

“Who is watching the watchdog?”: ethical perspectives of sharing health-related data for precision medicine in Singapore

Tamra Lysaght (✉ tlysaght@nus.edu.sg)

National University of Singapore <https://orcid.org/0000-0002-7125-4206>

Angela Ballantyne

National University of Singapore

Vicki Xafis

National University of Singapore

Serene Ong

National University of Singapore

Gerald Owen Schaefer

National University of Singapore

Jeffrey Ling

National University of Singapore

Ainsley J Newson

The University of Sydney

Ing Wei Khor

National University of Singapore

E Shyong Tai

National University of Singapore

Research article

Keywords: data security, health data, precision medicine, governance

Posted Date: October 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-24953/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on November 19th, 2020. See the published version at <https://doi.org/10.1186/s12910-020-00561-8>.

Abstract

Background We aimed to examine the ethical concerns Singaporeans have about sharing health-data for precision medicine (PM) and identify suggestions for governance strategies. Just as Asian genomes are under-represented in PM, the views of Asian populations about the risks and benefits of data sharing are under-represented in prior attitudinal research.

Methods We conducted seven focus groups with 62 participants in Singapore from May to July 2019. They were conducted in three languages (English, Mandarin and Malay) and analysed with qualitative content and thematic analysis.

Results Four key themes emerged: nuanced understandings of data security and data sensitivity; trade-offs between data protection and research benefits; trust (and distrust) in the public and private sectors; and governance and control options. Participants were aware of the inherent risks associated with data sharing for research. Participants expressed conditional support for data sharing, including genomic sequence data and information contained within electronic medical records. This support included sharing data with researchers from universities and healthcare institutions, both in Singapore and overseas. Support was conditional on the perceived social value of the research and appropriate de-identification and data security processes. Participants suggested that a data sharing oversight body would help strengthen public trust and comfort in data research for PM in Singapore.

Conclusion Maintenance of public trust in data security systems and governance regimes can enhance participation in PM and data sharing for research. Contrary to themes in much prior research, participants demonstrated a sophisticated understanding of the inherent risks of data sharing, analysed trade-offs between risks and potential benefits of PM, and often adopted an international perspective.

Introduction

Precision medicine (PM) aims to provide more personalised care to patients, taking account of genomic, environmental and behavioral information.¹ In general, medical knowledge has been largely built on physiological, epidemiological and clinical trial data, reflecting the health status and treatment response of entire communities. The overarching aim of PM is to use genomic analyses and data analytics to build a finer grained and tailored approach to predicting disease progression and treatment response for individual patients.^{2 3} For this paradigm to work, researchers would require access to comprehensive clinical data and biological material from many people to be able to identify patterns and stratify small groups of patients. Key examples of PM initiatives include the United States the *All of US* which seeks to recruit one million volunteers to share genomic samples and data for research;⁴ and the China Precision Medicine Initiative with an estimated investment of US\$9.2 billion over the next 15 years (2016–2030).⁵

PM initiatives typically seek broad consent when participants enrol and do not seek specific consent for every time the data or samples are shared with a new research team. Health data sharing for research

under this consent model should operate within the social licence and in line with public expectations. The concept of a social licence refers to a privilege to operate, implicitly granted by society to an organisation or profession, often in the absence of explicit consent.⁶ It refers to a practice that broader publics are willing to accept as morally and socially permissible. Identifying the boundaries of the social licence is difficult; and this is often most evident when activities overstep this boundary and there is a subsequent public backlash.⁷ A breach of public trust and expectations can have wide-ranging negative consequences for a PM initiative. Such concerns materialised in the wide-spread rejection of the proposed UK Care.data project 2014.^{8,9} It is thus important to characterise the boundaries and nature of the social licence for health data sharing for research in advance.

In this study, we examine public expectations and concerns about health data sharing for PM in Singapore. This small city-state of 5.7 million inhabitants has a high gross domestic product, a robust and comprehensive health system, and strong investment in biomedical research.^{10 11} Singapore has invested significantly in developing PM with the large-scale whole-genome sequencing of Singapore's multi-ethnic Asian population (SG10k).¹² Some of this data has been linked to patients' electronic medical records (EMR) and used to support clinical care under the SingHealth Duke-NUS Institute of Precision Medicine (PRISM) initiative.¹³ Singapore has the capacity to upscale this initiative across the public healthcare system, which is served by a national IT provider that can integrate health information from multiple sources, including EMR, biobanks and administrative databases.¹⁴ This combination of comprehensive health information, and expertise in genomics research and clinical implementation,¹⁵ puts Singapore in a strong position to initiate a nation-wide PM programmes and contribute to similar initiatives internationally.

Despite these strengths, however, several major high-profile data breaches in Singapore have the potential to undermine efforts to establish an effective PM programme. These included unauthorised access to 1.6 million patient records held at SingHealth (one of the three major health clusters) and a leak involving the names of over 14,000 individuals named on the HIV positive registry.^{16,17} These leaks involved identifiable clinical data; even though they were not directly related to precision medicine, or even research, they may have increased public anxiety around health data sharing. Maintaining the social licence for PM thus requires responding to community concerns about data security and ethical issues. This study aimed to examine those concerns and seek views on potential governance strategies for managing them.

