight to Stop a Looming Crisis

Alzheimer’s is a global crisis. Nearly 44 million people worldwide suffer from dementia, and this number will spike to 115 million by 2050. The crisis is not particular to the rich world. Nearly 60 percent of the burden of dementia is in low- and middle-income nations. And this percentage will rise.

The global crisis demands a global solution.

The cost of care is catastrophic. It drives millions of households below the poverty line, and it places enormous burdens on health systems and public budgets. According to Alzheimer’s Disease International, the total costs of care equal a staggering $604 billion annually, or 1 percent of global GDP. This level of spending is simply unsustainable. Left unchecked, Alzheimer’s is a grave threat to the global economic system.

More and more, governments and multi-governmental organizations are recognizing the looming public health, fiscal, and economic damage Alzheimer’s disease and dementia will cause. In April 2012, the World Health Organization (WHO) identified dementia as a “public health priority,” and the G7, under UK Prime Minister David Cameron’s leadership, has declared that we launch a global “fightback” against Alzheimer’s. Around a dozen countries – including the United States, France, Great Britain, Australia and South Korea – have adopted national plans or strategies to address the disease as well as its public health and financial impacts. In its first-ever National Plan to Address Alzheimer’s disease, the United States dedicated itself to prevent and effectively treat the disease by 2025. This bold goal was subsequently adopted by the G7.

Barriers to Progress

While billions are spent every year on Alzheimer’s research, progress is slow. Hopeful patients and families have not seen a new drug reach the market in over a decade. More than one hundred years after the disease was first classified by Dr. Alois Alzheimer, we still do not know the molecular basis of the disease, and there are no drugs in the market that address the disease’s underlying pathology. Even with the much-needed increase in leadership against Alzheimer’s, there remain incredible barriers to breakthroughs in diagnosis, treatment, and care delivery, including:

- There is no clear and prioritized Alzheimer’s research agenda.
- The process to develop an Alzheimer’s therapy is long and costly.
- Private investment in R&D is a huge gamble and becoming increasingly risky.
- Financing mechanisms and incentives – both public and private – neither allow for risk-sharing in early phase therapy development nor support the development of cost-saving care interventions.
- Alzheimer’s remains highly stigmatized, and its diagnosis is widely feared. This stigma hampers early detection, diagnosis, and intervention, and it discourages patients from becoming critically needed partners in clinical research.
The situation with Alzheimer’s, however, is not without bright spots. One thing that is well established within the field is that Alzheimer’s is a multifactorial disease. That is, the disease develops as a result of the interaction of genetic, non-genetic, and environmental factors. As such, research on Alzheimer’s could significantly develop if we had the ability to analyze large data streams and multiple observation points. This kind of analysis would enable researchers to develop a more sophisticated understanding of the disease’s causes, treatments, and potentials for better care delivery.

But what is needed first is the ability to analyze massive amounts of data: behavioral, genetic, environmental, epigenetic, clinical, and more. Right now, for the first time in human history, we have the volume of data and the analytical tools to begin this project.

Indeed, big data may be the greatest weapon to wield in the global fight against Alzheimer’s.

The big data revolution is in its early days, and most of the potential is still unclaimed.


What is Big Data?

“Big data” refers to information that is too large, varied, or high-speed for traditional methods of storage, processing, and analytics. The term “big data” does not refer to a single technology. It is an umbrella term used to signify a set of next-generation tools and techniques that enable data storage, management, analytics, and integration.

The rise of big data is the result of a confluence of events. On one hand, the costs of collecting, storing, and analyzing enormous quantities of data have dropped dramatically. On the other hand, there is an exponential increase in the amount of health-related data being generated. With genomics and “smart sensors” and other new technologies, consumers of all types have opened a world of possibilities for scientific discovery and diagnostic and therapeutic product development.

This confluence opens the door for healthcare research and delivery. For example, in 2012, Google predicted a flu outbreak nearly two weeks before the Center for Disease Control and Prevention by analyzing search queries. Last year, Target used big data to identify customers who were pregnant (in some cases before the customers were certain themselves). And this is just the beginning.
Big Data in Health: Real-World Application

The accolades for big data are everywhere. Big data is claimed to be “a revolution,” “a cure to healthcare’s ills,” and “the best shot” at breakthrough progress. But what does that mean in practical, actionable terms?

