

## Big Data, Big Problems: How Should Singapore Approach the Privacy Challenge of Big Data Research?

### Introduction

Miss Tham is a single 47-year old woman who earns close to the median national salary. She is already feeling that her age is catching up to her. Miss Tham falls ill more often and she observes that friends her age are prone to chronic illnesses such as diabetes and cancer. As she gets closer to the retirement age, she worries that she will not have enough savings to tide her through the last phase of her life. She wonders if she will have to work longer and retire later. A more worrying scenario that often comes to her mind is whether she will even be able to work much longer, and what if she is hit with a chronic illness?<sup>1</sup>

A group of researchers claimed that they could offer solutions to Singapore's healthcare expenditure problem by using big data in health and biomedical research and is lobbying the Singapore government to grant them access to all Electronic Health Records (EHR) and all digital data. Singapore was already sitting on an insurmountable amount of data, thus making access to it was not a technical issue. However, a privacy campaign group, Big Brother Watch, opposes this proposal due to privacy and security concerns. The recent scandal surrounding big data in the U.K. solidified their scepticisms – Google was given questionable access to healthcare data of up to 1.6 million patients, including sensitive data about HIV-positive status, details on drug overdoses and abortions.<sup>2</sup>

Further, nearly one million patients had their confidential medical data shared with third parties despite opting out of the care.data database<sup>3</sup> due to slow processing by the Health and Social Care Information Centre.<sup>4</sup>

You are tasked to head a special task force to look into this issue and prepare a policy recommendation to the Prime Minister's Office. Taking into consideration Singapore's current situation and the data controversy in the U.K., how should Singapore proceed in its agenda to adopt big data in its health and biomedical research?

---

<sup>1</sup> Characters are purely fictitious to describe the lives of Singaporeans.

<sup>2</sup> Ben Quinn, "Goggle given access to healthcare data of up to 1.6 million patients", *The Guardian*, May, 4, 2016, <https://www.theguardian.com/technology/2016/may/04/google-deepmind-access-healthcare-data-patients>

<sup>3</sup> An National Health Service (NHS) England initiative to upload patients' medical records to the national Health and Social Care Information databases with the objective to centralise all existing health records to provide a more holistic picture of the healthcare given to patients. Care.data would be used to find more effective ways of managing illnesses, monitor the risk of disease spread and streamline inefficiencies. Every patient was automatically included in the system unless they explicitly opted out.

<sup>4</sup> Sarah Knapton, "How the NHS got it so wrong with care.data", *The Telegraph*, July, 7 2016, <http://www.telegraph.co.uk/science/2016/07/07/how-the-nhs-got-it-so-wrong-with-caredata/>

---

This case has been written by Harisan Unais Nasir and Kershia Tan Wei under the guidance of Markus Labude, Centre of Biomedical Ethics, National University of Singapore. The case does not reflect the views of the sponsoring organization nor is it intended to suggest correct or incorrect handling of the situation depicted. This case is based on actual events but characters and events are fictional. The case is not intended to serve as a primary source of data and is meant solely for class discussion. This case was awarded a Distinguished Prize in the Lee Kuan Yew School of Public Policy Case Writing Competition 2017/2018.

## Singapore's Demographic "Time-Bomb"

In 2018, the percentage of senior citizens (i.e. above 65 years old) was equal to that of juniors (i.e. below 15 years old) and further projections showed that the percentage of seniors would overtake the juniors by 16.2% by 2030<sup>5</sup> (Exhibit 1). In comparison with other countries, Singapore's 2030 situation would be worse than Japan's current conditions<sup>6</sup> (Exhibit 2), and its dependency ratio would fall from 1: 5.7 to 1:2.6 by 2030.<sup>7</sup> Inevitably, the growing ageing population would strain Singapore's healthcare system.<sup>8</sup> In 2015, Singapore's Current Health Expenditure (CHE) per capita was the highest among ASEAN countries (i.e. US\$2,280) and was also comparable with other Asia-Pacific developed countries such as Republic of Korea (US\$2,013), New Zealand (US\$3,554) and Japan (US\$3,733)<sup>9</sup>. Nonetheless, what was interesting is that although the reported figures of the developed countries were high, Australia, Japan and New Zealand had exhibited a decreasing trend in their CHE per capita (Exhibit 3).

With Singaporeans living longer and having fewer children, its population was rapidly ageing and shrinking. This created a daunting challenge for Singapore's healthcare sector as well as management of its national resources, with its repercussions increasingly felt today. With an ageing population, demand for healthcare and related healthcare services would increase – healthcare expenditure for the ageing population was expected to hit \$44 billion (2013: \$17 billion) and hospital admissions to hit 791,000 (2013: 15,000) annually by 2030.<sup>10</sup> Thus, efforts to help reduce costs in healthcare services and research were underway.

## Singapore's Pro-Biomedical Research Approach

Singapore has been supportive with regard to biomedical research. In the early 2000s, Singapore started its push towards becoming a biomedical hub. The move was part of a larger effort to transform Singapore into a knowledge-based economy.<sup>11</sup> A biomedical hub was seen as having the potential not only to create jobs, but jobs that require high levels of technical skill and salaries commensurate to that skill. Demand for high-skilled jobs also fitted well into Singapore's population demographics. Singapore's population was highly educated. It was also home to two research-intensive universities that were ranked in the top 20 globally and were the best two universities in Asia by some measures; no small feat for a country this size.<sup>12</sup> However, a population that was highly-educated would also demand a different type of work in order to be fulfilled. Thus, for both economic and social reasons, biomedical research was strongly supported in Singapore.

The increasing gross expenditure on research and development (R&D) was a promising sign of the biomedical research's contribution to Singapore's gross domestic product (GDP). In 2015, R&D

---

<sup>5</sup> Francis Tan, "Singapore: Reaching A Critical Demographic Crossroad", *UOB Global Economics & Market Research*, December 6, 2017, 2, [http://www.uobgroup.com/assets/pdfs/research/MN\\_171206.pdf](http://www.uobgroup.com/assets/pdfs/research/MN_171206.pdf)

<sup>6</sup> Francis, "Singapore", 3.

<sup>7</sup> Francis, "Singapore", 9.

<sup>8</sup> Private and public healthcare costs have risen at a rate of 18% and 12% yearly respectively (<http://www.straitstimes.com/singapore/health/prevention-is-better-than-cure-as-costs-keep-rising>)

<sup>9</sup> World Health Organisation, *Global Health Expenditure Database*.

<sup>10</sup> Rachel Tan and Andy Sim, "Singapore Healthcare", *DBS Asian Insights*, July 4, 2016.