Background

There is a growing body of international empirical literature exploring the views of patients and various publics about sharing clinical data and biological samples for health research. The emerging consensus shows broad conditional support for sharing health data for research. Four common themes are reported in this literature: (1) the necessity of appropriate governance structures; (2) that research aims to produce public benefit; (3) the reluctance to allow commercialisation of public data sets, and (4) concerns about

privacy and data security. A recent meta-analysis of this literature found the following recurrent issues: social value of the research, privacy, risk-minimisation, data security, control, transparency and engagement, trust, responsibility and accountability.¹⁸

Prior empirical research in Singapore indicates patient preferences for broad consent models to manage genome data sharing for research and the return of cancer profiling results (provided this is accompanied with adequate explanation).¹⁹ Broad consent involves consent for an unspecified range of future research projects, subject to limited content and/or process restrictions, collected at the point of sample or data donation; it is popular for bio-banking globally.²⁰ It contrasts with specific consent models where consent is sought from donors each time data are accessed for specific research projects. While specific consent provides data donors with greater individual-level control over who gets access to the data, broad consent models are more common because they reduce the costs of research and burdens on donors being repeatedly re-contacted for consent.²¹

There are three key limitations of the current empirical literature about attitudes to data sharing. First, the findings quickly become outdated because the technology and data ecosystem are changing so rapidly. As technology develops and the public gains more experience with the benefits and risks associated with data sharing, public understanding, expectations and concerns about data sharing are re-calibrated. For example, models of data de-identification in research continue to evolve. As more personal data becomes available in the public sphere there is greater likelihood of de-identified data being re-identified so de-identification methods and data regulation have to adjust. De-identification is a process of detecting and removing identifiers (e.g., personal names and national identification numbers) that directly or indirectly point to a person. True anonymization of datasets is becoming technically unachievable; and in several respects is undesirable for PM which ultimately seeks to integrate genomic research and clinical care.²²
²³ Therefore effective de-identification is an essential tool for protecting the privacy of participants in PM research.²⁴

Second, much of the existing literature investigates patient and public *preferences* about data sharing and patient control of data access and use.^{25 26 27 28 29 30} This focus on preferences is arguably misdirected. Taylor and Taylor found that people under some circumstances were still willing to accept “non-preferred” data sharing models.³¹ Therefore, rather than asking about preferences, it may be more relevant to ask: what do people expect from participating in data driven medical research and what costs are they willing to bear to gain the purported benefits from data-driven medical research? This question is relevant because there is an inherent trade-off between competing values in managing access to data for research: most obviously between protecting privacy and generating public health benefits from data-driven research.³² Research that simply asks participants what they want or prefer may generate an idealised but impractical model that does not sufficiently account for these trade-offs. Few prior studies explicitly explore which data sharing practices people are *willing* to accept, all things considered, as opposed to what they prefer.^{34 34 35}

Furthermore, the views of Asian populations around these issues are under-represented in the literature. Just as European genomes are over-represented in genomics,³⁶ the majority of qualitative and quantitative studies about public attitudes and perspectives of genomics have been conducted in Europe, North America and Australia. Low rates of Asian minority participation in biobanks have been attributed to cultural views that blood is sacred³⁷ and fear of discrimination.³⁸ According to the Global Alliance for Genomics and Health, governance frameworks for the responsible sharing of genomic and health-related data should be culturally appropriate and sensitive to local norms, practices and perspectives.³⁹ It cannot be assumed that the perspectives people have about data sharing in Asian countries are the same as those reported elsewhere.⁴⁰ Therefore, we aimed to examine those perspectives in this qualitative study of the ethical concerns that Singaporeans have about sharing health-related data and what governance strategies could be put in place to address them.

Methods

A commercial market research firm was contracted to recruit participants, conduct the focus groups, and transcribe the discussions for analysis. The firm recruited participants with consecutive sampling from several online platforms (e.g. Facebook & Gumtree) using advertisements and an anonymous survey to screen prospective participants for their eligibility, availability and demographic profile (i.e. age, sex, education, ethnicity). Only Singaporean citizens or permanent residents over the age of 21 years were eligible to participate. This restriction was to reflect the recruitment parameters for precision medicine initiatives such as SG10K¹² in Singapore. Participants gave their written informed consent and received a supermarket voucher of SGD 50.00 for participation.

The study team developed the discussion guide in consultation with the research firm. The guide covered three topics: (1) concerns about the storage and sharing of data for PM; (2) potential governance strategies to mitigate or alleviate concerns; and (3) the kinds of benefits participants would expect from PM. At the commencement of each focus group (FG), participants were shown videos explaining the concepts of genes and precision medicine, and describing that data would be stored and shared only in a de-identified format but would be encoded to allow re-identification of an individual in the future, if needed (e.g. to withdraw them from a database or contact them with health-relevant findings). It also discussed the potential security risks from an unauthorised data breach and showed two cases that describe the potential uses and benefits of PM for treating individual patients. The videos shown to participants are available online.⁴¹

Focus groups were conducted at the premises of the market research firm in three languages (English, Mandarin, and Malay) and were observed by at least one member of the project team. The duration of each focus group was between 90-130mins. The facilitator avoided using the term *anonymous* and emphasised that data for PM would be stored and shared in a de-identified format. The firm transcribed the discussion from audio recordings and translated the non-English discussions into English (checked with reverse translation). They also removed any personal identifiers from the transcripts before

sharing them with the study team for analysis. Transcripts were analysed in NVivo 12 (QSR International) using qualitative thematic analysis.⁴² The coding was primarily shared between two senior researchers (TL and VX) with assistance from two graduate students (SO and JL) until thematic saturation was reached (where no new themes relevant to the research aims were emerging). Content was coded inductively for topics, which were progressively collapsed into broader thematic categories, and refined over several meetings between the coders and study team.

Results

Seven focus groups were conducted from May to July 2019: five in English, one in Mandarin, and another in Malay. Each focus group lasted between 90 and 120 minutes and had six to eleven people, totalling 62 participants for the study (see Figure 1 for demographic description of participants). The median age range of 38-45 years. The majority were ethnic Chinese (49), with Malay (7) being the next largest group, and Indian (5) and others (2) making up the rest. The majority of participants were female (63%) and most had either a diploma (37.1%) or university degree (48.4%).

Despite efforts made to make clear the distinction between personal and de-identified data, and to avoid the term *anonymous*, participants regularly reverted to discussing personally identifiable data and anonymity. For example, immediately following a discussion where participants explicitly acknowledged that data would be de-identified, another expressed their concern that “It can be misused. It’s my personal information so I would like to guard it” (FG0521_P1). Although initially out of scope, some comments relating to the sharing of personal information are included in the results as they reflect the participants’ concerns about data sharing, and it was impossible to completely disentangle conversations about personal and de-identified data.

The following results are organised around four key themes: nuanced understandings of data security and data sensitivity; trades-offs between data protection and research benefits; trust (and distrust) in the public and private sectors; and governance and control options.

