DATA-SHARING IN CANADIAN MEDICAL EDUCATION RESEARCH

Consensus Recommendations and Principles for Governance

JANUARY 2022

TABLE OF CONTENTS

ABOUT THIS DOCUMENT

The risks, benefits, values, principles, and recommendations for ethical, inter-institutional, data-driven medical education research in Canada presented here were developed in collaboration with Canadian medical education stakeholders and data stewards through 5 knowledge synthesis workshops aimed at building consensus. The workshops were augmented by pre- and post-event consultations; invited presentations from data sharing experts from the fields of ethics, higher education, library science, and epidemiology/health services; and an observer group of representatives of physician trainee advocacy organizations that ensured trainee perspectives were integrated into the final consensus statement.

FUNDING

This work was supported by the Social Sciences and Humanities Research Council Connection Grant, Education Investigator Award from the Department of Family & Community Medicine (University of Toronto), and the Department of Family Medicine (McMaster University).

CORRESPONDING AUTHORS

Lawrence Grierson, PhD
Associate Professor, Department of Family Medicine
McMaster University
100 Main St. W., Hamilton, ON, L8P 1H6
(905) 525-9140 x22738
griersle@mcmaster.ca

Kulamakan Kulasegaram, PhD
Associate Professor, Department of Community and Family Medicine
Temerty Chair, Learner Assessment & Evaluation
Temerty Faculty of Medicine,
University of Toronto
200 Elizabeth Street, 1ES-565., Toronto, Ontario, M5G 2C4
(416) 340-3646
mahan.kulaseagram@utoronto.ca

CITATION

Grierson, L., Kulasegaram, K., Button, B., Lee-Krueger, R., McNeill, K., Youssef, A., Cavanagh, A. Data-Sharing in Canadian Medical Education Research: Recommendations and Principles for Governance. Hamilton, Ontario: Jan. 6, 2022. Available from: www.dataconnection.ca

SUMMARY

By way of a SSHRC-sponsored Connection Project, entitled: "Research Data Management Capacity Building Initiative: Health Professions Education", stakeholders and data stewards from across the Canadian medical education community participated in a series of 5 successive invitation-only knowledge synthesis workshops that aimed to advance collective thinking and build consensus as it pertains to the collection and use of health professions education data for the purposes of medical training, health policy, and health services research, evaluation, and/or quality improvement (which we refer collectively to as "research" within this document).

These stakeholders and data custodians included representatives of physician training, physician licensure, physician certification, physician regulation, training program accreditation, and physician and physician-in-training advocacy organizations involved in the medical professional development trajectory. Through consideration of the potential benefits and risks associated with inter-institutional data-driven research, the series resulted in agreement about key principles and a set of 3 recommendations for the ethical utilization of education data for the purposes of medical education, health policy, and health services research in Canada.

Key Principles

Privacy & anonymity	Informed consent	Autonomy
Non-discrimination	Appropriate data collection	Appropriate research purposes

Recommendations

Inter-institutional data-driven education research should include:

1. The establishment of a governance body that includes knowledgeable representatives of all relevant stakeholders, ensuring the appropriate research purposes, transparency, informed consent, respect for data sovereignty, ethical review, and proactive oversight.
2. The involvement of a trusted data management facility, which enables the technical and procedural aspects of data sharing.
3. A commitment to managing data according to best practices that include the articulation of institutional strategies, the development of data management plans, and contemplation for data deposits.

A CALL TO ACTION

Studies using longitudinal medical education data are not currently the norm. However, the need for these studies will continue to grow as stakeholders increasingly call for evidence of accountability and impact of medical education. The status quo for the governance and oversight of this type of medical education scholarship is fraught with risks. The capacities and models necessary for sensitive and responsive governance need to be developed in the context of medical education and its stakeholders. Moreover, without thoughtful reflection and action on the risks and benefits of data sharing, our field will lag and perhaps create the conditions for unreflective and potentially harmful work. Therefore, the purpose of this consensus statement is to serve as a framework for addressing the difficult challenges of governance, ethics, accountability, and technical issues as we move into a state where 'big data' research becomes the norm, not the exception.

