International Journal of Population Data Science (2024) 9:1:24

International Journal of Population Data Science

Journal Website: www.ijpds.org

Research data use in a digital society: a deliberative public engagement

Kimberlyn M. McGrail[1,2,*], Jack Teng[2], Colene Bentley[3], Kieran C. O'Doherty[4], and Michael M. Burgess[2,5]

Submission History

Submitted:	21/11/2023
Accepted:	07/07/2024
Published:	14/10/2024

[1]Centre for Health Services and Policy Research, University of British Columbia, Vancouver, British Columbia, Canada
[2]School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada
[3]BC Cancer, Vancouver, British Columbia, Canada
[4]Department of Psychology, University of Guelph, Guelph, Ontario, Canada
[5]W. Maurice Young Centre for Applied Ethics, University of British Columbia, Vancouver, British Columbia, Canada

Abstract

Background

Sources of public and private data and ways to link them continue to evolve. This offers new opportunities for research, and new reasons for data-holding organisations to form partnerships. While research using these data can be beneficial, there is also a potential for negative consequences for some individuals or groups, including unintended or unanticipated effects. It is important to consult the public on how we might achieve both opportunities to link different types of data for research purposes, and protections against the misuse of data and the possibility of negative consequences.

Methods

Combining data sources for research was the topic of four days of deliberation held in British Columbia, Canada in late 2019. Public deliberation events bring diverse groups of people together to give direct input to policy makers, through carefully structured in-depth discussion on issues that are controversial and/or a source of public concern. Participants discussed whether data from electronic medical records should be used for research purposes, whether it is acceptable to combine data from public and private sources, who should authorise its use in research, and how a public advisory group on data use might be structured.

Results

Over four days, 29 residents of BC developed 17 deliberative conclusions that can be grouped into four broad topic areas: balancing benefit and potential harms when linking data; the protections that are expected to govern use of data; the type of authorisation required; and how the public should be involved in an ongoing way. Overall, the public is very supportive of research as long as oversight and controls are in place, including ongoing input from members of the public.

Conclusion

Deliberative conclusions from this event provide essential public input on the use of linked data for research, in particular when those data come from multiple sources. This is important information as policy-makers continue to develop legislation and practices around the use and linkage of both public and private sources of data.

keywords
public deliberation; data intensive science; data governance

---

[*]Corresponding Author:
Email Address: kim.mcgrail@ubc.ca (Kimberlyn M McGrail)

https://doi.org/10.23889/ijpds.v9i1.2372

Introduction

Digital technologies and the collection of data are ubiquitous. Wearable technology, internet-connected devices, and routine data collection when we use services all amass data that can be used for research [1–4]. These data are collected by or through various public and private entities such as researchers, governmental agencies, health care systems, other public and private service providers, app developers and device manufacturers from pacemakers to exercise monitors to cars [5, 6]. In many cases, data can be linked across sources, creating an ability to follow individuals' experiences across different systems, service use, and/or over time [7].

Digital data are readily accessible and can easily be moved, shared and used. New sources of data and links across data sources and types offer new opportunities for population data science [8], and new motivations to create partnerships across data-holding organisations, for example in the form of data trusts [9]. In health systems and research, access to and use of linked data can be used to provide important insights and evidence on a range of issues, for example to evaluate the effect of primary care reform [10, 11], to assess the impact of new technology implementation such as telehealth / virtual care [12–14], and to focus attention on specific conditions such as dementia [15–17].

British Columbia, similar to the rest of Canada, has a largely publicly-funded healthcare system that is delivered through a mix of public (e.g. hospitals) and private (e.g. pharmacists, most physicians) providers [18]. Despite the potential public benefit that could come from combining public and private sources of data for research, these same data could also be used in ways that negatively affect some individuals or communities. Unintended consequences could lead to stigmatisation of people or (for example) not being eligible for extended insurance coverage because of their health status. Moreover, because new analytical techniques and digital innovations are always being developed, future uses of data are unknown and hard to predict [19]. The potential harms from misuse of data and the possibility of revealing private information [20] must be mitigated as part of establishing data linkage practices.

While these risks are always present, existing research shows that members of the public are generally very supportive of data linkage and data use, as long as appropriate oversight and controls are in place [21–23], though this consensus decreases if the question includes private sector access [24]. Less is understood about public views when combining sources of data from the public and private sector, and using data that includes more sensitive and detailed personal information, for example from electronic medical records. Understanding public views is critical because the data being used are derived from individuals, these data are often collected without direct consent that identifies future uses [25], and trust is critical to the effective functioning of health and other systems [26–28]. Engaging with publics is also consistent with commitments to person-centred data-intensive research [29].

Combining data sources for research was the topic of a public deliberation event held in Vancouver, British Columbia (BC) over two weekends in late 2019. Participants considered rules and regulations that might ensure trustworthy governance of how new sources and types of linked data are used in the future. The aim of this deliberation was to inform the development of new data linkage activities and resources for researchers, and to ensure alignment of policies with public sentiment.

Methods

Public deliberation

Public deliberation events are informed by political theory on deliberative democracy. The idea is that a diverse group of citizens comes together to give informed input, through carefully structured in-depth deliberations, on issues that are controversial or a source of public concern [30]. The goal is for participants, working together, to come to conclusions about policy that accommodate their varied perspectives. The process of public deliberations enhances democratic legitimacy of decisions, and if deliberative conclusions are followed they can help ensure programs, actions, and decisions will be accepted [31].

The approach used for this public deliberation differs from other public consultations in several ways. Public consultations collect peoples' individual views, while deliberations create collective conclusions that reflect how participants think diverse interests are best accommodated [32]. Participants in deliberation events are given background information, ahead of time and during the event, including expert speakers who provide varying viewpoints on the subject matter. Discussions and the way participants create their deliberative conclusions are more in depth, with the whole process lasting several days [33], in this case four days over two weekends that are two weeks apart. The purpose of public deliberations is not to convince participants of any given position or bring them to consensus on the issues being discussed, but for participants to deliberate among themselves and reach either collective statements or deliberative conclusions that are acceptable to them as a group. The depth of engagement and participant-driven recommendations were key reasons we chose this method of public engagement.