These solutions augment and enhance enterprise database platforms, offering a means for storing and joining both structured data (information easily stored in a database or spreadsheet, such as sales figures) and unstructured data (content mined from emails, tweets, images, etc.). In an industry that routinely “declares war” on diseases and speaks in absolute terms such as “eradicate,” excitement often precedes success. But with big data, successes are starting to mount.

Research

Biopharmaceutical researchers have begun to aggregate information in medical databases, and some are even sharing data obtained through their research with other healthcare organizations, and, in some cases, with the very people who are the source of the data. For Alzheimer’s drug development, this sharing could be pivotal. Current Alzheimer’s data is stored disparately throughout the globe, and the inability for it to “talk” is prohibitive. McKinsey summarizes the fragmentation best: “The US health care sector is dotted by many small companies and individual physicians’ practices. Large hospital chains, national insurers, and drug manufacturers, by contrast, stand to gain substantially through the pooling and more effective analysis of data.”

This is an enormous missed opportunity. As the following case studies suggest, the potential for big data to drive breakthroughs is unprecedented.

Case Study: Infectious Disease. Researchers at Toronto’s Hospital for Sick Children have applied big data to save the lives of premature babies. By converting a set of already-collected vital signs into an information flow of more than 1,000 data points per second, the researchers created an algorithm to predict which children are most likely to develop a life-threatening infection. Now, doctors can act earlier and better treat these patients.

Case Study: Project Data Sphere. Oncology drug development has seen tremendous change in recent years. One sign of success is that oncology drug approvals are becoming more and more common. Through Project Data Sphere, companies researching in oncology have started sharing patient-level clinical trial data. The aggregation of this data – with respect to barriers including intellectual property rights, privacy, cost, and other challenges – has created new insights for cancer researchers.

Case Study: Eyewire. For making sense of large datasets, some groups have turned to large crowds. Eyewire, developed by Sebastian Seung and his colleagues at MIT, is an example of complex neuronal maps solved with the help of a game anyone can play. It would take an estimated four trillion hours to reconstruct a human brain map from electron micrographs, but with hundreds of thousands of users starting with a smaller patch of brain (mouse retina), the task seems more approachable. Another game called Foldit uses the same principle of crowdsourcing large data analytics to help solve the structure of proteins.
Case Study: Genome-Wide Association Studies (GWAS). One application of mining large datasets that has been particularly productive in the research community is the search for genome-wide associations. GWAS rely on analysis of DNA segments across vast patient populations to search for DNA variants associated with a particular disease. To date, GWAS analyses have identified a handful of promising genetic associations with Alzheimer’s disease, including ApoE4.

Care Delivery
The United States spends nearly 20% of its GDP on healthcare, far more per capita than any other country in the world. It is absolutely incumbent upon healthcare systems to control costs while also maintaining the highest levels of care. With the aging of the population – and the projected increases in Alzheimer’s – this will require new ways of operating. Mobile applications tracking patient trends could prove valuable for informing healthcare systems of care needs that could forestall more serious interventions later.

Case Study: Aggregating Conversations to Improve Disease Management. Insurer Cigna captures and analyzes call center speech-to-text data in an effort to improve disease management interventions that occur over the phone. Specifically, the conversations between Care Managers and Cigna members can be recorded and translated into text and associated with the chronic condition that is the primary purpose of the call. By doing so, the member may mention a condition or symptom that the health plan did not already know about related to the condition, or the member may describe some symptoms that Cigna can automatically capture and attach to the member’s file for future reference or perhaps as feedback regarding or articulating a poor member experience, reducing overall member engagement. This information allows Cigna to be able to take action on otherwise overlooked or non-documented information.

Patient Engagement

Case Study: Expediting Clinical Trial Enrollment. Recruiting patients into clinical trials has become increasingly difficult and slow. In response, companies have begun to use vast amounts of consumer data in order to find the most likely clinical trial participants. As an example, a small biotech needed to enroll 9,000 patients into a study, and, given precedent, they anticipated that enrollment would take two years. However, by leveraging data – from shopping habits to lifestyle choices – they decreased the enrollment time to six months.