These values, principles, and recommendations are not intended to serve as a formal governing policy for inter-institutional data-driven education research in Canada. We recognize that each data steward and medical education institution in Canada operates within the context of a unique organizational and jurisdictional reality that dictates the data management processes that they must consider in pursuit of research evidence that supports their missions. It is our hope, however, that institutions and researchers acknowledge the relevance of these values, principles, and recommendations, embracing them as they collaborate in data-driven research across the medical education sector. We acknowledge that operationalization will be a challenge. We believe that firm principles can guide flexible implementation in each unique context.

We call upon the leaders of the Canadian medical education community to i) consider and adopt these principles and ii) to study the implementation of these principles and the enabling recommendations as appropriate.

THE RESEARCH DATA MANAGEMENT CAPACITY BUILDING INITIATIVE

Medical education is an interdisciplinary field that advances the theory and practice of education for future physicians. During the course of medical training, education programs generate vast amounts of data on students. This includes application and admissions information, assessment performance (e.g. tests), attendance records, professionalism reports, participation in extracurricular activities, licensing examination results, certification scores, records of clinical activities, locations of practice after graduation, and various other metrics from numerous institutional sources such as medical and health professional schools, licensing and certification authorities, regulatory bodies, and provincial health human resource agencies. Typically, these data are collected but not used beyond their pre-specified purposes (for e.g., to determine competence). It is even rarer that these data are deliberately linked across the periods of training and practice.

By way of a SSHRC-sponsored Connection Project, entitled: "Research Data Management Capacity Building Initiative: Health Professions Education", stakeholders and data stewards from across the Canadian medical education community participated in a series of 5 successive invitation-only knowledge synthesis workshops that aimed to advance collective thinking and build consensus as it pertains to the collection and use of health professions education data. In particular, the series brought together individuals from the fields of higher education, library and data sciences, and epidemiology/health services, which have made considerable advances in addressing data quality and data standards, to share their insights and experiences for consideration by those in the Canadian medical education field. The goal of this series was to determine key principles for the appropriate use of education data for the purposes of medical training, health policy, and health services research. This involved consideration for the wide array of duties and potential consequences that may accompany inter-institutional data sharing for scholarship.

The objectives of the Connection Project were to:

1. Facilitate the flow of knowledge pertaining to data ethics and policy among researchers, administrators, and stakeholders within the medical education community.
2. Support knowledge mobilization around data governance, quality, and management for research between individuals from the fields of library science, epidemiology/health services, and higher education that have developed solutions to similar challenges, and the medical education community.

The intended outcome of the Connection Project was a consensus statement for the ethical utilization of administrative education data for research in Canadian medical education that was consistent with the proposed Tri-Agency Policy on Research Data Management.

OUR APPROACH

The goals, perspectives, benefits, risks, values, recommendations, and principles of governance described in this document were derived via review of the literature; informed through pre-workshop consultations with Canadian medical education institutions that collect, generate and hold education data, medical education stakeholders, and education data science research leaders (see Appendix A for an overview of pre-workshop consultation outcomes); reflected through our own experience as education scientists working within the medical education research community in Canada; and advanced through the SSHRC-sponsored Connection Project with a national group of relevant stakeholders. This event was organized as a series of 5 successive knowledge synthesis workshops that aimed to advance collective thinking and build consensus. The workshops were held on March 25th, April 1st, April 7th, April 14th, and April 21st, 2021. Workshops were each 3 hours in length. In each session, the meeting organizers oriented the participants' attention to the proposed Tri-Agency Policy on Research Data Management, the status of the emerging synthesis, and the workshop activities.

Each workshop was augmented by an invited presentation from a relevant expert in the fields of ethics, higher education, library science, and epidemiology/health services; each of whom had either unique perspective on the ethical challenges of inter-institutional education data sharing or whom had made considerable progress in data sharing within their own field. The invited speakers of the Connection Project were Dr. Victoria McKinnon, representing learner perspectives, (McMaster University, OMSA, CFMS); Dr. Lisa Schwartz, a leading bioethicist (McMaster University); Dr. Daniel Corral, an experienced big data researcher in higher education, (University of Toronto); Dr. Elizabeth Wenghofer, who has conducted several longitudinal data projects linking education to clinical and health systems outcomes, (Laurentian University); and Jeff Gardner and Nick Rochlin (Portage Network; University of British Columbia) who provided expert perspectives on the technical and logistical issues in data sharing. These expert perspectives served to stimulate participant thinking regarding the key features and foundations for an agreement regarding the inter-institutional use of education data for the purposes of medical education research; including the wide array of duties and potential consequences that may accompany inter-institutional data sharing for research. In each workshop, participants engaged in session-specific facilitated group activities that were geared towards the development and refinement of the consensus agreement. In each case, these activities involved reflective breakout group exercises, which were followed by facilitated large group dialogue. Collectively, these activities aimed to promote empathy for all stakeholders, foster ideation, and provide opportunities for productivity directed towards agreement development. Notably, each session was attended by an observer group, composed of representatives of relevant physician trainee advocacy organizations. This group reflected on the workshop's progress, sharing their perspectives during a dedicated roundtable activity at the end of each session. This reflection ensured that ethical orientations pertaining to the trainee perspective were integrated into the final document. Workshop agendas, details about speakers and their presentations, the Connection Project workbook, and other key resources that supported the sessions can be accessed and reviewed at www.dataconnection.ca