Recruiting participants

A direct marketing company was engaged to mail letters of invitation to 10,000 randomly selected households from all parts of BC (the research team did not select the households). Household addresses were provided to the marketing company by Canada Post. Interested residents were asked to complete an online questionnaire on their demographic information, availability to participate in the deliberation, and to confirm they were not privacy professionals. Nearly 300 BC residents responded and met the eligibility criteria.

In order to capture a diverse range of life experience and perspectives, the online questionnaire asked for information on age, sex, income, self-identified ethnicity, and geographic region residence in the Canadian province of British Columbia (BC). In this case, we refer to diversity not as a statistical representation of BC (which would tend to value majoritarian views [34]) but as an attempt to have the broadest possible set of population-level views as part of the deliberation [35]. The only group explicitly excluded were people who work as privacy professionals; in public deliberation, it is best to avoid including "experts," (with respect to the topic at hand) because participants tend to defer to them, which can undermine the deliberation process [36].

Potential participants were then selected from the eligible pool through a stratified random process. We made special effort to recruit hard to reach groups, such as 18-24 year-olds, Indigenous persons, and those living in remote regions of BC, since these groups tend to be under-represented in public discourse and their perspectives are likely different and thus important to informing the overall deliberations [37]. The costs of travel, meals and accommodation were covered by project funding. Each participant also received a $125 honorarium per day of attendance.

Informing participants

Participants were encouraged to bring their opinions, values, and ideas about the topic to the deliberation but were not required to have prior knowledge of the use of linked data for research. To support their discussions and ensure they all had the same basic knowledge, a plain language information booklet [38] was provided two to three weeks before the meeting, and expert speakers gave presentations on the first day. The goal was to provide a broad range of views on the issues central to the deliberation so participants could be well-informed when engaging with and responding to other participants.

The information booklet was developed using academic literature and in consultation with expert advisors. It described what linked data are, how they are collected, what regulations are in place for sharing them, and issues and concerns surrounding their use. The intent was for the material to be presented at a grade 10 reading level and was reviewed by a member of the public not otherwise involved with the project. A glossary provided definitions of technical terms. The booklet was provided in both digital and physical formats.

The speakers at the event were selected to reflect a range of both expertise and positions on the issues. They were invited to speak and encouraged to bring their passion and reason to the event. The aim was to have a balanced and nuanced view of these issues through the diversity of views of speakers, without relying on a single speaker to provide that balance. The speakers included the President and CEO of Life Sciences BC, a health sector leader with expertise in privacy and research ethics, staff counsel of the BC Civil Liberties Association, and a researcher who uses electronic medical records in their research. The speakers also took part in an informal panel and answered participant questions.

Deliberation questions

Deliberative questions were developed by the research team, informed by consultations with data stewards and policy makers on current and relevant data-based issues, to ensure the deliberations focused on topics that can be actionable by policy-makers. The questions were: 1. Under what conditions is it acceptable to use data from electronic medical records for research? 2. Under what conditions is it acceptable to combine private sector and public sector data for research? 3. Who needs to authorise research that combines public and private data? 4. What are important features of an ongoing citizen advisory for decisions about data-based research in BC?

Deliberative process

The event format followed deliberative public engagement methods developed by Burgess, O'Doherty et al. [39] which have been used in previous deliberation events [40–42]. The four-day event was run by a large group facilitator who specialises in deliberative events (C. Bentley), and small group facilitators trained by the large group's facilitator. It took place on two non-consecutive weekends, with the break in between intended to give participants time to reflect on their discussions and discuss the topic with their families and friends. An overall flow of the two weekends is shown in Figure 1.

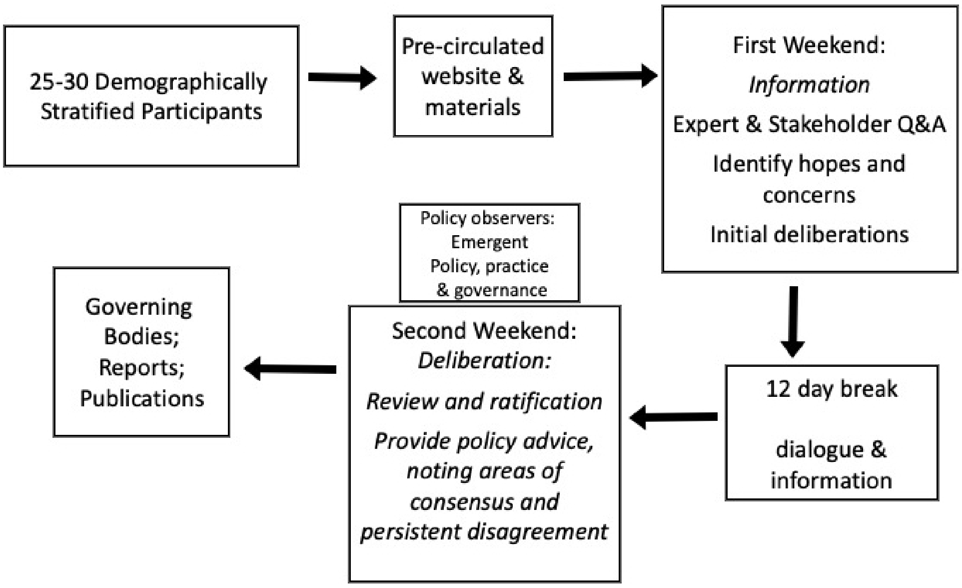
The first day of the event focused on informing participants about using and sharing linked data for research and introducing them to the process of public deliberation. Expert speakers gave presentations and answered participants' questions. Participants were told if they needed more information the research team would find it and present it to the group on subsequent days.

The second and third days were used to discuss the deliberation questions in both small and large groups. The four deliberative questions were given to participants to discuss in four separate small group sessions (i.e. one question per session), with the small groups remaining the same through both weekends. Each included scenarios designed by the research team to encourage discussion. Participants were asked to start by discussing the deliberative question and scenario and to gather feedback, identifying a broad range of viewpoints and what they would recommend as a way to proceed.