<sup>11</sup> Ministry of Health, "Achievements of the Biomedical Sciences Initiative, Annex D".

<sup>12</sup> Quacquarelli Symonds World University Rankings 2017.

expenditure in biomedical and related sciences was recorded at \$1.9 billion, an 18.8% increase compared to the previous year (2014: \$1.6 billion)<sup>13</sup> and, as at end-2016, the biomedical sector had produced products worth more than \$16 billion for global markets.<sup>14</sup> The increasing R&D expenditure, coupled with the government's commitment to further increase investments to \$19 billion for the next five years in research,<sup>15</sup> was a huge testament of the unlocked benefits of the sector. Consequently, with increased investment, came increased number of high-skilled jobs created. Overall, in 2015, the number of R&D jobs in Singapore had reached 44,669, signalling a 6.1% increases compared to 2014.<sup>16</sup> Of these employees, 10,300 and 8,321 were PhD and Masters holders respectively, of which both recorded more than 6% growth from the previous year,<sup>17</sup> moving in tandem with Singapore's transformation into a knowledge-based economy.

Moreover, Singapore expected biomedical research to make a big impact on its economy. The Ministry of Health was building up the biomedical research sector as the 'fourth pillar' of the economy.<sup>18</sup> In other words, it was not only that Singapore provided the right environment for big data research through its investments in R&D, it was also expected that biomedical research would make up a big part of its GDP output. The use of big data might drive the biomedical research economy to this goal. The increase in GDP could then be used to shoulder the additional healthcare expenditure that it would need in the near future.

### **Big Data in Health and Biomedical Research in Singapore**

Big data was not a new term but was popularised in recent years with the help of advancements in technology and analytics. Data was not just limited to official records but also included other digital data sources such as online web searches, posts and images. Given the increasing availability and vast potential of real-time digital data, the 21st century was dubbed the 'golden information age' – a potential that was so huge and unmissable that the world was, arguably, standing on the brink of a fourth industrial revolution.

The health and biomedical research industries were no exception as beneficiaries to this explosion of data. In fact, the healthcare industry had been a large biomedical data generator. Since 2011, Singapore had undertaken efforts to bring together all patients' records such as medication and laboratory reports from different healthcare providers by establishing the National Electronic Health Record (NEHR) system.<sup>19</sup> Advocating "One Patient, One Health Record", the NEHR was progressively introduced to public and private healthcare institutions across Singapore and was suggested to be

---

<sup>13</sup> SingStat, Taken from Agency for Science, Technology and Research, "Research and Development Expenditure by Area of Research, Annual", last updated January 2, 2018

<sup>14</sup> Economic Development Board Singapore, "Pharmaceuticals and Biotechnology", 2018, <https://www.edb.gov.sg/en/our-industries/pharmaceuticals-and-biotechnology.html>

<sup>15</sup> Chang Ai-Lien, "Singapore commits record \$19b to R&D", *The Straits Times*, January 9, 2016, <http://www.straitstimes.com/singapore/singapore-commits-record-19b-to-rd>

<sup>16</sup> Agency for Science, Technology and Research Singapore (A-STAR), "National Survey of Research and Development in Singapore 2015", December, 2016, <https://www.a-star.edu.sg/Portals/81/Data/News%20And%20Events/Publications/National%20Survey%20of%20R&D/Files/RnD%202015.pdf>

<sup>17</sup> A-STAR, "National Survey".

<sup>18</sup> Ministry of Health, "Achievements of the Biomedical Sciences Initiative, Annex D

<sup>19</sup> The NEHR is owned by the Ministry of Health and managed by the Integrated Health Information Systems.

mandatory for all healthcare institutions once the new Healthcare Services Bill was introduced.<sup>20</sup> With the implementation of the NEHR system, not only were authorized healthcare professionals able to access patients’ digitised records to assist them in forming a more holistic and longitudinal view of the patients’ healthcare history, the system also provided an opportunity to build further analytics capabilities for bio-surveillance, population health management or cutting readmission.<sup>21</sup>

As a society that was technologically advanced, wired and had a high mobile penetration rate, much of an individual’s personal information was stored in large databases. This would be further expanded with the introduction of the NEHR system that aimed to collect and retain health information of patients nationwide. Information would already be stored and researchers would not need to perform the extra costly step of collecting them from each individual.

### Types of Data Available in the Healthcare Sector

In simple terms, within the healthcare industry, there were generally three groups of data (Table 1). The first could be categorised as “administrative” data. This included health records such as drug prescriptions, doctor’s diagnosis and number of clinical visits. Other derivative information such as clinical bills and insurance data were also captured. The second was biomedical data such as genomic data and data from clinical trials. Genomic studies were part of “big data” due to the sheer size of the information stored in our genes and due to the large sample sizes required to make sound conclusions. Individual genes were “big” in the sense that they contained a large volume of information. Genetic studies were also “big” because they required a large sample size. Lastly, one other point of concern was patient-generated, non-health data such as tweets, Facebook and even fitness trackers.

Table 1: Examples of the types of data used in biomedical and health research.

| Administrative Data                     | Biomedical Data                                     | Non-health data  |
|---|---|------------------|
| Electronic Health Records               | Genomic Data  | Credit Card data |
| Clinical and administrative information | Clinical Trials                                     | Facebook data    |
| Public and Private insurance data       | Patient-generated data with health-tracking devices | Twitter feeds    |
|   |   | Geospatial data  |

The first instance of using “non-health” data for the purpose of healthcare could be traced back to 1854 in the infamous Broad Street cholera outbreak. In combining geospatial information about infected patients, John Snow was able to conclude that the spread of cholera was not due to “bad air” or “miasma” as prevailing doctors thought. Instead, he figured out by mapping the clusters of

<sup>20</sup> Chan Luo Er, “National electronic patient database soon to be mandatory for healthcare providers”, *Channel NewsAsia*, November 8, 2011, <https://www.channelnewsasia.com/news/singapore/national-electronic-health-record-mandatory-clinics-9385666>

<sup>21</sup> Manik Bhandari, “Is data Singapore’s next big bet?”, *The Straits Times*, last modified February 9, 2017, <http://www.straitstimes.com/opinion/is-data-spores-next-big-bet>

cholera outbreaks that it was due to a contaminated water pump.<sup>22</sup> This was one early example of how non-health data such as geolocation information could produce findings in healthcare when combined with other information. This made the division between “health” and “non-health” data contentious. In more recent cases, mobile phone data had also been used to predict depression rates.<sup>23</sup> One thing that seemed clear was that with the right amount of analysis, traditional non-health data could become relevant and important to health research.