Nuanced understandings of data security and sensitivity

While participants were generally comfortable with the idea of storing and sharing de-identified data, concerns were raised about data security. The primary concern was about the potential misuse of data (e.g. for malevolent purposes, such as bioterrorism, discrimination or profiteering) and the resulting harm that might arise from a data breach. However, some participants appeared to have a good understanding of the various security measures that could be put in place to protect the identity of volunteers in a PM program and to reduce the likelihood of an unauthorised breach. Some drew on their experiences and knowledge of previous data breaches with references to SingHealth and the HIV registry cases arising frequently and without prompting. These events gave participants a reference point for exploring the issues of data sensitivity and security. They also recognised that no data security systems were

impervious to attacks and no data holders, including trusted government agencies, can guarantee data security.

Participants were generally more worried about the security of what they perceived to be sensitive data. People suffering stigmatised conditions in Singapore, such as infectious disease like HIV or mental illness, were viewed as being more vulnerable to an unauthorised disclosure than someone who is in good health or is suffering conditions such as cancer.

If I have like generic health condition, perhaps like high cholesterol or heart issues, I wouldn't be so worried about my information being shared. But if [...] I have HIV or things like that, that are more sensitive, then I would be concerned. But if its generic health situations, I, for me, personally, I won't be that concerned. (FG0530_P4)

However, opinions varied on the sensitivity of genetic health information. Some participants were relatively unconcerned about sharing genetic data. Others were worried about insurers, employers or school selectors using (personally identifiable) genetic data against the interests of the data donor:

[If you] happen to have some genetic diseases that could result in you having a higher risk of a certain disease and what happens if this information is sold to the insurance companies and then you will be anti-selection, you won't be covered for a certain disease or when your coverage will be very high. Then that would be a concern. (FG0530_P7)

Participants also recognised that some people might be vulnerable to unauthorised data breaches due to their environment or circumstances, rather than the sensitivity of the data itself. Public figures (e.g. celebrities and politicians) were viewed as greater targets for 'hackers' than ordinary citizens.

Trades-offs between data protection and research benefits

There was broad support for sharing data with researchers at academic and healthcare institutions (both local and international) as they were perceived as conducting socially valuable research. Participants recognised and articulated the implicit trade-offs between protecting data from a breach, on the one hand, and promoting access to the data to generate potential benefits on the other hand.

We have to think about the benefits that you can reap from this programme, having the genomic data and sharing it through the national body as well as at an international level. That is basically for the goodness of humankind in the world. So the issue of transfer of information, stolen data, hackers, either for political use or used by pharmaceutical companies. They're bad, unfortunately [...] that is a general concern. But I feel that the benefits, weighing the pros and cons, it is better to have this data. (FG0521_P2)

The purpose of data sharing was an important factor in the discussions. The introductory video provided examples of possible health benefits to individuals and their families. However, participants also went on to highlight a wider range of societal benefits for what was frequently termed, unprompted, the "greater good" (e.g. FG0527_P3&P6, FG0524_P2&P3, FG0530_P6&P2). These benefits included the development

of new medicines, improved or more affordable healthcare, and improved understanding of genetic diseases (including hereditary cancer). Participants recognised the genomic research in Singapore could be especially valuable to minority ethnic groups in Singapore and Asia who may not be well represented in other PM initiatives.

Participants were concerned about parties with *authorised* access using the data for purposes contrary to the intentions or expectations of the data donors. Hence, it was important for them to understand the reasons why data would be shared with anyone inside or outside of the institution holding the data:

I guess, for me I think it's not so much the party accessing the data, but rather how is the data being used for and for what purpose. So knowing that, then I'm able to make a better decision as in whether I want to participate. ...if it's from a big pharmacy company, then I think it may be for a commercial gain, but again it still help people. So I guess it's still the purpose, how the data being used, the purpose what is it used for. (FG0524_P11)

As indicated in the quote above, many participants expected that they would be notified prior and/or asked for consent for data-sharing. Some mentioned Singapore's personal data protection laws (i.e. the PDPA) and expressed in strong terms that their consent should be sought before data were shared for any purpose. However, many others were willing to trust the institution storing the data to only allow access to trustworthy third parties for worthwhile purposes of generating benefit, and would not need to provide new consent to each research project.

So, if I [...] give you all my data, ok, so actually I trust you enough. (FG0530_P10) Trust (and distrust) in the public and private sectors

When asked about who they would trust to hold the data, participants suggested academic research institutions, and healthcare providers. Government agencies came through as trusted entities for ensuring data security. Given the health-relatedness of the data being stored, the Ministry of Health (MOH) specifically was frequently mentioned. This was despite some recognition that MOH had jurisdictional responsibility for data accessed in previous breaches. In addition to MOH, government defence and security agencies were also mentioned along with statutory boards and public-private partnerships that regulate digital and information technologies in Singapore (e.g. iHIS⁴³ and GovTech⁴⁴). Trust was not absolute, but rather participants expressed a nuanced approach, differentiating between different government agencies and difference functions (e.g. data security, use and oversight).

On the other hand, there was significant debate amongst participants about sharing data with researchers at private or commercial entities, such as pharmaceutical companies. Some participants highlighted the fact that the pharmaceutical industry contributes to social value by developing new medicine; others were more ambivalent about the high costs of some medications in Singapore. Furthermore, participants were uncertain as to how the government could ensure that individuals or the public healthcare system would benefit from data-sharing with industry.

Commercial companies, ultimately they [...] make money, so [...] they [may] take your information, develop a certain diagnostic kit and then sell it off to other companies without benefitting the whole of Singaporeans and we are actually providing them with the data to come up with the kit, so I think [...] they must bring some of the benefits back to the Singapore community. (FG0530_P7)

Overall, there was strong reluctance to share health data with private insurance companies. There was much discussion about the potential for insurance companies to use data to deny coverage to certain ethnic groups and/or increase the cost of insurance plans. By comparison, some participants supported sharing health data with insurance companies, arguing that they could use the data to benefit patients by tailoring plans for specific diseases or expanding coverage for patients not currently covered. Others did not see the need to share the data with insurance companies at all, noting that they could instead work with public research institutions as trustworthy actors to develop equitable insurance plans based on aggregated data.