Back in the large (full) group, the purpose was for participants to develop recommendations for decision-makers developing policy on data-use issues. The large-group facilitator elicited the issues and perspectives raised in the small groups by listening and reflecting sentiments using the participants' own words. She then helped formulate the points raised into preliminary statements, which participants edited until they arrived at a collective statement that represented one conclusion of their deliberations. Participants then voted for or against the statement, or abstained, with each person explaining their reasons for taking their position. The extent of agreement (or disagreement) is documented so that policy-makers are able to identify topics or activities that need to be approached with greater caution because there is greater disagreement among publics' views. Put another way, a group that managed to converge on common conclusions in some cases while not finding convergence in others suggests there is a substantive controversy that is not yet resolved; this is important information for policy-makers. See Appendix 2 for the complete list of deliberation questions and the scenarios used for each of the questions.

On the fourth day, the deliberative conclusions were summarised and reviewed, and participants had the chance to vote again if new information and perspectives had changed their minds. The meeting concluded with a panel of experts who are in positions to influence policy on sharing linked data for research. The panel included an Assistant Deputy

Figure 1: The Structure of a deliberative process



Minister in charge of digital platforms for government, the scientific director of an organisation that provides linked data for approved research projects (K McGrail), and the chief privacy officer for the Ministry of Health. The panel members listened to the deliberative conclusions and then discussed them with participants.

While in previous deliberations, the policy statements were called "recommendations", here we concluded that word did not adequately describe what the participants developed. In some cases, the participants made statements that reflected a point of view they all held and wished to communicate to the policy-makers. In others, where there was not agreement, the participants voted on differently worded versions of statements to test whether their ideas converged. For simplicity, we refer to all of these as "deliberative conclusions."

The proceedings were audio recorded and transcribed. What is presented here represents a summary the collective statements that participants themselves arrived at and voted on during the process of deliberation itself. Here, we do not present quotes from individual participants, as would be common in other forms of qualitative research. This is important because for the purpose of reporting the outcomes of the deliberation, we give priority to the wording of statements that were explicitly considered by the group as a whole, rather than the statements of individuals throughout the process [43, 44]. Additional analyses of the statements made by participants individually are currently underway and will be reported separately. See Appendix 1 for the event's schedule.

Results

Twenty-nine members of the BC public participated in the deliberation on the first weekend, with 28 participating on the second weekend (one person left because of a family obligation). Participants were nearly evenly split between those who identified as male or female, with one "other", spanned all age groups, with 15 indicating a university degree or higher in education, 16 identifying as white, three as Indigenous and others as East Asian, South East Asian, African American, mixed race, and other. Participants were from all parts of the province. Further details of the demographic composition of the participants are in Appendix 3. Italics are used in the text below to indicate direct quotes of participants' deliberative conclusions (i.e. the collective statements developed by the deliberants, not of participants' individual statements).

As participants considered the deliberation questions, they developed ideas on a variety of related topics. They produced 17 deliberative conclusions for consideration by policy-makers. All of these and the voting associated with them are provided in more detail in a report on the deliberation that was provided to policy-makers [45]. To facilitate interpretation and reporting, these are grouped into four emergent categories:

1. Balancing benefit and potential harms when linking data;
2. Expected protections;
3. Required authorisation; and
4. Ongoing public involvement.

Balancing benefit and potential harms when linking data

Participants confirmed general support for linking and using data, with some important considerations that underlie many of the deliberative conclusions. First, there was universal agreement that *combining public and private data is acceptable if due consideration is given to the mitigation of discrimination*. The types of discrimination raised were broad, for example age, people applying for insurance, and people with certain kinds of health conditions who are seeking employment.

There was broad but not universal agreement that it is *acceptable to combine public and private data when the research has more potential for public benefit than risk*. There were a few who suggested that this framing might deter the private sector from useful research, and others felt that only talking about "potential" public benefit was too open to interpretation.

Most but not all participants agreed that it is *important that protections do not unduly restrict research using data*. This conclusion arose from participants' concern that excessive regulation could prevent research being done. Most participants thought it was important to maintain researchers' ability to pursue their interests (i.e., interest-driven academic projects). There were also concerns expressed that review processes might be unjustifiably rushed and protections would not be considered sufficiently, and there was some discussion of how to balance the importance of both research and privacy.

Expected protections

How to regulate, control and manage the sharing and use of linked data (i.e., data governance) were issues that arose continually throughout the two weekends. Generally, participants' conclusions called for secure and trustworthy processes to ensure potential abuses would be prevented, or caught and stopped. They concluded with nearly universal agreement that *it is acceptable to combine public and private data when the data linkage is done by a trusted third party*, i.e., an independent body. In addition, a strong majority wanted that body to be free from ties to government, because of underlying mistrust of government intentions and the risk that political interference will thwart research.

While supportive of this kind of linkage, participants also concluded (again almost universally) that *the use of data must be peer reviewed and appropriate for the proposed research, taking into consideration privacy issues*. This conclusion resulted from concerns over the risk to privacy some research presents (for example, research on small communities or rare diseases, where it might be possible to identify the people involved). Participants felt strongly that research proposals must be reviewed to identify privacy risks, and suggested that assessment could be part of peer reviews.

Participants concluded that the *independent oversight review committees should have authority to hold researchers accountable for the appropriate use of data and impose consequences for non-compliance*. Discussions included that there might be fines or revocation of privileges to use data, though there were some participants who were uncomfortable with a committee having authority to punish and wanted consequences to be more than just financial.

Participants discussed whether the independent review committees should be under government oversight, but ultimately most did not support this conclusion. The opposing majority felt the government could not be trusted to oversee data use and also wanted to separate the powers of the government and the oversight committee. The minority who supported government oversight said the government had people's best interests in mind and are representative of the people who voted for them. They also said basic trust of the government is necessary because it would be up to government to intervene if there were disagreement within the review committee.

Required authorisation

Several deliberative conclusions reflect how participants' discussion shifted from the general concept of authorisation to the more specific topic of consent, particularly informed consent. Participants were concerned about consent throughout the deliberation. Their concerns about informed consent were related to data ownership and people's power to retain control of their data. While not specifically raised in the booklet or by speakers, genetic data was also a key concern for participants who felt it was unique among forms of data.

Participants concluded (not universally) that *informed consent should be considered when using genetic data due to higher risk of unjustified discrimination*. The main concern was discrimination based on a genetic profile. A few, however, thought "discrimination" was inherently wrong, and were uncomfortable with the notion of "unjustified" discrimination. Other formulations of this deliberative conclusion did not increase consensus.