## Big Data Applications

*Bioinformatics applications*, where biological data was analysed at a molecular level to enable more personalised medicine and healthcare services. The advent of big data techniques had allowed for larger data repositories, more sophisticated computing infrastructure and more efficient data manipulation tools to store and analyse biological information.<sup>24</sup> As genomic and biomedical data accumulated, bioinformatics applications were helping health providers and researchers manage and analyse data. This presented an opportunity for better identification of disease susceptibility genes, development of new drug treatments, more accurate predictions of adverse reaction risks and possible reclassification of diseases based on genetic characteristics.<sup>25</sup>

*Clinical informatics applications*, which utilised clinical information and information technology for the delivery of healthcare services.<sup>26</sup> Unlike bioinformatics, data analysis using clinical informatics used both structured and unstructured data to develop specific ontologies.<sup>27</sup>

*Imaging informatics applications*, which utilised medical images, alongside EHR, to optimize clinical diagnosis and treatment.<sup>28</sup> For example, using artificial intelligence and big data to supplement the analysis of cardiac imaging and medical data, it was found that cases of inappropriate use of diagnostic imaging were reduced from 10% to 5%.<sup>29</sup>

*Public health information*, whereby data was used for three core functions: (1) assessment which involved collecting and analysing data to track and monitor health status, thereby leading to, (2) decision making and policy development, and (3) assurance that was used by relevant authorities to validate their targeted goals<sup>30</sup>. An example was infectious disease surveillance. Using big data for

---

<sup>22</sup> Judith Summers, *“Soho -- A History of London's Most Colourful Neighborhood”*, Bloomsbury, London, (1989): 113-117. Retrieved from: <http://www.ph.ucla.edu/epi/snow/broadstreetpump.html>

<sup>23</sup> Saeb, S., M. Zhang, C.J. Karr, et al, “Mobile Phone Sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study”, *J Med Internet* 17, no.7 (2015): e175. doi:10.2196/jmir.4273.

<sup>24</sup> Jake Luo, Min Wu, Deepika Gopukumar and Yiqing Zhao. “Big Data Application in Biomedical Research and Health Care: A Literature Review”, *Biomedical Informatics Insights*, no.8 (2016): 2, doi: 10.4137/Bii.s31559

<sup>25</sup> Himanshu Singh. “Bioinformatics: Benefits to Mankind”. *International Journal of PharmTech Research* 9, no.4 (2016): 243.

<sup>26</sup> Jake, Min, Deepika and Yiqing, “Big Data Application”, 4.

<sup>27</sup> Jake, Min, Deepika and Yiqing, “Big Data Application”, 4.

<sup>28</sup> Jake, Min, Deepika and Yiqing, “Big Data Application”, 6.

<sup>29</sup> Steven E. Dilsizian and Eliot L. Siegel. “Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment”. *Curr Cardiol Rep* 16, no. 441(2014): 3.

<sup>30</sup> Shortliffe EH and Cimino JJ. *Biomedical Informatics*. Berlin: Springer; 2014.

global infectious disease surveillance, Hay et al.<sup>31</sup> developed a system, using online social media and epidemiological information, to provide real-time risk monitoring, of which real life application was demonstrated by Young et al.<sup>32</sup> for HIV cases. Using social media, 553,186,016 tweets were collected and more than 9,800 HIV risk-related keywords were extracted<sup>33</sup>. Their results demonstrated the potential impact of social media in helping to monitor global disease occurrences.<sup>34</sup>

## Arguments for the Usage of Big Data for Health and Biomedical Research

### *The Clinical Trend Argument*

Evidence-based medicine was not without its criticism. However, one could not deny its epistemological strength. As in any other scientific field, it would be difficult for a clinician to justify his findings and medical decisions based on contentious means such as intuition. An evidence-based approach would stand more strongly against critics as the data in which decisions rely on would have to be refuted. As the healthcare sector moved towards more evidence-based medicine, whereby treatment for individual patients were specifically curated based on available evidence, big data stood to play a more prominent role. Health practitioners were moving away from traditional self-judgement and increasingly relying on evidence in making treatment decisions.<sup>35</sup> The nuances in subpopulations, which were previously not readily apparent in small data sample sizes, could now be more visible through aggregating individual data sets into big data algorithms, thus providing more robust evidence to health practitioners.<sup>36</sup>

### *The Time-Sensitivity Argument*

Time is of essence when conducting health and biomedical researches. With big data, researchers had access to huge volume of medical records, images and test results to assist them in making better-informed decisions faster. Biomedical studies were “big” in nature – the more data points we had, the more accurate our decisions became. Using large datasets, researchers at London’s Institute of Cancer Research identified genes that correlate with cell shape features in breast cancer cells, which subsequently assisted physicians to administer treatment accordingly.<sup>37</sup> Such complex analysis would typically take decades but with the help of big data, the time period could be reduced to mere months.

### *The Altruism Argument*

Another argument for the use of big data, including sharing of data sets, in health and biomedical research was that individuals wanted to see progress in science. Typically, individuals with conditions with no available cure would benefit greatly from such progress. The integration of large clinical and

---

<sup>31</sup> Hay, Simon I., Dylan B. George, Catherine L. Moyes, and John S. Brownstein. "Big Data Opportunities for Global Infectious Disease Surveillance." *PLoS Medicine* 10, no. 4 (2013): e1001413.

<sup>32</sup> Young, Sean D., Caitlin Rivers, and Bryan Lewis. "Methods of using Real-Time Social Media Technologies for Detection and Remote Monitoring of HIV Outcomes." *Preventive Medicine*, no. 63 (2014): 112-115.

<sup>33</sup> Young, Rivers and Lewis, "Methods", 113.

<sup>34</sup> Young, Rivers and Lewis, "Methods", 113.

<sup>35</sup> Basel Kayyali, David Knott and Steven Van Kuiken, "The big-data revolution in US health care: Accelerating value and innovation", *McKinsey & Company*, April, 2013, <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care>

<sup>36</sup> Basel, David and Steven, "The big-data revolution".

<sup>37</sup> Mike Larrobino, "3 Ways Big Data Affects Biomedical Research", *Copyright Clearance Centre*, March 21, 2017, <http://www.copyright.com/blog/3-ways-big-data-affects-biomedical-research/>

biomedical data sets was an incredibly powerful resource to expedite medical discovery. Take the fight on Alzheimer as an example. Though total eradication of Alzheimer might be a long way to go, substantially reducing its impact was within reach. Having access to enormous data sets, researchers could now better understand how Alzheimer deviates from normal ageing, how the brain changed and how these changes differed based on gender, lifestyle, ethnicity or genetics.<sup>38</sup> Early identification of risk had become more probable and with comprehensive brain scans, genetics, medical records and memory testing results, health practitioners would be able to calculate the “risk score” and identify new targets for treatment.<sup>39</sup> There was potential for these successful results to be replicated for other diseases that currently did not have any available cure.