In total, the workshops were attended by approximately 50 individuals, representing a wide variety of institutions and interests, with representation from the Association of Faculties of Medicine of Canada, Medical Council of Canada, College of Family Physicians of Canada, Royal College of Physicians and Surgeons of Canada, Federation of Medical Regulatory Authorities of Canada, Canadian Post-MD Education Registry, Canadian Federation of Medical Students, Black Medical Students Association of Canada, and the medical education and medical education research communities associated with Canada's 17 medical training institutions (Appendix B).

Throughout each session, the discussion was mediated such that the group considered the unique perspectives, constraints, and opportunities inherent to inter-institutional education data sharing and what these mean for medical education research in Canada.

This document reflects our critical synthesis of these discussions. During the workshops, we solicited ideas and feedback on pertinent propositions, and compiled field notes pertaining to the participants' responses and discussions. At the conclusion of each session, the team of facilitators reviewed and categorized these notes according to relevant themes, questions, and issues; a descriptive process aimed at capturing what was said by participants in the workshop. As notes coalesced, consensus ideas were identified, as were those ideas for which consensus remained contentious, ambiguous, or unclear. These ideas were then used by the series facilitators to drive discussions across subsequent sessions; such that, over the course of Connection Series, the following goals, benefits, risks, values, and recommendations were distilled.

Prior to its final presentation, draft iterations of this document were edited in response to gracious feedback provided by representatives of the Association of Faculty of Medicine of Canada, College of Family Physicians, Royal College of Physicians and Surgeons of Canada, Canadian Association for Medical Education, Black Medical Students Association of Canada, Society of Rural Physicians of Canada, Resident Doctors of Canada, Federation of Medical Regulatory Authorities of Canada, and Canadian Federation of Medical Students as well as Dr. Jean-Michel Leduc (University of Montreal) and Dr. Christina St-Onge (Sherbrooke University) (see Appendix C for a schematic overview of the workshop series).

SHARED GOALS AND PERSPECTIVES

The various stakeholders and data custodians in the medical professional development trajectory include physician training, physician licensure, physician certification, physician regulation, training program accreditation, and advocacy for physicians and physicians-in-training. While each represents its own separate function, through the workshop series we recognized a shared goal:

> *A self-regulating health human resource and Canadian healthcare system that is effective in constantly meeting the evolving healthcare needs of all persons who access it, as well as the professional aspirations of its physician constituents.*

The Connection Project conversations also highlighted that Canadian medical stakeholders and data custodians share 3 key perspectives:

- Medical education is an important factor in meeting the above-stated goal.
- Education data of all forms has the potential to support the generation of knowledge that may enhance education practice and policy. The lack of data sharing can impede timely and meaningful response to advancing medical education practice.
- Data-driven research has previously contributed to the adoption of policy and practice that has promoted harm and accentuated social inequity; and that potential exists in the current context. Perspectives on education and education research in this domain that are uninformed by historical inequalities, inconsiderate of the rights and priorities of marginalized communities, and not mindful of the value of autonomy can cause harm.

Inherent in these perspectives is a tension between balancing the potential good to be gained from inter-institutional data-driven research and the harm that this type of work may cause to physicians-in-training and physicians, their professional goals, and healthcare delivery in many communities in Canada.

BENEFITS OF INTER-INSTITUTIONAL DATA-DRIVEN EDUCATION RESEARCH

The conversation about inter-institutional data sharing for education research is pertinent because numerous advances in education data science serve as facilitators for this type of research: medical education institutions now collect and archive vast amounts of digitized medical education data; more and more tools are being developed to support the collection of standardized data across institutions; and developments in data infrastructure and research methods now permit rapid, large-scale analyses. Accordingly, even though there is still a need to establish appropriate information technology, data infrastructure, and organizational readiness, there is growing motivation and momentum throughout the medical education research community to leverage this new potential for its many benefits.