On data from electronic medical records, many but far from all participants felt *that it is acceptable to use all medical data (including EMR) without consent, provided there is an option to opt out*. Those in favour thought the conclusion would enable important research while providing sufficient opt-out provisions. Those opposed were concerned that individuals might be identified, and did not want this type of data to be used without their knowledge.

Nearly all participants concluded that *research that combines public and private data should be authorised by an independent multidisciplinary committee(s) that include but are not limited to expertise in ethics, law, commerce, science, and data management*. It was important to some to include "independent" in this statement.

Participants discussed but did not reach broad agreement on how data from small communities should be handled. There was extensive discussion about people living in smaller communities having a higher risk of being identified in research. Some felt these situations require informed consent. Others felt data should be treated equally regardless of community size and thought existing ethics rules were sufficient to protect the privacy of people in small communities.

Ongoing public involvement

There was broad agreement that there should be an ongoing public advisory relating to the use of linked data sets for research. There was also broad but not universal agreement that *an ongoing public advisory should be comprised of a diverse group of British Columbians, with diversity reflecting ethnicity, socioeconomic status, age, gender, education, geographic location, language, and disability*. Those who were not in full agreement felt the deliberative conclusion was a bit vague, or that diversity is good in theory but too difficult to implement in practice. Different considerations and wording did not increase consensus.

Participants felt nearly universally that *it is important that an ongoing public advisory meet face to face, but with an option to meet through other communication technologies as available*.

They recognised it would be expensive to bring advisory members from across BC to meet face-to-face, but felt that meeting in person was important to develop rapport and a working relationship among members. They further (again nearly universally) felt that such *an advisory should meet a minimum of twice a year and more as needed*. They concluded *that an ongoing public advisory should have membership terms of two years with 50% turnover annually*. They felt there should be turnover to prevent people from becoming complacent and ultimately less likely to challenge current rules or arrangements. Rotating people off would mean ensuring half the group remain and can share their knowledge with the new members.

Discussion

Consistent with previous deliberations on health data, participant conclusions here show strong and universal support for supporting research that use linked data, including data from multiple sources [21–23]. Equally consistent with existing literature, this support was context-specific rather than unconditional. The deliberative conclusions developed by participants in BC fall into four interdependent categories. There is broad and clear support for research using linked data, but always with specific context or expectations. Data must always be protected, users must meet standards (required authorisation) before they are allowed to access them, and there is an expectation of ongoing public input.

In many cases, the participants indicated that their support for a given deliberative conclusion was dependent on a previous deliberative conclusion. For example, the acceptability of using medical and EMR data without consent was contingent on having an oversight committee that assessed the research for ethical, privacy, and security concerns. In other words, there was a general sentiment that multiple, inter-related policies and procedures must be in place to ensure all of the considerations for data use are covered.

Participants readily identified that the use of sensitive linked data from public and private sources creates both opportunities and obligations. There was an emphasis from participants on the importance of the independent nature of the review committee and the trusted third party status of the data linking organisation. This is consistent with emerging literature around data intermediaries, data institutions and data trusts [9, 46, 47], which are collectives or organisations that organise and make multi-source data available through transparent and trustworthy systems that include multi-stakeholder governance [48].

All of this focuses on maximising the public good (the opportunity) that can come from data-intensive research. In many cases, participants' objection to certain proposed regulations was based on the concern that they would lead to unnecessary barriers to conducting promising research and thus deter the research from being conducted altogether. This was represented in discussions around conditions requiring informed consent, such as with research involving residents of small communities. At the same time, participants recognised a need to protect individuals in vulnerable populations (e.g. Indigenous peoples and people with rare diseases) and in small communities.

Although the participants voted on the deliberative conclusions, the total of those votes should not be over-interpreted. Voting was a tool the large-group facilitator used to assess convergence on a position and to identify the reasons in support of it. For example, while some participants who voted against a deliberative conclusion were truly opposed to it, others who voted against it may have had issues only with the conclusion's wording, sometimes because they wanted to take an even stronger position. Voting was used not to create quantitative information, but to provide contextual qualitative information to support understanding of the context and content of discussions.

Limitations

Public deliberations are a way to engage deeply with members of the public on topics that have trade-offs and that are not suitable for surveys or other more limited forms of involvement. While this approach has many benefits, the trade-off is that it can only touch directly a small number of people (in this case about 30). In addition, participants who have gone through a deliberative process are no longer representative of the general public since they have been informed about the issues and reflected deeply on them. It is therefore important that the conclusions of a public deliberation are not interpreted as reflective of the opinions of the general population. Rather, the deliberative conclusions should be interpreted as an important avenue of normative input for policy that draws its value from the diversity of the positions considered by the deliberants and the degree to which trade-offs between conflicting values and positions were addressed in the deliberative process. The previous successes of this method [49–52], and participation (including observation of the deliberation) by policy makers mitigates concerns raised by these considerations.

The process used in this type of deliberation tends to produce deliberative statements that have some level of agreement amongst participants, though that agreement is often not universal, as seen here. There is a risk that deliberative statements do not draw out divergent views held by a minority of participants, especially when facilitators drive too strongly toward consensus. In our process this limitation was addressed through the voting process so that the facilitator could explicitly identify opposing views and invite deliberants to express them. Further, when there was persistent disagreement, even after deliberation, there was no further push to consensus and the disagreement was allowed to stand and reported as an outcome.

This deliberation did not include a question on Artificial Intelligence, as that was not a topic of great concern when this work was developed in 2019. Given changes since then, it will be important for future deliberations and other forms of public engagement to elicit public views as they relate to data linkage and data use given recent development of AI.

Conclusion

The possibilities for research using linked data are evolving quickly. Such research creates opportunities to produce knowledge that benefits the public, but it also creates risks, including the possibility of privacy breaches and identification as well as community harm from ongoing stigma and harmful stereotypes. These concerns need to be addressed by policy makers, while also considering the diversity in the perspectives and interests of the public [53–55]. Doing so will help to ensure that policies, procedures, and guidelines balance the benefits with considerations of publics about possible harms. Paying close attention to the ideas of encouraging research with context-specific guardrails, appropriate authorisation, and ongoing input from publics will help to produce data sharing processes that are safe, trustworthy, and acceptable to the public.