Governance and control options

Participants were asked to suggest strategies to ensure responsible and trustworthy approaches to storing and sharing health data. Participants made numerous suggestions for data security measures that would help to alleviate their concerns, such as block chain technology, data masking, two-step verification, and tiered access restrictions. The need for a strong regulatory environment was acknowledged. Some participants also felt that data should not be accessed without the specific consent of volunteers, although it is unclear if they were talking about personally identifiable data or de-identified data. Being able to withdraw from the PM initiative appeared to be important for many participants; although exactly what that meant (e.g. severing all links to data, deleting all records from the database, declining participation in a particular study) was not explored.

In addition, some participants suggested that independent oversight of data sharing in Singapore across sectors and agencies would be an important component of a trustworthy governance system. Several options were proposed to fulfil this oversight function, including non-government organisations or an intergovernmental agency within Singapore. Some suggested cross-sectorial “semi-government, private sector” (FG0528_P5) collaborations. International oversight bodies were also suggested in the context of government agencies gaining access to data:

Who is watching the watchdog? [...] there has to be checks and balances from all different agencies... Or even, not even the agency, I think even bigger body. A bigger international body. ... Like a United Nations or some nations right? Or Interpol, for example, I do not know. Beyond just our country. (FG0527_P8)

Thus, although there was a general sense that government agencies could be trusted to regulate the research institutions and private entities that access and use data within Singapore, there was also some unease with how the government itself might use the data in the absence of external oversight or checks and balances. The reasons for the apparent distrust were not probed and preferences for particular oversight mechanisms were not explored in any further depth.

Discussion

Key findings show that participants in this study were generally comfortable with the idea of storing and sharing de-identified genomic and medical data (e.g. from EMRs) with researchers from universities and healthcare institutions, both in Singapore and overseas. However, this broad support was conditional on the likelihood of producing a tangible health benefit for patients and the adoption of acceptable de-identification and data security processes. Participants also raised a range of concerns about data sharing, the conditions under which they would accept it taking place and the trade-offs they made between risks and potential benefits.

Participants in this study were mostly tertiary educated. Thus, it is unsurprising that they demonstrated nuanced and sophisticated understandings of the ethical issues surrounding data security and sharing for PM. Participants took a pragmatic approach to data sharing and understood degrees of data sensitivity and limitations of security measures. They also appreciated that this was related to, but different from, data identifiability. That said, there remained persistent slippage between discussion of personal identifiable data and de-identified data, *even amongst this highly educated group* of people. The difficulty in holding personal data separate from de-identified data may have been because the distinction was not clear in the mind of participants and the short videos shown to participants may have been insufficient to facilitate deeper understandings. However, it also reflects ongoing academic, regulatory and technical debate about the conceptual boundaries of personal and non-personal data, as well as methods of identification, de-identification and re-identification, and the persistent confusion between these terms.^{45 46}

Participants did not expect a guarantee of anonymity or data security and they were willing to accept some risks in order to gain the potential benefits of PM. Participants were most motivated to accept these inherent risks for research aimed at producing public benefit. Finally, participants offered rich and diverse suggestions to alleviate concerns about data sharing. Participants preferred a multi-tiered approach including control options for individuals (e.g. the ability to withdraw from the PM initiative), regulatory limits on users and uses of the data, as well as independent oversight. Numerous suggestions were made about the types of organisations that could be trusted in an oversight role with the government being cited most prominently.

The results of our research are novel in four respects. First, compared to prior research⁴⁷ our results demonstrated a high level of awareness of the inherent risks associated with data collection, the impossibility of ensuring data security to avert **all** risks, and challenges of de-identifying data. Previous research has demonstrated broad support for data sharing for health research¹⁹ so long as data is de-identified^{48 49} and privacy is protected;⁵⁰ both of which are increasingly difficult conditions to guarantee. Participants in this study also expressed concerns about data security and identifiability, but demonstrated a nuanced and sophisticated understanding of data security. Participants recognised that all security systems were vulnerable to an attack and differentiated between a data breach that

amounted to a technical privacy invasion versus a data breach resulting in harm. Importantly, support for data sharing was not conditional on a guarantee of data security.

Second, participants weighed data protection and individual patient control on the one hand (personal risk) and the benefits of sharing the data to promote health research on the other (public good). Only a small proportion of the existing empirical research captures the public's efforts to evaluate these trade-offs.^{47 51} Participants in our study engaged in these complicated trade-offs unprompted. Understanding how people make these assessments is important because they may be concerned about data security; but may be nonetheless willing to accept the inherent risks of data collection, storage and sharing if this facilitates medical progress. In our study, participants framed the debate as a question of trade-offs and opportunity costs without being prompted. This balancing of harms and benefits revealed some sophistication in participants' understanding of the ethical issues and the trade-offs that would be necessary to generate value from PM. Our results suggest that public understanding of biobanking and data sharing in technologically developed countries such as Singapore may be deepening and maturing.

Third, participants were relatively unconcerned with the prospect of international data sharing. Previous research has shown general public reticence to share data across national borders, and especially with respect to sharing genomic data⁵² or biological samples⁵³ and/or data relating to Indigenous communities.⁵⁴ Participants in this study, however, recognised the benefits of international data sharing to drive research and innovation, highlighted the value of generating knowledge that would benefit all humankind. Some even suggested international oversight as a potential governance option for PM, suggesting an international rather than exclusively domestic approach to PM. This attitude may reflect the role of Singapore as a global hub for international commerce, business and travel.

Finally, participants expressed relatively high levels of trust in government authorities to ensure data security. This perspective contrasts with other countries where confidence in governments to safeguard data security and oversee the responsible use of health-related information is less striking. Governmental and legal responses to the prior data breaches at SingHealth and the HIV registry, which included substantial investigations and rapid adoption of reform measures,⁵⁵ may have influenced this attitude. However, it also speaks to cultural norms and the socio-political relationship that Singaporeans have with the government and ministerial agencies. Singaporeans have experienced soft authoritarian governance for over 55 years where a single party has been in power since independence; providing political stability, economic prosperity, security, and access to housing, healthcare and education, in exchange for restricted liberties and personal freedoms.⁵⁶ For the most part, the citizenry has accepted this situation and trusts the pragmatic paternalism that is characteristic of Singapore. This degree of trust would, in all likelihood, pass onto a national PM initiative.