### *The Collective Duty Argument*

The data that was generated by the NEHR system benefited the citizens that it served. By having such a system, citizens would have a better healthcare delivery system. Doctors could access the database, see a patient’s medical history in full detail and prescribe the relevant therapies that are appropriate. It also saved cost for citizens. If someone had taken a medical test in one hospital but decided to change hospitals after a few visits, his new doctor could easily access the test results and the patient need not repeat the test at the new facility. All these benefits were only possible because the government had invested into data collection and storage systems as well as training public health workers. It was argued that because these investments were made by the government, citizens had an obligation to share their health data.<sup>40</sup> Citizens received a wide-range of benefits from the NEHR and other data services that the government provided, thus, they seemed to have an obligation to return in-kind at least by allowing their health data to be used for research for the further benefit of the community.

## **Arguments Against the Usage of Big Data for Health and Biomedical Research**

### *Biomedical Research Ethics*

Traditionally, biomedical research had been guided by the ethical theory of principlism. This theory demanded that researchers balanced four basic ethical principles: (1) Respect for Persons – treating agents as autonomous agents (2) Beneficence – research should maximise potential benefits (3) Nonmalificence – human subjects should not be harmed (4) Justice – the benefits and risks of research should be fairly distributed. These principles were first formalised in the Belmont Report in 1978 by the United States National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.<sup>41</sup>

The era of big data technologies presented a threat to current practices in the research community. Firstly, the principle of beneficence and our need to respect the autonomy of individuals were in tension with one another. From the principle of beneficence, big data research should be promoted due to its potential to benefit everyone. However, as we will show below, issues of privacy have

---

<sup>38</sup> Bill Gates, “Bill Gates: We must share data to fight Alzheimer’s”, *Financial Times*, December 14, 2017, <https://www.ft.com/content/dc27da0a-e020-11e7-a0d4-0944c5f49e46>

<sup>39</sup> Bill Gates, “Bill Gates”.

<sup>40</sup> Ballantyne, Angela and Schaefer, Owen. “Consent and the ethical duty to participate in health data research.” *Journal of Medical Ethics*, (2018): doi: 10.1136/medethics-2017-104550

<sup>41</sup> National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Department of Health, Education and Welfare (DHEW), September 30, 1978. *The Belmont Report*. Washington, DC: United States Government Printing Office.

cropped up with the use of big data. The issue of privacy, consent and autonomy did not come apart so easily. In order to respect a person's autonomy, researchers must gain informed consent from participants. Ensuring the privacy of data subjects has been part-and-parcel of consent agreements in research. As much as possible, we would want research participants to be autonomous in their decision to be part of any research project. Respecting the autonomy of research participants has been a crucial core value that has driven much of today's best practices in research. In turn, this has brought benefits to the research community as well. Citizens with a high level of trust in researchers might be more willing to participate in research programs.

### *The Impossibility of Privacy*

The type of data used in big data research sometimes makes privacy impossible on several accounts. Firstly, the richness of data is sometimes necessary for research purposes. Unlike traditional research that begins with a hypothesis followed by the collection of data, big data research sometimes began with the analysis or collection of data in order to look for possible trends and relationships. This made anonymization and de-identification problematic. By removing parts of data that were traceable to the data subject, we also risked removing data that might have valuable research potential.<sup>42</sup> Some types of data used in biomedical research were almost impossible to anonymise. One clear example was genomic data. Genetics were inherently identifiable due to its uniqueness and it was almost impossible to remove identifiable parts of your genes without losing some important information. Some claims of anonymization had been debunked in some instances, which questioned the very possibility of privacy and de-identification in big data research.<sup>43</sup> Secondly, data linkages could sometimes reveal the identity of data subjects. One study showed that by using various publicly available data sources, researchers could narrow down the identity of data subjects from a Facebook study that were supposed to be anonymised.<sup>44</sup>

Keeping data under strict lock and key was not the ideal solution to the lack of privacy. It would also be counter-intuitive to the very goal of big data research. Big data research required the use of many different sources of data and keeping research data immensely private erected a barrier towards data sharing. Data sharing itself provided immense benefits and the way forward was to guide data sharing in an ethical manner, rather than to stop it altogether. Several medical research projects that centred on data sharing were already ongoing and were likely to grow.<sup>45</sup> Furthermore, as highlighted by the principle of beneficence, we ought to maximise the potential benefits of the contributions of research subjects. However, it was reasonable to expect that there were fears that the research participants took on a disproportionate amount of risk and would be harmed when their data was made too widely available.<sup>46</sup>

---

<sup>42</sup> Janet Currie, "Big Data" Versus "Big Brother": On the appropriate use of large-scale data collections in pediatrics." *Pediatrics* 131, Supplement, (2013):S127–S132. doi:10.1542/peds.2013-0252c.

<sup>43</sup> Ohm (2010), Narayanan & Felten (2014) argue extensively that de-identification is ineffective. Gymrek et al. (2013) also showed that it was possible to trace participants in a supposedly "de-identified" open genome database.

<sup>44</sup> Michael Zimmer, "More on the "Anonymity" of the Facebook dataset—It's Harvard College." October 3, 2008, <http://michaelzimmer.org/2008/10/03/more-on-the-anonymity-of-the-facebook-dataset-its-harvard-college/>

<sup>45</sup> Poldrack & Gorgolewski (2014) discusses the progress of data sharing of neuroimaging data to maximise the contribution of human subjects.

<sup>46</sup> Sharona Hoffman, "The Promise and Perils of Open Medical Data." *The Hastings Centre Report* 4, no.1 (2016):6-7.

### *Group-Level Harms*

Moreover, anonymization of individual datasets was not the silver bullet to solve these issues. Even if anonymization and de-identification could work, there was still a possibility of group-level harms affecting certain social groups. Findings from big data research in health could show that a certain race, socio-economic background or even geographical living space correlated with a certain health outcome. What this means is that big data research would predict outcomes for groups instead of individuals in the case of anonymised data. The revelation of relationships between such groups and their health could result in unwanted discrimination against people who belong to these groups. Therefore, groups still faced potential harms even if their data could not be traced back to them individually.<sup>47</sup> Thus, regardless of whether data could be anonymised, such instances of group-level harms posed a certain risk on data subjects. In a multiracial society like Singapore, such discrimination could be extremely harmful to the social fabric. Singapore had no anti-discrimination laws concerning issues such as employment and insurance.<sup>48</sup> This problem was exacerbated by the fact that the Singapore government regularly issued public data that was broken up by race. For instance, as seen in Exhibit 4, the proportion of Singaporeans with diabetes was revealed in 2017 and broken down based on race. This way of revealing statistics could be potentially harmful to certain racial groups as they might end up being discriminated against with higher insurance premiums based on their race alone.