Acknowledgements

We would like to acknowledge the funding support from the Canadian Institutes for Health Research (CIHR), the ongoing support and input from our policy partners and other collaborators, and the generosity of the participants in this deliberation.

Ethics statement

All procedures for this research project were reviewed and approved by the UBC Behavioural Research Ethics Board (Certificate # H19-01765).

Statement of conflict of interest

The authors declare no competing interests.

Data availability

Transcripts from this event are availability upon request to the corresponding author.

References

1. D. R. Witt, R. A. Kellogg, M. P. Snyder, and J. Dunn, "Windows into human health through wearables data analytics," *Curr. Opin. Biomed. Eng.*, vol. 9, pp. 28–46, Mar. 2019, https://doi.org/10.1016/j.cobme.2019.01.001.
2. A. K. Yetisen, J. L. Martinez-Hurtado, B. Ünal, A. Khademhosseini, and H. Butt, "Wearables in Medicine," *Adv. Mater.*, vol. 30, no. 33, p. 1706910, 2018, https://doi.org/10.1002/adma.201706910.
3. L. T. Dahl *et al.*, "The SPOR-Canadian Data Platform: a national initiative to facilitate data rich multi-jurisdictional research," *Int. J. Popul. Data Sci.*, vol. 5, no. 1, p. 1374, https://doi.org/10.23889/ijpds.v5i1.1374.
4. K. Jones, D. Ford, S. Thompson, and R. Lyons, "A Profile of the SAIL Databank on the UK Secure Research Platform," *Int. J. Popul. Data Sci.*, vol. 4, no. 2, p. 1134, https://doi.org/10.23889/ijpds.v4i2.1134.
5. L. B. Othmane, H. Weffers, M. M. Mohamad, and M. Wolf, "A Survey of Security and Privacy in Connected Vehicles," in *Wireless Sensor and Mobile Ad-Hoc Networks: Vehicular and Space Applications*, D. Benhaddou and A. Al-Fuqaha, Eds., New York, NY: Springer, 2015, pp. 217–247. https://doi.org/10.1007/978-1-4939-2468-4_10.
6. K. C. O'Doherty *et al.*, "If you build it, they will come: unintended future uses of organised health data collections," *BMC Med. Ethics*, vol. 17, no. 1, p. 54, Sep. 2016, https://doi.org/10.1186/s12910-016-0137-x.
7. A. Young and F. Flack, "Recent trends in the use of linked data in Australia," *Aust. Health Rev.*, vol. 42, no. 5, pp. 584–590, Aug. 2018, https://doi.org/10.1071/AH18014.
8. K. M. McGrail *et al.*, "A position statement on population data science: The science of data about people," *Int. J. Popul. Data Sci.*, vol. 3, no. 1, 2018, https://doi.org/10.23889/ijpds.v3i1.415.
9. P. A. Paprica *et al.*, "Essential requirements for the governance and management of data trusts, data repositories, and other data collaborations," *Int. J. Popul. Data Sci.*, vol. 8, no. 4, Art. no. 4, Sep. 2023, https://doi.org/10.23889/ijpds.v8i4.2142.
10. M. R. Lavergne *et al.*, "Do changes in primary care service use over time differ by neighbourhood income? Population-based longitudinal study in British Columbia, Canada," *Int. J. Equity Health*, vol. 21, no. 1, p. 80, Jun. 2022, https://doi.org/10.1186/s12939-022-01679-4.
11. N. Austin *et al.*, "Team-based primary care reforms and older adults: a descriptive assessment of sociodemographic trends and prescribing endpoints in two Canadian provinces," *BMC Prim. Care*, vol. 24, no. 1, p. 7, Jan. 2023, https://doi.org/10.1186/s12875-022-01960-z.
12. D. C. Butler *et al.*, "Changes in general practice use and costs with COVID-19 and telehealth initiatives: analysis of Australian whole-population linked data," *Br. J. Gen. Pract.*, vol. 73, no. 730, pp. e364–e373, May 2023, https://doi.org/10.3399/BJGP.2022.0351.
13. T. Kiran *et al.*, "Virtual Care and Emergency Department Use During the COVID-19 Pandemic Among Patients of Family Physicians in Ontario, Canada," *JAMA Netw. Open*, vol. 6, no. 4, p. e239602, Apr. 2023, https://doi.org/10.1001/jamanetworkopen.2023.9602.
14. C. Dahlgren, M. Dackehag, P. Wändell, and C. Rehnberg, "Determinants for use of direct-to-consumer telemedicine consultations in primary healthcare—a registry based total population study from Stockholm, Sweden," *BMC Fam. Pract.*, vol. 22, no. 1, p. 133, Jun. 2021, https://doi.org/10.1186/s12875-021-01481-1.
15. J. Stevenson-Hoare, A.-K. Schalkamp, C. Sandor, J. Hardy, and V. Escott-Price, "New cases of dementia are rising in elderly populations in Wales, UK," *J. Neurol. Sci.*, vol. 451, p. 120715, Aug. 2023, https://doi.org/10.1016/j.jns.2023.120715.