Building trustworthy governance

Solutions to some of these issues may be found in models of systemic governance. Recognising the ethical deficiencies that informed consent provides as a mechanism in preserving autonomy

and protecting privacy in data-intensive research, systemic governance refers to an oversight pipeline that accounts for the multiplicity of actors and stakeholders involved in an entire health data eco-system.⁵⁷ Vayena and Blasimme propose a normative framework of six principles for systematic oversight: adaptivity, flexibility, monitoring, responsiveness, reflexivity, and inclusiveness. This framework fits within an adaptive governance approach that allows policymakers to respond to changing technological and social conditions, and align regulation with public expectations, by including multiple actors in the design and monitoring of the oversight pipeline.⁵⁸ It encourages collaboration between regulators and the regulated as well as social learning across different levels and hierarchies of governance structures.

Securing the social license for PM initiatives to operate will require trustworthy systems of governance. As participants in one discussion asked, “who watches the watchdog?”. This question is a critical reflection upon the capacity of any single agency to protect and promote the common good.⁵⁹ As our results and the prior literature suggests,⁷ the social license for PM programmes is likely to be contingent upon assumptions that participation is voluntary and use of the data is governed by broadly accepted social and cultural values, such as reciprocity and public benefit. Independent agencies and oversight mechanisms play an important role in holding others accountable when breaches occur; and, therefore, are a significant component of a trustworthy system of governance. Examples of possible oversight roles include the National Data Guardian for Health and Social Care in England,⁶⁰ the Privacy Commissioner in New Zealand,⁶¹ and the Australian Government’s Data Governance Framework 2020.⁶²

A critical aspect of the governance system is attention to ‘who’ holds decision making power within the system. This issue is particularly relevant to PM programmes because they are long term propositions where the available technologies will change over time, along with potential usages and users; often in highly unpredictable ways. The norms and values implicit to the social license are also liable to change, or at least be re-interpreted and re-prioritised over time, and in ways that may differ from the original configuration when a programme is established. Furthermore, social injustices that expose segments of the population to vulnerabilities from data breaches and misuse can aggregate and become exacerbated in unforeseen ways. Governance approaches must be adaptive rather than static. It matters who has a role in interpreting and applying governance frameworks, in defining the public interest and appropriate uses and users of precision medicine data.⁶³

Ostrom has argued that nested governance allows oversight systems to be scaled appropriately in reaction to changing circumstances.⁶⁴ Multiple levels of governance captures the benefits of centralisation whilst still being responsive to specific social-ecological challenges. One example is the need for data governance systems to respond to growing demands for Indigenous data sovereignty.⁶⁵ For example, New Zealand has a centralised research database called the Integrated Data Infrastructure (IDI) holding comprehensive microdata about people, households and businesses and governed by Statistics New Zealand (Stats NZ).⁶⁶ In response to local concerns about control over the use of Māori data, Stats NZ has adopted an additional layer of review for research using Māori data, called Ngā Tikanga

Paihere.⁶⁷ This review process was developed by leading Māori scholars and is an example of power sharing within a data governance model.

Limitations and future research

Caution should be taken in extrapolating findings from these focus groups to the wider population in Singapore and beyond. As with any qualitative study, the results are not intended to be generalizable and may not reflect the views of all Singaporeans. Recruiting from social media platforms may only reach a segment of the population with access to the technology, and our study only included citizens and permanent residents. Approximately 30% of the population in Singapore are foreigners with many being unskilled transient workers whose perspectives and understandings may differ greatly from those included in our study. Almost half participants in our study were tertiary educated with university degrees compared with 32.4% of the population; and where 25% of the resident population have not attained secondary education.⁶⁸ Higher levels of education within our research population may explain their familiarity with data security, previous data leaks, and ability to engage in sophisticated trading-offs between the benefits and risks of data sharing for PM. The lack of conceptual clarity around identification and de/re-identification also suggests that care should be taken in drawing conclusions about what data Singaporeans will be comfortable sharing, especially under a broad consent model.

Yet these findings provide important insights into ethical concerns that are held amongst Singaporeans and the trade-offs they may be willing to accept for data sharing. They have also identified several measures that could be applied in building a culturally appropriate trustworthy governance framework for PM in Singapore. Further research is needed to elaborate on the measures that Singaporeans would consider most important in the configuration of governing norms and values, and to spell out more explicitly, where the moral boundaries of the social license will lie.

Conclusion

Maintenance of public trust in data security systems and governance regimes can enhance participation in PM and data sharing for research. Our research sought to explore the boundaries of the social licence for data sharing for PM in Singapore. We were interested specifically in what data sharing arrangements participants would be willing to accept; as opposed to models they might prefer. Despite recent data security breaches, government agencies remain preferred guardians for health data but there was recognition that they too must be subject to oversight. The purpose of data sharing emerged as a primary factor affecting willingness to share data. Participants were comfortable with data sharing with academic and healthcare institutions (in Singapore or overseas) because they trust such researchers to produce socially valuable knowledge, products and interventions. There was general hesitancy to share data with industry and insurance companies but also recognition of the important role pharmaceutical companies and insurers play in drug development and healthcare. Participants suggested that a data sharing oversight body would help strengthen public trust and comfort in data research for PM in Singapore.

Abbreviations

PM	Precision Medicine
EMR	Electronic Medical Records
PRISM	SingHealth Duke-NUS Institute of PRecISion Medicine

Declarations

Ethics approval and consent to participate

The National University of Singapore Institutional Review Board approved this study (S-19-045E). All participants gave their informed consent to participate in this study.