### *The Difficulties of Informed Consent*

In traditional research programmes, participants were informed about a specific research program and gave their informed consent for a specific purpose. Some researchers had noted that this method of gaining consent was impractical for big data research. Take the example of using centralised electronic health records. Singapore's upcoming health database would be extremely valuable for research. If we were to follow the traditional rules of informed consent, researchers would have to approach each and every single person in this database, inform them of the research program and ask for their consent. Such a task would be extremely costly or even impossible if researchers want to make full use of the database. However, the data was already available and there might be several benefits if research was done.

Datasets could also be re-used for many other purposes down the line and it was impossible to tell what it can potentially be used for in the future as there is an "impossibility of certainty concerning future uses of data".<sup>49</sup> It seemed that it would also be a great waste to collect data for just one research project and never use it again. For instance, the sequencing of a person's genome might come at a great cost and would be a great benefit to science and medicine if the data were used more than just once. Thus, there were significant limits in our ability to fully inform data subjects about what their data would ultimately be used for in the future.

### *The Use of Non-traditional Data in Big Data Health Research*

Big Data also made use of datasets that might not come with a participant's explicit consent to involve themselves in a specific research program. For example, researchers might use consumer purchasing trends and social media habits of persons without their explicit consent for the specific use of medical

---

<sup>47</sup> Docherty, "Big data: Ethical perspectives." *Anaesthesia* 69, no.4 (2014): 390–391. doi:10.1111/anae. 12656.

<sup>48</sup> TAFEP, "Why is there no legislation on workplace discrimination in Singapore?" <https://www.tafep.sg/3-why-there-no-legislation-workplace-discrimination-singapore>

<sup>49</sup> Mittelstadt & Floridi, "The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts." *Science and Engineering Ethics* 22, no.2 (2016):303-41.

research. Social media companies like Facebook and Twitter usually demanded that individuals agree to end-user license agreements (EULA) where they forfeited their right to their data before they could access the company's' services. However, such broad agreements did not amount to the level of informed consent that we would want to achieve in the medical research community. Although EULAs satisfied legal requirements, for health and biomedical research purposes, they did not meet a stronger ethical requirement. Within the field of research, these agreements did not amount to informed consent because they failed to provide adequate information to the data subjects on what would be done with their data. There were key difficulties that were inherent in big data research that was not prevalent in traditional research methods.

Ignoring these issues could lead to severe public backlash. One example of how privacy and consent issues led to a severe backlash is the care.data initiative in the U.K. The care.data programme aimed to centralise data from general practice surgeries and hospitals across the country to create a database of patient records for the use of healthcare delivery and research. However, it faced problems and backlash from various stakeholders over the issue of privacy that led to its early demise.<sup>50</sup> In one study, researchers listed several major concerns that were raised by citizens about the care.data programme.<sup>51</sup> Three of these were related to privacy and trust: lack of transparency; lack of respect for confidentiality and privacy; erosion of trust in GPs and the healthcare system. These factors eventually led to a pushback against the care.data programme and the open nature of the policy was eventually reversed. In order to avoid the problems that care.data faced, any national data policy that emphasised an open data approach had to take privacy, confidentiality and citizen trust seriously.

### **Data Privacy, Trust and Accountability in Singapore's Context**

In Singapore, data privacy laws only recently took effect with the legislation of Protection of Personal Data Act (PDPA) in 2012. Amongst other things, the PDPA legislated that organisations had to ask for consent before collecting data and were responsible in making "reasonable security arrangements to protect personal data they possess(ed) or control(led)".<sup>52</sup> However, these new regulations did not extend to ministries and government agencies. There are a few reasons why this was so.

Firstly, as part of the PDPA, persons could retrieve any information that any organisation had about them and could demand to change that information in any way they saw fit. This would not be ideal for government agencies to have. Due to security issues, governments regularly kept information on people that needed to be kept confidential. However, this did not change the fact that public agencies need to be held responsible in safeguarding information that was kept. There were already cases of security lapses involving public agencies and services in Singapore. For instance, in 2014, 1,500 SingPass accounts were breached.<sup>53</sup> SingPass allowed Singapore residents to perform various transactions with public agencies such as filing income taxes and accessing medical records. Following

---

<sup>50</sup> Carter et al. "The social licence for research: why care.data ran into trouble", *Journal of Medical Ethics* (2015) 41, 5 (2015):404-9.

<sup>51</sup> Sterckx et al. "You hoped we would sleep walk into accepting the collection of our data: controversies surrounding the UK care.data scheme and their wider relevance for biomedical research" *Medicine, Health Care and Philosophy* 19, (2016):177-190.

<sup>52</sup> "Data Protection and You", Personal Data Protection Commission. <https://www.pdpc.gov.sg/individuals/Overview>

<sup>53</sup> Ng Jing Yng, "1,560 SingPass user accounts breached" Todayonline. June, 5, 2014. <http://www.todayonline.com/singapore/1560-singpass-user-accounts-breached>

this breach, SingPass moved to a 2-factor authentication system to improve security. However, the damage had already been done.

Within Singapore’s public health system, there had already been cases of unauthorized access of patient data by healthcare providers.<sup>54</sup> In one case, a plastic surgeon had abused the information system of a public hospital, Singapore General Hospital, to access information of a person without authorisation to do so. A complaint was lodged with the Singapore Medical Council (SMC), however, no formal inquiry was made and a mere “letter of advice” was issued to the perpetrator.<sup>55</sup> Already, privacy breaches garnered a call for stronger punishments and enforcement from the state, especially in the case of one car-sharing company which had its consumer data stolen.<sup>56</sup>

Trust and accountability were intrinsically related. To retain the trust of people, agencies that shared, stored and collected data had to be held responsible for their safekeeping. This was of serious concern as we entered what some might call the 4th Industrial Revolution. A lack of trust between the public and the state would hinder national policies that aimed to expand data collection and data sharing. Without trust, data sharing initiatives might fail as they would face the same backlash as care.data in the UK. A policy that leaned too heavily towards privacy protection might restrict the possibility of accessibility and dissemination of data and information. Privacy and trust had to go hand in hand. Simply having privacy safeguards would still be insufficient if citizens did not trust the organisation that handled their information. Lack of trust toward data guardians would still lead to a backlash, regardless of whether adequate privacy safeguards were in place. Thus, building trust between citizens and institutions that held sensitive information would be necessary.