Aggregation

Case Study: Partners HealthCare. The Boston-based healthcare system, which already brings together many of the country’s best hospital systems, is now connecting its data-collecting systems to enable real-time queries, analytics, and reports at the point of care.
Real-World Use Cases for Big Data in Alzheimer’s Disease

It is more important than ever for companies in the healthcare and life sciences industries to leverage big data. Vast amounts of information are being created and delivered by electronic health records, mobile health apps, wireless medical devices, and more. Below, several use cases are discussed in order to highlight how big data analytics can improve outcomes and enable critical insights:

Semantic Analysis and Machine Learning for Early Detection of Alzheimer’s

Recent research showed that when people complain about cognitive functioning, and when these complaints are corroborated by caregivers, it may be a sign for early Alzheimer’s detection. This research, however, required both in-person interviews with caregivers and manual coding of their observations. This process, of course, is extremely time-consuming and prohibitive at scale.

However, new technologies can automatically analyze both the patients’ complaints and their caregivers’ observations, and they can blend semantic analysis with behavioral patterns. Automating the process would not only enable large-scale research, but it also would open the door for alternative corroborating evidence, like things we do with our digital devices. Additionally, if we could analyze how Alzheimer’s patients change their digital behaviors as the disease progresses, we could potentially develop a predictive model to detect onset and stage.

Case Study: Lumosity. Mobile applications or software that can track cognitive function over time could greatly reduce the amount of time doctors spend diagnosing patients. One of many such programs, Lumosity, provides a brain game platform, and data from user scores could be valuable to correlate with diagnosis and conversion to Alzheimer’s disease. This may lead to efficient early detection, which is essential if we are to intervene in neuronal loss before it becomes irreversible.

Clustering for Patient Recruitment

In the Alzheimer’s clinical trials space, patient recruitment into the trials is one of the biggest challenges a sponsor faces. The availability of big data analytics can now offer a change in the way patients are recruited. Alzheimer’s patients and their caregivers are increasingly using online resources to search for medical-related information and to post to online forums. Big data analytics can gather this data and use advanced algorithms (i.e., clustering) to gain insights into patient density and disease state. This would help sponsors make informed decisions about targeting advertising and contracting investigative sites to run the trial.
Big Data Analysis to Drive New Insights

Patient-level medical records provide information that enables us to understand both occurrences and sequencing of patient events, clinical diagnoses, and prescribed medications. Through big data analytical solutions, we can analyze these massive and disparate datasets at the macro-level to gain insights into patterns, trends, correlations, and clusters of medical and demographic information. Furthermore, with the increasing quantity of genomic information, big data will be needed to process and derive meaning from the large, complex volume of information. With information that is captured in real-world clinical settings and genomic sequences, researchers will be able to analyze new dimensions of evidence and outcomes. This is a missing piece in the journey to find new cures and advance the delivery of care.

A recent Alzforum article captures both the issue and the opportunity:

“The genetic studies we are doing now are limited by sample sizes,” noted Zaven Khachaturian, editor of the journal Alzheimer’s & Dementia, in Washington, DC. “We need to have very large samples because the genetics and the complexity of the disease is enormous.” This kind of “big data” analysis would be an excellent element in international AD efforts, agreed George Vradenburg, who chairs the advocacy group USAgainstAlzheimer’s [and Convener of The Global CEO Initiative on Alzheimer’s Disease] based in Washington, DC. Combining the genomes, blood chemistry profiles, and brain images of a large number of subjects should help researchers understand who is at risk for dementia and what biological pathways to target with treatments, he said.

Structured and Unstructured Data to Develop a Holistic Member Profile

Given the ongoing importance of patient engagement within the healthcare sector, organizations are evaluating other potential ways to attribute members to identify key preferences, experiences, and information. Also, with the number of Medicare & Retirement members currently with early onset and full Alzheimer’s, it will be extremely important to understand the household and family factors of a member. Healthcare companies can better align treatments paths and maintenance programs by evaluating attributes and clinical profiles across members, and this analysis can guide population health management.