16. H. J. Welberry, H. Brodaty, B. Hsu, S. Barbieri, and L. R. Jorm, "Measuring dementia incidence within a cohort of 267,153 older Australians using routinely collected linked administrative data," *Sci. Rep.*, vol. 10, no. 1, p. 8781, May 2020, https://doi.org/10.1038/s41598-020-65273-w.
17. L. E. Williamson, J. Leniz, E. Chukwusa, C. J. Evans, and K. E. Sleeman, "A population-based retrospective cohort study of end-of-life emergency department visits by people with dementia: multilevel modelling of individual- and service-level factors using linked data," *Age Ageing*, vol. 52, no. 3, p. afac332, Mar. 2023, https://doi.org/10.1093/ageing/afac332.
18. L. Hedden and K. McGrail, "The best defence is a good offence: Ensuring equitable access to primary care in Canada," *Healthc. Manage. Forum*, vol. 36, no. 5, pp. 293–298, Sep. 2023, https://doi.org/10.1177/08404704231182260.
19. V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Reprint edition. Boston: Harper Business, 2014.
20. A. H. Seh *et al.*, "Healthcare Data Breaches: Insights and Implications," *Healthcare*, vol. 8, no. 2, p. 133, May 2020, https://doi.org/ 10.3390/healthcare8020133.
21. J. Teng, C. Bentley, M. M. Burgess, K. C. O'Doherty, and K. M. McGrail, "Sharing linked data sets for research: Results from a deliberative public engagement event in British Columbia, Canada," *Int. J. Popul. Data Sci.*, vol. 4, no. 1, 2019, https://doi.org/10.23889/ijpds.v4i1.1103.
22. S. Kalkman, J. Van Delden, A. Banerjee, B. Tyl, M. Mostert, and G. Van Thiel, "Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence," *J. Med. Ethics*, vol. 48, no. 1, pp. 3–13, Jan. 2022, https://doi.org/10.1136/medethics-2019-105651.
23. M. Aitken, J. de St. Jorre, C. Pagliari, R. Jepson, and S. Cunningham-Burley, "Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies," *BMC Med. Ethics*, vol. 17, no. 1, p. 73, Dec. 2016, https://doi.org/ 10.1186/s12910-016-0153-x.
24. The Wellcome Trust, "The One-Way Mirror : Public attitudes to commercial access to health data," no. March, pp. 1–161, 2016.
25. K. McGrail, M. J. Schull, and P. A. Paprica, "The public needs to know why health data are used without consent," The Conversation. Accessed : May 20, 2024. [Online]. Available: http://theconversation.com/the-public-needs-to-know-why-health-data-are-used-without-consent-123669.
26. T. Bubela, C. M. Flood, K. McGrail, S. E. Straus, and S. Mishra, "How Canada's decentralised covid-19 response affected public health data and decision making," *BMJ*, p. e075665, Jul. 2023, https://doi.org/10.1136/bmj-2023-075665.
27. K. McGrail, J. Morgan, and A. Siddiqi, "Looking back and moving forward: Addressing health inequities after COVID-19," *Lancet Reg. Health Am.*, vol. 9, p. 100232, May 2022, https://doi.org/10.1016/j.lana.2022.100232.
28. P. H. A. of Canada, "Expert Advisory Group Report 3 : Toward a world-class health data system." Accessed: May 20, 2024. [Online]. Available: https://www.canada.ca/en/public-health/corporate/mandate/about-agency/external-advisory-bodies/list/pan-canadian-health-data-strategy-reports-summaries/expert-advisory-group-report-03-toward-world-class-health-data-system.html
29. M. Aitken *et al.*, "Consensus statement on public involvement and engagement with data intensive health research," *Int. J. Popul. Data Sci.*, vol. 4, no. 1, 2019, https://doi.org/10.23889/ijpds.v4i1.586.
30. R. E. Goodin and J. S. Dryzek, "Deliberative Impacts: The Macro-Political Uptake of Mini-Publics," *Polit. Soc.*, vol. 34, no. 2, pp. 219–244, 2006, https://doi.org/10.1177/0032329206288152.
31. M. M. Burgess, "From 'trust us' to participatory governance: Deliberative publics and science policy," *Public Underst. Sci.*, vol. 23, no. 1, pp. 48–52, 2014, https://doi.org/10.1177/0963662512472160.
32. K. O'Doherty, F. P. Gauvin, C. Grogan, and W. Friedman, "Implementing a public deliberative forum," *Hastings Cent. Rep.*, vol. 42, no. 2, pp. 20–23, 2012, https://doi.org/10.1002/hast.28.
33. S. Solomon and J. Abelson, "Why and when should we use public deliberation?," *Hastings Cent. Rep.*, vol. 42, no. 2, pp. 17–20, 2012, https://doi.org/ 10.1002/hast.27.
34. D. Steel, N. Bolduc, K. Jenei, and M. Burgess, "Rethinking Representation and Diversity in Deliberative Minipublics," *J. Deliberative Democr.*, vol. 16, no. 1, Art. no. 1, Aug. 2020, https://doi.org/10.16997/jdd.398.
35. K. C. O'Doherty and M. M. Burgess, "Public deliberation to develop ethical norms and inform policy for biobanks: Lessons learnt and challenges remaining," *Res. Ethics*, vol. 9, no. 2, pp. 55–77, Jun. 2013, https://doi.org/10.1177/1747016113488858.
36. H. Longstaff and M. M. Burgess, "Recruiting for representation in public deliberation on the ethics of biobanks," *Public Underst. Sci.*, vol. 19, no. 2, pp. 212–224, 2010, https://doi.org/10.1177/0963662508097626.
37. R. W. Pong, M. DesMeules, and C. Lagacé, "Rural-Urban disparities in health: How does Canada fare and how does Canada compare with Australia?," *Aust. J. Rural Health*, vol. 17, no. 1, pp. 58–64, 2009, https://doi.org/10.1111/j.1440-1584.2008.01039.x.

38. "Website for Research Data Use in a Digital Society: A Deliberative Public Engagement." [Online]. Available: www.https://www.popdata.bc.ca/bcdatadeliberation2019.

39. M. Burgess, K. O'Doherty, and D. Secko, "Biobanking in British Columbia: discussions of the future of personalized medicine through deliberative public engagement," *Pers. Med.*, vol. 5, no. 3, pp. 285–296, 2008, https://doi.org/10.2217/17410541.5.3.285.

40. K. C. O'Doherty, M. K. MacKenzie, D. Badulescu, and M. M. Burgess, "Explosives, genomics, and the environment: Conducting public deliberation on topics of complex science and social controversy," *SAGE Open*, vol. 3, no. 1, pp. 1–17, 2013, https://doi.org/10.1177/2158244013478951.

41. C. Bentley, S. Costa, M. M. Burgess, D. Regier, H. Mctaggart-cowan, and S. J. Peacock, "Trade-offs , fairness , and funding for cancer drugs : key findings from a deliberative public engagement event in British," pp. 1–13, 2018.

42. S. M. Dry *et al.*, "Community recommendations on biobank governance: Results from a deliberative community engagement in California," *PLoS ONE*, vol. 12, no. 2, pp. 1–14, 2017, https://doi.org/10.1371/journal.pone.0172582.