In response to the privacy breaches that the country had faced and the possibility that medical data could now be opened to the public, an NGO called Big Brother Watch was formed by concerned citizens. The members of this group were largely against what they saw as further breaches of their privacy. Below is a transcript of a meeting a colleague had with them to hear their concerns.

### Snippets of Transcript of Meeting with Big Brother Watch<sup>57</sup>

| S/N | Description   |
|-----|---|
| 1   | Sabrina, 55, is a transgender Singaporean. She has been going to a private doctor regularly for any medical check-ups. Although her doctor has known about her sex change for a long time, he has kept it strictly confidential. Jenny fears that this might change as the Ministry of Health looks to set-up the centralised Electronic Health Records (EHR) database. She fears that her privacy will be breached, either maliciously or accidentally, and the fact that she is |

<sup>54</sup> Elena Chong, “Plastic surgeon fined \$13,000 for computer misuse”, *The Straits Times*, February 15, 2017, <http://www.straitstimes.com/singapore/courts-crime/plastic-surgeon-fined-13000-for-computer-misuse>

<sup>55</sup> Wong Tien Hua, “SMA's Letter to MOH on patient privacy and confidentiality under NEHR”, Singapore Medical Association. March 2, 2017. <https://www.sma.org.sg/UploadedImg/files/Publications%20-%20SMA%20News/4906/CN%203.pdf>

<sup>56</sup> Daniel Liu,. “With Uber, PDPC Must Show Force, or Risk Losing Credibility”, *ctrlfail*, December 18, 2017. <https://ctrlfail.com/2017/12/18/uber-pdpc-must-show-force-risk-loosing-credibility/>; Gabriel Olano, “Uber data breach strengthens calls for tougher privacy rules in Singapore”, *Insurance Business Mag*. Dec 1, 2017. <https://www.insurancebusinessmag.com/asia/news/breaking-news/uber-data-breach-strengthens-calls-for-tougher-privacy-rules-in-singapore-86426.aspx>

<sup>57</sup> Fictional conversation.

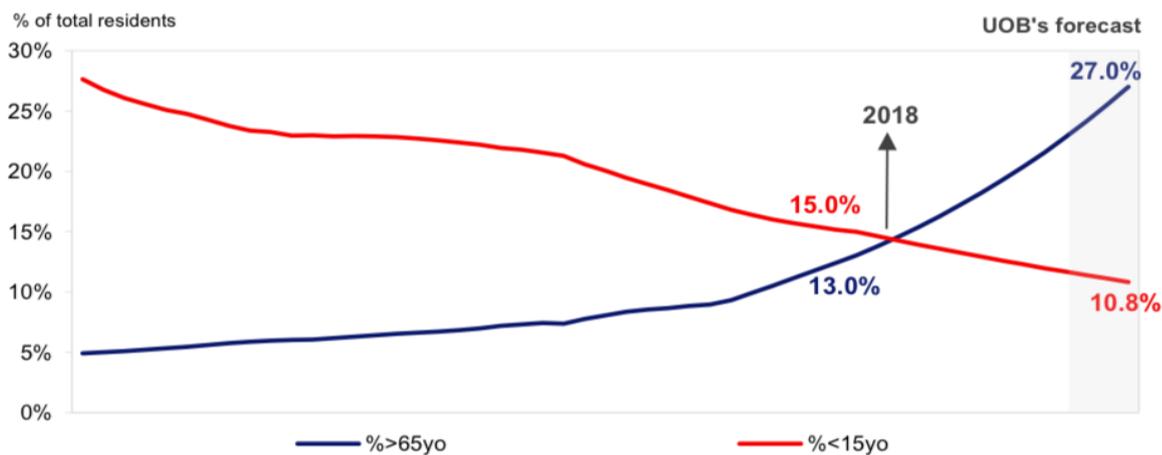
|   |  |
|---|--|
|   | transgender will be found out by her friends, colleagues or future potential employers since it is still not widely socially acceptable.   |
| 2 | Regina, 24, was one of the few people whose Singpass account was breached in 2014. Ever since, she has been living with a deep paranoia as she fears that the people who stole her data might be targeting her since they know all about her from where she lives to where she works. Although security has been tightened with Singpass, the damage to her privacy had already been done. She wants to prevent further centralisation of personal data in large repositories as it creates very lucrative targets for malicious hackers.  |
| 3 | Alex, 32, is a doctor at a local private hospital. He regularly sees patients who turn to the private hospitals specifically because they want their medical conditions to be kept strictly confidential. Recently, due to the announcement that private doctors may have to submit medical records to the central database, his patients have expressed worries that their information will no longer be kept private. As a doctor, the well-being of his patients is his utmost priority and feels duty-bound to keep them out of such stressing predicaments. Thus, he is advocating for medical records to be kept out of centralised databases and out of researcher’s hands. |
| 4 | Clara, 28, is an activist against racial discrimination in Singapore. She is concerned with what she sees as the potential use of Big Data to enact group-level harms on minority groups in Singapore. She fears that big data research might be used to justify discrimination against minorities, especially in the case of employment and insurance.  |

### Policy Options

The policy options regarding the use of big data in biomedical research are tricky and ranges from the extremely cautious to the unbridled use and access of personal information. It could be recommended that all personal data should remain uncollected or kept strictly under lock and key with strong punishments levied on those who contravene such rules. However, such a move would prevent Singapore’s economy and healthcare sector from benefitting from big data. On the other hand, being too liberal with the use and sharing of individual data can lead to public backlash. Given the potential of big data to boost the biomedical research economy and the sensitivities surrounding privacy, what should your recommendations to the Prime Minister’s Office be?