Consent for publication

Not applicable

Competing interests

E Shyong Tai is Director of the Centre for Precision Health at the National University Hospital System, which has received grants from the National Medical Research Council and the National Institutes of Health to conduct studies that involve analysis and sharing of de-identified genetic data. E Shyong Tai is also the co-chair of the Precision Medicine Steering Committee (appointed by the Ministry of Health) with specific responsibility for the Public and Community Trust Workgroup. Owen Schaefer is a member of that Workgroup and Ing Wei Khor is partly funded by the Ministry of Health, Singapore, to support their work. Ainsley Newson is also an Editorial Board member of the BMC Medical Ethics. No other author has any financial competing interests to declare.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available for reasons of personal privacy, but are available from the corresponding author on reasonable request.

Funding

This research is supported by the Ministry of Education, Singapore, under its Social Science Research Council Thematic Grant (MOE2017-SSRTG-028), and the Precision Medicine Coordination Office, Ministry of Health, Singapore. Neither funder played any role in the planning and development of the research, the collection or analysis of data, or reporting of the results.

Author Contributions

Tamra Lysaght (study design, data analysis, supervision of research students, manuscript drafting and submission); Angela Ballantyne (literature review, data analysis, manuscript drafting); Vicki Xafis (study design, data analysis, manuscript review and editing); Serene Ong (literature review, study design, data analysis, manuscript review and editing); Owen Schaefer (study design, manuscript review and editing); Jeffrey Ling (data analysis); Ainsley Newson (data analysis, manuscript review and editing); Ing Wei Khor (study design, data analysis, manuscript review and editing); E Shyong Tai (study design, manuscript review and editing)

Acknowledgements

We thank Professor Patrick Tan and Yasmin Bylstra from the SingHealth Duke-NUS Institute of PRECISION Medicine (PRISM) for their insights and clarifications on the genome sequencing research programmes in Singapore, and Prof Helena Legido-Quigley from Saw Swee Hock School of Public Health, National University of Singapore, for input into the research design.

References

1. Schaefer, G.O., Tai, E. & Sun, S. Precision Medicine and Big Data. *ABR* (2019) 11: 275. <https://doi.org/10.1007/s41649-019-00094-2>
2. Marquart J, Chen EY, Prasad Estimation of the Percentage of US Patients With Cancer Who Benefit From Genome-Driven Oncology. *JAMA Oncol.*2018;4(8):1093–1098. doi:10.1001/jamaoncol.2018.1660
3. Chowkwanyun M1, Bayer R1, Galea "Precision" Public Health - Between Novelty and Hype. *N Engl J Med.* 2018 Oct 11;379(15):1398-1400. doi: 10.1056/NEJMp1806634. Epub 2018 Sep 5; Aron DC.Precision medicine in an imprecise and complex world: Magic bullets, hype, and the fuzzy line between health and disease. *J Eval Clin Pract.* 2019 Dec 21. doi: 10.1111/jep.13306.
4. National Institute of Health (2020). What is the Precision Medicine Initiative? Available at: <https://ghr.nlm.nih.gov/primer/precisionmedicine/initiative> Accessed on 31 March,
5. Sanders S, Oberst J (Eds): Precision Medicine in In American Association for the Advancement of Science, Washington DC:Science/AAAS Custom Publishing Office; 2016, p. 41. https://www.sciencemag.org/sites/default/files/custom-publishing/documents/Bioyong- Precision-Medicine-supplement_Final.pdf
6. Allen J, Adams C, Flack F. The role of data custodians in establishing and maintaining social licence for health research. *Bioethics* 2019;33(4):502–10.
7. Carter P, Laurie GT, Dixon- Woods The social licence for research: why care. data ran into trouble. *J Med Ethics* 2015;41(5):404–9.
8. Temperton J, 2016. NHS care.data scheme closed after years of controversy *Wired* 6 July 2016 <https://www.wired.co.uk/article/care-data-nhs-england-closed>

9. Carter P, Laurie GT, Dixon- Woods The social licence for research: why care. data ran into trouble. *J Med Ethics* 2015;41(5):404–9.
10. Department of Statistics, Singapore (2020). Population and Population Structure. Available at: <https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data> Accessed on 31 March,
11. National Research Foundation (2019) Health and Biomedical Sciences. Available at: <https://www.nrf.gov.sg/rie2020/health-and-biomedical-sciences>. Accessed on 31 March, 2020.
12. Wu et al. (2019) "Large-Scale Whole-Genome Sequencing of Three Diverse Asian Populations in Singapore." *Cell* 179(3): 736-749.e715.
13. Bylstra, Y., S. Davila, W. K. Lim, R. Wu, J. X. Teo, S. Kam, T. Lysaght, S. Rozen, B. T. Teh, K. K. Yeo, Cook, P. Tan and S. S. Jamuar (2019). "Implementation of genomics in medical practice to deliver precision medicine for an Asian population." *npj Genomic Medicine* 4(1): 12.
14. Chin, N. C. (2015). Big Data Making A Great Difference in Healthcare. Available at: https://www.ihis.com.sg/Latest_News/News_Article/Pages/Big_Data_Making_A_Great_Differen.aspx Accessed on 31 March, 2020.
15. Editor (2018). UK Biobank data on 500,000 people paves way to precision medicine. *Nature* 562, 163-164.
16. Tham I: Personal info of 1.5m SingHealth patients, including PM Lee, stolen in Singapore's worst cyber attack. *The Straits Times*, Singapore, 20 July
17. Chang A-L: Data of 14,200 people with HIV leaked online by US fraudster who was deported from Singapore. *The Straits Times*, Singapore, 28 Jan
18. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *J Med Ethics*. 2019 Nov 12. pii: medethics-2019-105651. doi: 10.1136/medethics-2019-105651.
19. Bylstra Y, Lysaght T, Thriyikraman J, Watson S, Tan P. (2017) Ethical frameworks for obtaining informed consent in tumour profiling: an evidence-based case for Singapore. *Hum Genomics*. Dec 8;11(1):31. .
20. Grady C, Eckstein L, Berkman B, et al. (2015) Broad Consent for Research With Biological Samples: Workshop Conclusions. *Am J Bioeth*. 15(9):34–42.
21. McGuire AL1, Oliver JM, Slashinski MJ, Graves JL, Wang T, Kelly PA, Fisher W, Lau CC, Goss J, Okcu M, Treadwell-Deering D, Goldman AM, Noebels JL, Hilsenbeck SG. To share or not to share: a randomized trial of consent for data sharing in genome research. *Genet Med*. 2011 Nov;13(11):948-55. Mikkelsen RB, Gjerris M, Waldemar G, Sandøe P. Broad consent for biobanks is best - provided it is also deep. *BMC Med Ethics*. 2019 Oct 15;20(1):71.
22. Lim, H. Y. (2017). *Data Protection in the Practical Context – Strategies and Techniques* Singapore, Academy Publishing Singapore;