43. K. C. O'Doherty and M. M. Burgess, "Engaging the Public on Biobanks: Outcomes of the BC Biobank Deliberation," *Public Health Genomics*, vol. 12, no. 4, pp. 203–215, Oct. 2008, https://doi.org/10.1159/000167801.

44. K. C. O'Doherty, "Synthesising the outputs of deliberation: Extracting meaningful results from a public forum," *J. Deliberative Democr.*, vol. 9, no. 1, Art. no. 1, Apr. 2013, https://doi.org/10.16997/jdd.160.

45. "BC Data Deliberation | Population Data BC." Accessed: Jun. 26, 2024. [Online]. Available: https://www.popdata.bc.ca/BCDataDeliberation#

46. N. Olorunju and R. Adams, "African data trusts: new tools towards collective data governance?," *Inf. Commun. Technol. Law*, vol. 0, no. 0, pp. 1–14, 2023, https://doi.org/10.1080/13600834.2023.2260678.

47. "Institutions for data governance: Building trust through collective action," in *World Development Report 2021: Data for Better Lives*, in World Development Report. , The World Bank, 2021, pp. 265–296. https://doi.org/10.1596/978-1-4648-1600-0_ch8.

48. Ada Lovelace Institute, "Participatory data stewardship: A framework for involving people in the use of data," Ada Lovelace Insitute, London, Sep. 2021. Accessed: Mar. 23, 2022. [Online]. Available: https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/.

49. K. C. O'Doherty, A. K. Hawkins, and M. M. Burgess, "Involving citizens in the ethics of biobank research: Informing institutional policy through structured public deliberation," *Soc. Sci. Med.*, vol. 75, no. 9, pp. 1604–1611, Nov. 2012, https://doi.org/10.1016/j.socscimed.2012.06.026.

50. K. O'Doherty, M. Burgess, and D. M. Secko, "Sequencing the salmon genome: A deliberative public engagement," *Genomics Soc. Policy*, vol. 6, no. 1, p. 15, Apr. 2010, https://doi.org/10.1186/1746-5354-6-1-15.

51. D. M. Secko, M. Burgess, and K. O'Doherty, "Perspectives on Engaging the Public in the Ethics of Emerging Biotechnologies: From Salmon to Biobanks to Neuroethics," *Account. Res.*, vol. 15, no. 4, pp. 283–302, Nov. 2008, https://doi.org/10.1080/08989620802388762.

52. P. H. Schwartz, K. C. O'Doherty, C. Bentley, K. K. Schmidt, and M. M. Burgess, "Layperson Views about the Design and Evaluation of Decision Aids: A Public Deliberation," *Med. Decis. Making*, vol. 41, no. 5, pp. 527–539, Jul. 2021, https://doi.org/10.1177/0272989X21998980.

53. M. Aitken, J. de St. Jorre, C. Pagliari, R. Jepson, and S. Cunningham-Burley, "Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies," *BMC Med. Ethics*, vol. 17, no. 1, p. 73, 2016, https://doi.org/ 10.1186/s12910-016-0153-x.

54. M. E. Turpel-Lafond, "In Plain Sight: Addressing Indigenous-specific Racism and Discrimination in B.C. Health Care," Government of British Columbia, Victoria, BC, 2021. Accessed: Nov. 21, 2023. [Online]. Available: https://www.bcchr.ca/sites/default/files/group-opsei/in-plain-sight-full-report.pdf.

55. A. Katz and A. Freier, "Recognition of inherent biases in administrative data," *Can. Fam. Physician*, vol. 68, no. 12, pp. 933–934, Dec. 2022, https://doi.org/10.46747/cfp.6812933.

Appendix 1

B. Event Schedule

DAY ONE: October 19th

8:00-9:00 AM	Breakfast and check-in
9:00-9:15 AM	Welcome address
9:15-10:00 AM	Participant and research team introductions
10:00-10:20 AM	Overview of the event and ground rules
10:20-10:40 AM	Break
10:40-11:00 AM	Sabrina Wong, PhD, Professor, School of Population and Public Health, University of British Columbia
11:00-11:20 AM	Wendy Hurlburt, CEO of LifeSciences BC
11:20-11:40 AM	Meghan McDermott, Staff Counsel, BC Civil Liberties Association
11:40-12:00 PM	Holly Longstaff, Director Privacy and Access, Provincial Health Services Authority
12:00-1:00 PM	Lunch
1:00-2:00 PM	Speaker panel discussion
2:00-2:10 PM	Introduction to Hopes and Concerns task and break-down of small groups
2:10-2:30 PM	Break (and reconvene in small groups)
2:30-3:30 PM	Small group discussions: Hopes and Concerns
3:30-4:15 PM	Large group discussions: Hopes and Concerns
4:15-4:30 PM	Overview of tasks and goals for Day 2

DAY TWO: October 20th

8:00-9:00 AM	Breakfast and check-in
9:00-9:30 AM	Overview of tasks and goals for the day, and introduction of deliberation question #1
9:30-10:30 AM	Small groups: Deliberation question #1: Under what conditions is it acceptable to use data from electronic medical records for research?
10:30-10:50 AM	Break
10:50-11:50 AM	Large group: Deliberation question #1
11:50-12:00 PM	Introduction to Deliberation question #2: Under what conditions is it acceptable to combine private sector and public sector data for research?
12:00-1:00 PM	Lunch
1:00-2:00 PM	Small groups: Deliberation question #2
2:00-2:20 PM	Break
2:20-3:30 PM	Large group: Deliberation question #2
3:30-3:45 PM	Are there questions we need to add to our agenda for Weekend 2?
3:45-4:00 PM	Overview of tasks and goals for weekend 2, check out

DAY THREE: November 2nd

8:00-9:00 AM	Breakfast and check-in
9:00-9:15 AM	Welcome back and overview of weekend
9:15-10:00 AM	Report back on questions from last weekend; introduction to deliberative question #3
10:10-11:10 AM	Small groups: Deliberation question #3: Who needs to authorize research that combines public and private data?
11:10-11:30 AM	Break
11:30-12:30 PM	Large group: Deliberation question #3
12:30-1:30 PM	Lunch
1:30-3:00 PM	Large group discussion
3:00-3:20PM	Break
3:20-3:30 PM	Introduction to deliberative question #4: What are important features of an ongoing citizen advisory for decisions about data-based research in BC?
3:30-4:30 PM	Small group: Deliberation question #4 on citizen advisories
4:30-4:45 PM	Overview of tasks for Day 4

DAY FOUR: November 3rd

8:00-9:00 AM	Breakfast and check-in
9:00-9:15 AM	Welcome back and overview of day
9:15-10:15AM	Large group: Deliberation question #5 on citizen advisories
10:15-10:35 AM	Break

10:35-11:35 AM	Large group: Review and revise recommendations
11:35-12:00 PM	Group photo!
12:00-1:00 PM	Lunch
1:00-1:30PM	Final questions: Is there anything we didn't get to?
1:30-2:30PM	Expert and policy panel discussion
2:30-2:50PM	Break
2:50-3:15 PM	Large group: Considerations from policy panel discussions
3:15-3:45 PM	Wrap up, check-out, and thank you!