Considerations for Use of Big Data in Alzheimer’s Disease

While the Alzheimer’s research and development community is eager to adopt cutting-edge techniques to advance the field, there are several important points to consider. First, it will be essential to ensure the privacy of patient records using sophisticated de-identification methods. This becomes trickier as DNA sequencing becomes routine, posing new challenges for anonymity. Secondly, standardization of collection and storage will be essential for analyzing data across time with many variables. Finally, data quality must be considered, particularly as new and improved methods emerge.
Taking Steps to Leverage Big Data Solutions

We are in the early days, but there are new efforts rising around big data and Alzheimer’s. The Organization for Economic Cooperation and Development (OECD) is beginning to identify research databases that can add to the big data field. OECD and Business Industry Advisory Council (BIAC) are collaborating to identify national policies that are best suited to achieve the value of open science and global biomedical data interchange. These policies are being developed to ensure that the privacy, security, and proprietary interests of individuals and innovators are protected. The G7, as part of the legacy workshop series following the 2013 Dementia Summit, will also undertake big data and Alzheimer’s.

During key proceedings in 2013, the CEOi, NYAS and Knowledgent agreed to lead forward a big data strategy. And in several ways, we have started. We have, with Sage BioNetworks, announced a series of Alzheimer’s Big Data Challenges to test the ability of new big data techniques to advance the scientific understanding of Alzheimer’s. Alzheimer’s Big Data Challenge #1, already announced, will seek to identify the best bouquet of predictors of cognitive decline and of neuroprotective genotypes. We have partnered with US national health authorities and others to extend the use of Blue Button capabilities to empower individuals to share their personal Alzheimer’s-related medical records for research purposes. We are also working to develop a global Alzheimer’s clinical trial platform, which will be underpinned by interconnected data collection and will enable mining of large data sets. These are helpful “use cases,” but more and others are needed.

Moreover, as these important early efforts in international organizations, governments, and private sectors move forward to meet the current needs, it is important to focus on what the future needs will be. This will become a central piece of our work.

To take the next step, we will invite experts across all sectors to participate in a Big Data Ideation Workshop.

In the Big Data Ideation Workshop, we will discuss and define approaches for evolving and augmenting Alzheimer’s-related research, drug development, evaluation, and care delivery using big data solutions. We will:

• Leverage cross-industry and big data expertise collaboratively to identify innovative ways to use the data we have available to us to further analytical insights.
• Define the information landscape across sectors to understand the:
  - Data each sector creates, purchases, or uses.
  - Data in use vs. not in use today.
  - Data we need to drive analysis.
• Address how technology solutions can bring together datasets that are created and owned by multiple parties.
• Identify ways to bring diverse datasets and technology together to evolve how data is used.

By exploring the potential of big data and developing a shared understanding of how to leverage it, we will be closer to finding a cure and improving the delivery of patient care.
About the Global CEO Initiative on Alzheimer’s (CEOi)

The Global CEO Initiative on Alzheimer’s Disease (CEOi) is a public-private partnership initiated by leading global corporations and involving non-profit and governmental organizations to identify and advance high-priority activities necessary to prevent and treat Alzheimer’s disease by 2025 and improve the quality of life for all affected by the disease. The CEOi has been formed to provide business leadership to the fight against Alzheimer’s. We intend to lend our voice to the social, fiscal, and political developments on Alzheimer’s in order to “change the game” by adding private-sector partnership and leadership.

For more information about CEOi, please visit www.ceoalzheimersinitiative.org.

About Knowledgent

Knowledgent is an industry information consultancy that helps organizations transform their information into business results through data and analytics innovation. Our expertise seamlessly integrates industry experience, data analyst and scientist capabilities, and data architecture and engineering skills to uncover actionable insights.

Knowledgent operates in the emerging world of big data as well as in the established disciplines of enterprise data warehousing, master data management, and business analysis. We not only have the technical knowledge to deliver game-changing solutions at all phases of development, but also the business acumen to evolve data initiatives from ideation to operationalization, ensuring that organizations realize the full value of their information.

For more information about Knowledgent, visit www.knowledgent.com.