23. Rocher, L., Hendrickx, J.M. & de Montjoye, Y. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* 10, 3069 (2019).
<https://doi.org/10.1038/s41467-019-10933-3>
24. Kayaalp M. Modes of De-identification. *AMIA Annu Symp Proc.* 2018;2017:1044–1050. Published 2018 Apr
25. Patil S, Lu H, Saunders CL, et al. Public preferences for electronic health data storage, access, and sharing - evidence from a pan- European survey. *J Am Med Inform Assoc* 2016;23(6):1096–106.
26. Shah N, Coathup V, Teare H, et Sharing data for future research- engaging participants' views about data governance beyond the original project: a direct study. *Genet Med* 2019;21(5):1131–8.
27. Willison DJ, Keshavjee K, Nair K, Goldsmith C, Holbrook AM; Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness investigators. Patients' consent preferences for research uses of information in electronic medical records: interview and survey data. *BMJ.* 2003 Feb 15;326(7385):373.
28. Goodman D, Johnson CO, Bowen D, et al. De- identified genomic data sharing: the research participant perspective. *J Community Genet* 2017;8(3):173–81.
29. Mathews DJH, Rabin JT, Quain K, Campbell E, Collyar D, Hlubocky FJ, Isakoff S, Peppercorn J. Secondary Use of Patient Tissue in Cancer Biobanks. *Oncologist.* 2019 Dec;24(12):1577-1583. doi: 10.1634/theoncologist.2018-0376. Epub 2019 Jun
30. Nair K, Willison D, Holbrook A, Keshavjee J *Health Serv Res Policy.* Patients' consent preferences regarding the use of their health information for research purposes: a qualitative study. 2004 Jan;9(1):22-7.
31. Taylor MJ, Taylor N. Health research access to personal Confidential data in England and Wales: assessing any gap in public attitude between preferable and acceptable models of *Life Sci Soc Policy* 2014;10(1).
32. Kaye The tension between data sharing and the protection of privacy in genomics research. *Annu Rev Genomics Hum Genet.* 2012;13:415-31. doi: 10.1146/annurev-genom-082410-101454. Epub 2012 Mar 9.
33. Xafis V. The acceptability of conducting data linkage research without obtaining consent: lay people's views and justifications. *BMC Med Ethics.* 2015;16(1):79. Published 2015 Nov 17. doi:10.1186/s12910-015-0070-4
34. Mählmann L, Schee Gen Halfmann S, von Wyl A, et al. Attitudes towards personal genomics and sharing of genetic data among older Swiss adults: a qualitative study. *Public Health Genomics* 2017;20(5):293–306.
35. Sanderson SC, Brothers KB, Mercaldo ND, et al. Public attitudes toward consent and data sharing in Biobank research: a large multi- site experimental survey in the US. *Am J Hum Genet* 2017;100(3):414–27.
36. Popejoy, B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* 538, 161–164 (2016). AND Crider KS, Reefhuis J, Woomert A, Honein MA. Racial and ethnic disparity in participation in DNA

- collection at the Atlanta site of the National Birth Defects Prevention Study. *Am J Epidemiol*. 2006;164(8):805–812. doi:10.1093/aje/kwj264
37. Dang JH, Rodriguez EM, Luque JS, Erwin DO, Meade CD, Chen , Jr Engaging diverse populations about biospecimen donation for cancer research. *J Commun Genet*. 2014;5(4):313–327. doi:10.1007/s12687-014-0186-0
 38. Wong M, Chia K, Yam W, Teodoro G, Lau K. Willingness to donate blood samples for genetic research: a survey from a community in Singapore. *Clin Genet*. 2004;65(1):45–51. doi:10.1111/j..2004.00192.x
 39. Global Alliance for Genomics and Health. "Framework for Responsible Sharing of Genomic and Health-Related Data: Development of " Paper presented at the 3rd Plenary Meeting of the Global Alliance for Genomics and Health, Leiden, The Netherlands, 2015.
 40. Fisher ER, Pratt R, Esch R, et al. The role of race and ethnicity in views toward and participation in genetic studies and precision medicine research in the United States: A systematic review of qualitative and quantitative studies. *Mol Genet Genomic Med*. 2020;8(2):e1099. doi:10.1002/mgg3.1099
 41. Trustworthy governance for sharing health-related data. Centre for Biomedical Ethics <https://medicine.nus.edu.sg/cbme/research/trustworthy-governance-for-sharing-health-related-data/>. Accessed 5 Oct
 42. Miles M, Huberman M. *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA, SAGE Publications;
 43. Integrated Health Information Systems. <https://www.ihis.com.sg/>. Accessed 24 Sep 2020.
 44. Government Technology Agency. <https://www.tech.gov.sg/>. Accessed 24 Sep
 45. Philipps M, Knoppers The discombulation of de-identification. *Nat Biotech*. 2016;34(11):1102– 3.
 46. Finck M, Pallas They who must not be identified—distinguishing personal from non-personal data under the GDPR. *Int Data Privacy Law*. 2020 10(1):11–36.
 47. Ipsos-MORI. 2016. The One-Way Mirror: Public attitudes to commercial access to health data. Wellcome Trust: UK. Available at: <https://www.ipsos.com/sites/default/files/publication/5200-03/sri-wellcome-trust-commercial-access-to-health-data.pdf> Accessed on 31 March,
 48. Xafis V. The acceptability of conducting data linkage research without obtaining consent: lay people's views and justifications. *BMC Med Ethics*. 2015;16(1):79. doi: 10.1186/s12910-015-0070- 4.
 49. McCormick N, Hamilton CB, Koehn CL, et Canadians' views on the use of routinely collected data in health research: a patient- oriented cross- sectional survey. *CMAJ Open* 2019;7(2):E203–9.
 50. Shah N, Coathup V, Teare H, et Motivations for data sharing- views of research participants from four European countries: a direct study. *Eur J Hum Genet* 2019;27(5):721–9.
 51. Kim KK, Sankar P, Wilson MD, et al. Factors affecting willingness to share electronic health data among California consumers. *BMC Med Ethics* 2017;18:25.