Appendix 2

C. Scenarios to aid discussion for the deliberation questions

Deliberative question 1: Under what conditions is it acceptable to use data from electronic medical records for research?

Scenario 1: A researcher is interested in investigating the potential link between medicine use to treat hypertension among older adults and later development of dementia. In order to do this research; she needs access to data for a large number of people. The data would include: age and sex; year of original diagnosis of hypertension, blood pressure readings over time; the names and doses of prescribed drugs; and signs, symptoms and/or measures of dementia. Some of these data could come from administrative data, but some (like blood pressure readings, measures of dementia) would only be available in electronic medical records.

Scenario 2: A researcher is interested in developing new genetic tests that could be used to identify whether people might have a bad reaction to a drug, or whether different choices in drugs might be more or less effective. This idea is a form of "precision medicine" but how well it might work is unknown; this is the reason for the research. This research needs as much detailed health history as possible, from electronic medical records and other sources, as well as a genetic profile that would be linked to those other data. If successful, the intent is to include genetic profiles in electronic medical records, and to commercialise the testing procedure.

Deliberative question 2: Under what conditions is it acceptable to combine private sector and public sector data for research?

Scenario 1: Driver factors such as speeding, distraction, and impaired driving play a major role in most fatal and injury crashes. Many drivers use medications that may slow their reaction time, cloud judgment and impair the psychomotor skills required for safe driving. This potential threat to road safety may be increasing due to an aging population and increased use of psychotropic (or impairing) medications. Currently, despite international efforts, the risk to road safety associated with most medications is poorly understood. This knowledge gap hinders the development of effective policy, social marketing campaigns, and medication warning labels targeting people who drive while using impairing medications. Researchers hope to address this knowledge gap with research using data on health care and deaths linked with data from ICBC.

Scenario 2: The likelihood of being diagnosed with cancer increases with age, and it is known that both genetic and environmental factors influence the risk of cancer. Researchers are interested in getting a better understanding of this risk, and specifically the influence of diet. They are proposing a study that would link data on demographics, health care services use, education, and occupation information with data on grocery shopping drawn directly from large supermarket chains. The desire is for a very large research study, so there is no intent to have direct contact with anyone whose data are used in the study.

Deliberative question 3: Who needs to authorise research that combines public and private data?"

Scenario 1: A group of family physicians is interested in replicating a study done in Spain that looks at differences in outcomes when people take hypertension medication at bedtime vs. in the morning. (The Spanish study, which was small, but suggested a 30% decrease in cardiovascular events.) This research requires linkage of EMR data and administrative data and would benefit from linkage to home health monitoring data held by a private company.

Scenario 2: Researchers are interested in better measures of and predictions of frailty and think that combining a number of different data sets will give them better information to create a good predictive tool. They want to use EMR data including clinical notes, plus patient-reported information, information from apps on mobility and activity, public and private information on care aides, administrative records, and information on social supports like home care to do their research.

Scenario 3: Policy makers have increasing concerns about early childhood experiences and health and educational outcomes in later life. Researchers want to link early development data, health data, education data, app-based fitness data, and grocery shopping habits, and ideally income and occupational information to understand different pathways and identify important events or triggers that might lead to better or worse outcomes. This research intends to include as large a population as possible and does not need direct contact with participants.

Deliberative question 4: What are important features of an ongoing citizen advisory for decisions about data-based research in BC?

Scenario: Imagine that the BC Government has come to you for advice about how to obtain **ongoing public input** into decisions about data-based research in BC. The provincial government is looking for advice from members of the BC public on challenging issues similar to the ones we have discussed over the two weekends of this deliberation, namely, about the possible uses and applications of data-based research. They want to find out what's important to people about these issues, what their values are and how we can make decisions that are acceptable to the people of BC.

Considering what you have heard so far in this deliberation and your own opinions about how an advisory could work, what advice would you give to the government?

Appendix 3

Participant demographic composition

We made a specific effort to recruit individuals who are younger (aged 18-24) as previous deliberations have found this group difficult both to recruit and retain. We also focused on recruiting people who identified as Indigenous, as there are distinct norms and practices around data and data sharing in Indigenous communities that were important to reflect in the deliberations. Finally, we made efforts to recruit people who live in remote regions of BC, selecting people based on their Metropolitan Influence Zone (MIZ), a measure that shows whether an area with a low population density is truly remote or located close to a metropolitan area

Gender:
Male: 15
Female: 13
Other: 1

Age:
18-24: 3
25-34: 4
35-49: 10
50-64: 6
65+: 6

Education:
University or above: 15
Some university: 5
High school: 3
College or apprenticeship: 6

Ethnic identity:
African American: 2
White: 16
East Asian: 6
First Nations: 2
Metis: 1
Mixed race: 1
South East Asian: 3

Geographic location based on the five geographic health authorities in British Columbia:
Vancouver Coastal: 8
Fraser Valley: 7
Vancouver Island: 4
Northern BC: 6
Interior BC: 4

Metropolitan Influenced Zone: This is a categorisation of geography that is based on communiting patterns between smaller and larger population areas. It is intended to capture the degree of influence of a larger city (e.g. proximity) to cities or towns with smaller populations. This is graded from metropolitan areas (1) to rural / remote areas far from any metropolitan zone.

1: 21
2: 3
3: 2
5: 2
6: 1
7: 1