Annex

**Exhibit 1 Percentage of Singapore’s Resident Population**



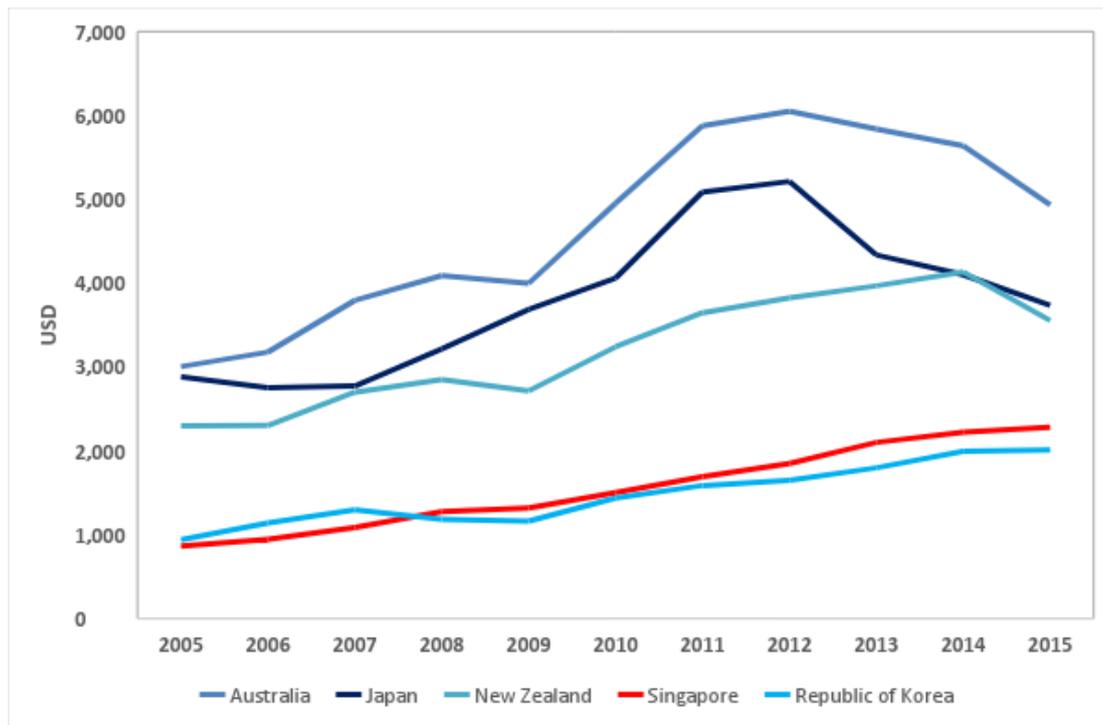
Source: Singapore Department of Statistics. Taken from UOB, “Singapore: Reaching A Critical Demographic Crossroad in 2018”, p2.

**Exhibit 2 Percentage of Population above 65 years old and below 15 years old Across Selected Countries**



Source: United Nations Population Division. Taken from UOB, “Singapore: Reaching A Critical Demographic Crossroad in 2018”, p3.

**Exhibit 3 Current Health Expenditure (CHE) per Capita Across Selected Countries**



Source: World Health Organisation, Global Health Expenditure Database.

**Exhibit 4 Diabetes in Singapore**

**DIABETES IN SINGAPORE**

**Singaporeans with diabetes**

Total  
**1 in 9**



**Breakdown by race (for those over 60)**

Over 60  
**3 in 10**



Chinese  
**2.5 in 10**



Malays  
**5 in 10**



Indians  
**6 in 10**



Infographic by Rafa Estrada



Source: Baker, Jalelah, "National Day Rally: 1 in 9 Singaporeans has diabetes; problem 'very serious', says PM Lee", *Channel NewsAsia*, August 20, 2017, <https://www.channelnewsasia.com/news/singapore/national-day-rally-1-in-9-singaporeans-has-diabetes-problem-very-9140176>

## Bibliography

- Agency for Science, Technology and Research Singapore (A-STAR). "National Survey of Research and Development in Singapore 2015." December, 2016. <https://www.a-star.edu.sg/Portals/81/Data/News%20And%20Events/Publications/National%20Survey%20of%20R&D/Files/RnD%202015.pdf>
- Ballantyne, Angela and Schaefer, Owen. "Consent and the ethical duty to participate in health data research." *Journal of Medical Ethics*, (2018): doi: 10.1136/medethics-2017-104550
- Bhandari, Manik. "Is data Singapore's next big bet?" *The Straits Times*. February 9, 2017. <http://www.straitstimes.com/opinion/is-data-spores-next-big-bet>
- Carter, Pam, Graeme T. Laurie, and Mary Dixon-Woods. "The Social Licence for Research: Why Care.Data Ran into Trouble." *Journal of Medical Ethics* 41, no. 5 (2015): 404.
- Chan, Luo Er. "National electronic patient database soon to be mandatory for healthcare providers." *Channel News Asia*. November 8, 2011. <https://www.channelnewsasia.com/news/singapore/national-electronic-health-record-mandatory-clinics-9385666>
- Chang, Ai-Lien. "Singapore commits record \$19b to R&D." *The Straits Times*. January 9, 2016. <http://www.straitstimes.com/singapore/singapore-commits-record-19b-to-rd>
- Chong, Elena. "Plastic surgeon fined \$13,000 for computer misuse." *The Straits Times*. February, 15, 2017. <http://www.straitstimes.com/singapore/courts-crime/plastic-surgeon-fined-13000-for-computer-misuse>
- Currie, Janet. "Big Data" Versus "Big Brother": On the appropriate use of large-scale data collections in pediatrics." *Pediatrics* 131, Supplement (2013): 127-132. doi:10.1542/peds.2013-0252c.
- "Data Protection and You", Personal Data Protection Commission. Retrieved from: <https://www.pdpc.gov.sg/individuals/Overview>
- Dilsizian, Steven E. and Siegel, Eliot L. "Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment". *Curr Cardiol Rep* 16, no. 441(2014): 1–8
- Docherty, A. "Big data: Ethical perspectives." *Anaesthesia* 69, 4(2014): 390-391. doi:10.1111/anae. 12656.
- Economic Development Board Singapore, "Pharmaceuticals and Biotechnology." 2018. <https://www.edb.gov.sg/en/our-industries/pharmaceuticals-and-biotechnology.html>
- Gates, Bill. "Bill Gates: We must share data to fight Alzheimer's". *Financial Times*. December 14, 2017. <https://www.ft.com/content/dc27da0a-e020-11e7-a0d4-0944c5f49e46>
- Gymrek, M. et al. "Identifying Personal Genomes by Surname Inference." *Science*. (2013). 339, pp. 321-324. <http://science.sciencemag.org/content/339/6117/321>
- Hay, Simon I., Dylan B. George, Catherine L. Moyes, and John S. Brownstein. "Big data opportunities for global infectious disease surveillance." *PLoS Medicine* 10, no. 4 (2013). *Academic OneFile* (accessed January 20, 2018). <http://link.galegroup.com.libproxy1.nus.edu.sg/apps/doc/A330252110/AONE?u=nuslib&sid=AONE&xid=4b504786>.
- Hoffman, Sharona. "The Promise and Perils of Open Medical Data." *The Hastings Centre Report* 4, 1 (2016):6-7.
- Kayyali, Basel, Knott, David and Van Kuiken, Steven. "The big-data revolution in US health care: Accelerating value and innovation." *McKinsey & Company*. April, 2013. <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care>
- Knapton, Sarah. "How the NHS got it so wrong with care.data." *The Telegraph*. July,7 2016. <http://www.telegraph.co.uk/science/2016/07/07/how-the-nhs-got-it-so-wrong-with-caredata/>