52. Majumder MA, Cook-Deegan R, McGuire AL. Beyond Our Borders? Public Resistance to Global Genomic Data Sharing. *PLoS Biol.* 2016;14(11):
53. Abdelhafiz AS, Sultan EA, Ziady HH et al. What Egyptians think. Knowledge, attitude, and opinions of Egyptian patients towards biobanking issues. *BMC Med Ethics.* 2019;20:57.
54. Walker J, Lovett R, Kukutai T, Jones C, Henry Indigenous health data and the path to healing. *Lancet.* 2017;390(10107):2022-3.
55. Cheok, J (2018) SingHealth hack 'worrying' for Singapore but govt response lauded. Available at: <https://www.businesstimes.com.sg/government-economy/singhealth-hack-worrying-for-singapore-but-govt-response-lauded> Accessed on 31 March, 2020; Mahmud, A.H (2019) Government accepts 5 measures to improve data security, to set up single contact for public to report breaches. Available at: <https://www.channelnewsasia.com/news/singapore/government-improve-data-security-contact-public-report-breaches-12130700>. Accessed on 31 March, 2020.
56. Chua BH, Murdoch U. *Communitarian Ideology and Democracy in Singapore.* London: Routledge; 2002.
57. Vayena E, Blasimme A. Health Research with Big Data: Time for Systemic Oversight. *J Law Med Ethics.* 2018;46(1):119-29.
58. Blasimme A, Vayena, E. Towards adaptive governance in big data health research: implementing regulatory principles. In Laurie G, editors. *Cambridge Handbook of Health Research Regulation.* Cambridge, UK: Cambridge University Press,
59. Ozdemir Innovating Governance for Planetary Health with Three Critically Informed Frames. *OMICS.* 2019;23(12), 623-30.
60. National Data Guardian. <https://www.gov.uk/government/organisations/national-data-guardian>. Accessed 5 Oct 2020.
61. Office of Privacy Commissioner (OPC) <https://www.privacy.org.nz/about-us/introduction/>. Accessed 5 Oct 2020
62. Data Governance Framework 2020. <https://www.aihw.gov.au/getmedia/a10b8148-ef65-4c37-945a-bb3effaa96e3/AIHW-Data-Governance-Framework.pdf.aspx>.
63. Ballantyne Adjusting the focus: A public health ethics approach to data research. *Bioethics.* 2019;33(3):357-66.
64. Ostrom, E. *Governing the Commons.* New York: Cambridge University Press;
65. Caron NR, Chongo M, Hudson M, et al. Indigenous Genomic Databases: Pragmatic Considerations and Cultural Contexts. *Front Public Health.* 2020;8:111. Published 2020 Apr 24. doi:10.3389/fpubh.2020.00111.
66. Integrated Data Infrastructure (IDI). <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>. Accessed 24 Sep
67. Ngā Tikanga Paihere. <https://data.govt.nz/use-data/data-ethics/nga-tikanga-paihere/#development>. Accessed 24 Sep

68. Department of Statistics, Singapore. 2020. Education, Language Spoken and Literacy. Available at: <https://www.singstat.govsg/find-data/search-by-theme/population/education-language-spoken-and-literacy/visualising-data> Accessed on 31 March, 2020.

Figures

Figure 1. Participants demographics. (a) Ethnicity. (b) Age. (c) Highest educational qualification attained. (PSLE: Primary School Leaving Examination. 'N', 'O' and 'A' levels: Singapore-Cambridge General Certificate of Education 'Normal', 'Ordinary' and 'Advanced' levels respectively).

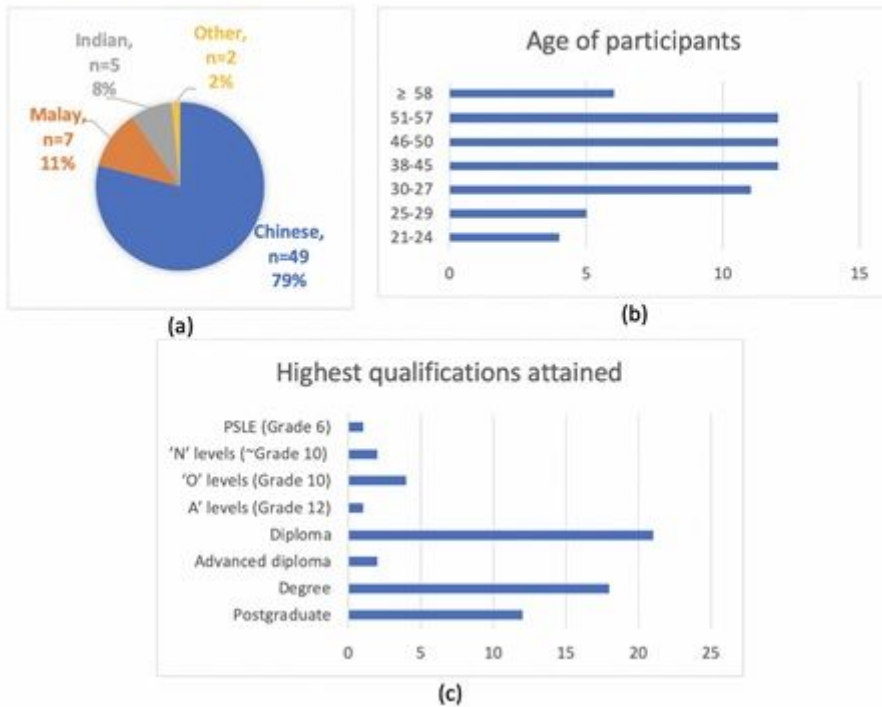


Figure 1

(caption in figure)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [COREQtable.docx](#)