- Larrobino, Mike. "3 Ways Big Data Affects Biomedical Research." *Copyright Clearance Centre*. March 21, 2017.  
<http://www.copyright.com/blog/3-ways-big-data-affects-biomedical-research/>
- Liu, Daniel. "With Uber, PDPC Must Show Force, or Risk Losing Credibility." *ctrlfail*. December 18, 2017.  
<https://ctrlfail.com/2017/12/18/uber-pdpc-must-show-force-risk-loosing-credibility/>
- Luo, Jake, Wu, Min, Gopukumar, Deepika and Zhao, Yiqing. "Big Data Application in Biomedical Research and Health Care: A Literature Review", *Biomedical Informatics Insights*, no.8 (2016): 2, doi: 10.4137/Bii.s31559
- Ministry of Health. "Achievements of the Biomedical Sciences Initiative, Annex D."  
[https://www.moh.gov.sg/content/dam/moh\\_web/PressRoom/Articles/2012/Annex%20D%20-%20Achievements%20of%20the%20Biomedical%20Sciences%20\(BMS\)%20Initiative.pdf](https://www.moh.gov.sg/content/dam/moh_web/PressRoom/Articles/2012/Annex%20D%20-%20Achievements%20of%20the%20Biomedical%20Sciences%20(BMS)%20Initiative.pdf)
- Mittelstadt, Brent Daniel and Luciano Floridi. "The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts." *Science and Engineering Ethics* 22, no. 2 (2016;2015): 303-341.
- Narayanan, Arvid, and Edward Felten. "No silver bullet: De-identification still doesn't work." (2014).  
<http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf>
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Department of Health, Education and Welfare (DHEW), September 30, 1978. *The Belmont Report*. Washington, DC: United States Government Printing Office.
- Ng, Jing Yng. "1,560 SingPass user accounts breached." *Todayonline*. June, 5, 2014.  
<http://www.todayonline.com/singapore/1560-singpass-user-accounts-breached>
- Olano, Gabriel. "Uber data breach strengthens calls for tougher privacy rules in Singapore." *Insurance Business Asia*. December 1, 2017. <https://www.insurancebusinessmag.com/asia/news/breaking-news/uber-data-breach-strengthens-calls-for-tougher-privacy-rules-in-singapore-86426.aspx>
- Ohm, Paul. "Broken promises of privacy: Responding to the surprising failure of anonymization." *UCLA Law Review* (2010) 57: 1701–1777.
- Poldrack, R. A. & Gorgolewski K. J. "Making big data open: data sharing in neuroimaging." *Nature Neuroscience* (2014): 17, 1510–1517
- Quacquarelli Symonds World University Rankings 2017. <https://www.topuniversities.com/university-rankings/world-university-rankings/2016>
- Quinn, Ben. "Google given access to healthcare data of up to 1.6 million patients." *The Guardian*. May, 4, 2016. <https://www.theguardian.com/technology/2016/may/04/google-deepmind-access-healthcare-data-patients>
- Saeb, Sohrab, Zhang, Mi, Karr, Christopher J, Schueller, Stephen M, Corden, Marya E, Kording, Konrad P and Mohr, David C. "Mobile Phone Sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study." *J Med Internet* 17, no. 7(2015): e175. doi:10.2196/jmir.4273.
- Shortliffe, EH and Cimino, JJ. *Biomedical Informatics*. (Berlin: Springer; 2014).
- Singh, Himanshu. "Bioinformatics: Benefits to Mankind". *International Journal of PharmTech Research* 9, no.4 (2016): 243.
- SingStat. Taken from Agency for Science, Technology and Research. "Research and Development Expenditure by Area of Research, Annual." last updated January 2, 2018.  
<http://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=14600>
- Sterckx, Sigrid, Vojin Rakic, Julian Cockbain, and Pascal Borry. "You hoped we would sleep walk into accepting the collection of our data: controversies surrounding the UK care.data scheme and their wider relevance for biomedical research." *Medicine, Health Care and Philosophy* 19, no. 2 (2016):177-190.
- Summers, Judith. "Soho -- A History of London's Most Colourful Neighborhood, Bloomsbury, London." (1989): pp. 113-117. <http://www.ph.ucla.edu/epi/snow/broadstreetpump.html>

- TAFEP. "Why is there no legislation on workplace discrimination in Singapore?" <https://www.tafep.sg/3-why-there-no-legislation-workplace-discrimination-singapore>
- Tai, Janice. "Elderly health costs to rise tenfold by 2030: Report." *The Straits Times*, August 25, 2016. <http://www.straitstimes.com/singapore/health/elderly-health-costs-to-rise-tenfold-by-2030-report>
- Tan, Francis. "Singapore: Reaching A Critical Demographic Crossroad.", *UOB Global Economics & Market Research*. December 6, 2017. [http://www.uobgroup.com/assets/pdfs/research/MN\\_171206.pdf](http://www.uobgroup.com/assets/pdfs/research/MN_171206.pdf)
- Tan, Rachel and Sim, Andy. "Singapore Healthcare." *DBS Asian Insights*. July 4, 2016. [https://www.dbsinsights.com/wp-content/uploads/2016/07/160704\\_insights\\_growing\\_old\\_with\\_you.pdf](https://www.dbsinsights.com/wp-content/uploads/2016/07/160704_insights_growing_old_with_you.pdf)
- Wong, Tien Hua. "SMA's Letter to MOH on patient privacy and confidentiality under NEHR", *Singapore Medical Association*. March, 2, 2017. <https://www.sma.org.sg/UploadedImg/files/Publications%20-%20SMA%20News/4906/CN%203.pdf>
- World Health Organisation, *Global Health Expenditure Database*, 2018.
- Young, Sean D., Caitlin Rivers, and Bryan Lewis. "Methods of using Real-Time Social Media Technologies for Detection and Remote Monitoring of HIV Outcomes." *Preventive Medicine*, 63, (2014): 112-115.
- Zimmer, Michael. "More on the "Anonymity" of the Facebook dataset-It's Harvard College." October 3, 2008. <http://www.michaelzimmer.org/2008/10/03/more-on-the-anonymity-of-the-facebook-dataset-its-harvard-college/>