In this regard, the identified benefits of doing this type of work include:

- Enhancing education quality improvement and evaluation efforts. Through data sharing research, institutions may come to a better understanding of their own unique context.
- Facilitating the success of social accountability missions. Through data sharing research, we may come to a better understanding of the relationship between education policies and practices and community level healthcare system effectiveness and health outcomes.
- Enhancing education research quality. Data sharing permits creates more opportunities for research and scholarship that leads to education theory that is broadly generalizable. This work may attend to numerous domains of education research: assessment, admissions, professionalism, competence, skill development, *et cetera.*

RISKS OF INTER-INSTITUTIONAL DATA-DRIVEN EDUCATION RESEARCH

Despite the technical wherewithal and an articulation of potential benefit, inter-institutional data sharing for education research has not yet become common practice. There are numerous reasons for this, and barriers to progress include technical issues pertaining to data equivalence; the temporal constraints and resource implications of ethical review and legal documentation; a lack of appropriately representative governance structures; and a lack of data management standards. At the root of these barriers is also a prevalent concern that the pursuit of the presumed benefit will see data used in ways that exposes individuals and institutions to risks, which pertain to both the procedural aspects of data-sharing as well as the way the potential outcomes of data-sharing research can impact individuals and communities.

Outcome Risks

- Data-driven research may be used to discriminate physicians-in-training and physicians (in several contexts) on the basis of personal or social identity characteristics.
- Data driven research promotes a culture of surveillance, which can impact physician-in-training and physician wellness and sense of safety. This is particularly the case for individuals from groups that have been historically disadvantaged through methods of surveillance.
- The outcomes of data-driven research may be erroneous and/or misinterpreted in potentially harmful ways. This may occur because of data inaccuracy or insufficiency; lack of consideration for appropriate context; and/or inappropriate acknowledgement of research limitations.
- Data sharing gives way to new forms of inter-institutional comparison, some of which will negatively impact certain institutions with respect to public perception. Data-sharing research may also create opportunities for misinformed scrutiny resulting from over emphasis on particular measurements or forms of data.

Procedural Risks

The development of inter-institutional datasets has the potential to contravene institutional commitments to:

- Privacy and confidentiality. Threats to data security (i.e., leaks, hacks, theft), can comprise individual and institutional safety and privacy, and carry the potential for loss of data or de-anonymization. Privacy and confidentiality can also be compromised via processes of residual identification (i.e., identifying individuals through incorporation of a greater constellation of data). Institutional commitments to privacy and confidentiality are often contemplative of the Fair Information Principles presented in the Canadian Standards Association's Model Code for the Protection of Personal Information (https://laws-lois.justice.gc.ca/eng/acts/p-8.6/page-7.html).
- Informed consent. Individuals and institutions are not always aware of the ways in which their data may be used. Processes of obtaining consent or determining authority over data use may not cover all possible future applications.
- Promoting data sovereignty, or the idea that data must be subject to the governance structures within the context or community in which it is collected. This can refer to conceptualizations of trans-national data flow; however, in the current context it is more relevant to notions of Indigenous autonomy from post-colonial states and the desires of other communities for similar data autonomy. Data collection, analysis, interpretation is never truly 'neutral' or objective and a disassociation of data uses from the communities that provided the data can amplify oppression and marginalization.

RELEVANT VALUES

The Connection event has highlighted that a key concern for all participants is the building and maintenance of trust between stakeholders and researchers. Trust is seen as a key factor in ensuring the safety of individuals, communities, and institutions. Through our conversations, we have distilled those values that the participating medical education stakeholders and data custodians agree are fundamental to building this trust:

- *Informed Consent*: The collection and sharing of data about an individual should take place with the knowledge of the individual. This value includes providing impacted individuals with appropriate notice of data uses, a transparent view of research practices, and the ability to choose to participate.
- *Appropriate Data Collection*: Monitoring is a key feature of institutional accountability and fundamental to education assessment practices. Accordingly, physicians-in-training and physicians understand that their progress will be subject to data collection. However, the collection of data on physicians-in-training and physicians should be limited to those metrics that support the core missions of the relevant institution.
- *Appropriate Research Purposes*: The pursuit of data-driven education research in an opportunistic manner and/or simply for the purposes of leveraging available infrastructure and/or enhancing research productivity carries with it heightened possibilities that data may be misused, misinterpreted, or mismanaged (i.e., as a function of the articulated procedural and outcome risks). In this regard, the pursuit of opportunistic data innovation may detract from addressing prevalent and firmly established issues. It may also contribute to a culture of research productivity that does not lead to meaningful changes in education practice or policy.
- *Privacy and Anonymity*: Individuals should remain anonymous or obscure within research datasets. In this regard, the amount of information collected and shared should be minimized to that which is required for the particular purpose. Almost all institutions and jurisdictions have policies in place to protect the privacy of individuals. Data-driven research should abide by and incorporate the policies and practices relevant to the institutions and jurisdictions involved in the project.
- *Autonomy*: Data-driven research should not be used to limit the options or developmental possibilities of learners. Predictive analytics may be used to constrain an individual's ability to govern their own education, by leading or nudging people in certain directions. Data driven research should not influence learners or communities in a way that constrains their ability to pursue thoughtful and deliberate choice.
- *Non-discrimination*: Data need to be used in ways that do not discriminate or perpetuate inequity. Analyses that make predictions about individuals based on constellations of their information can perpetuate prejudices, profiling, and discrimination while also accentuating social stratification.

RECOMMENDATIONS

With the above articulated, we offer the following recommendations. Notably, our goal here is not to establish a specific policy, procedure, or approach to data sharing. Rather, the intention is to provide clarity for all parties in how inter-institutional data sharing in medical education for various purposes should be conceptualized. This is intended as a framework of a foundational set of recommendations, principles, and specific perspectives that should be contemplated and applied to any data sharing activity in medical education. In this regard, we recognize that the context of inquiry may have considerable influence on the way in which the following are enacted. These ideas may underpin formal relationships and serve as the basis for data sharing agreements that mediate inter-institutional education research; or they may guide the development of statements of reflexivity that serve to make transparent the ways in which data were managed and research was conducted. The hope is, regardless of the context, that researchers and those who manage and oversee education data are attentive to these recommendations and principles as they embark on their inquiry.

Notably, the recommendations address data sharing of existing data but aim also to be comprehensive in addressing new data collected during the implementation of medical education (see Appendix D for notes concerning the collection and sharing of social identity data).

Recommendations

We recommend the following:

- A governance body (or bodies) should be established in accordance with the principles of governance that follow. The role of this body would be to oversee data sharing agreements, educate the community about safe and appropriate data sharing, and review and audit of relevant data sharing academic activity. This governance body should include representation of the perspectives of learners, practicing physicians, data stewards, regulators, education institutions, researchers, and community members including knowledgeable representatives of marginalized and under-represented communities, as is appropriate for the data being shared and the research questions being pursued. Where resourcing or other constraints prevent consistent representation, mechanisms should be in place to consult the relevant stakeholders as needed. Additionally, this body should provide the function of additional ethical review to data sharing projects beyond local review and, accordingly, may also include subject matter experts in data management and privacy legislation. The governance body should be legitimately empowered to enact and oversee data sharing that aligns with the rules, policies, and laws of the sovereign jurisdictions in which the to-be-shared data have been collected.
- A trusted data management facility (or facilities) should be established to enable the technical and procedural aspects of data sharing in medical education research as well as to apply the procedures and policies approved by the governance body. This facility may administer de novo data sharing agreements, assume existing ones where governance and legislation approve, and work with existing data managers where relevant. The facility would also be responsible for quality assurance of the data, reporting to the governance body on safety of data and adherence to protocols. This facility would also ensure and oversee, as necessary, the disaggregation and return of data, as well as data forgetting (i.e., expiration of data). Such a facility may or may not have a role in data archiving.

- Data should be managed according to standards aligned with best practices, including those described in the proposed Tri-Council Policy on Research Data Management (i.e., considerations for articulated institutional strategies, data management plans, and data deposits). Further to these practices, we recommend that data shared for the purposes of analysis should minimize the use of individual identifiers. Special identifiers should be used for matching, and the maximal number of identifiers should be removed prior to analyses. There may be situations in which institutional markers may also be removed from data sets prior to analysis. There may be situations wherein the governance permits inclusion of this information; depending on the nature of the research question, the assent of the participating stakeholders, and the risks involved. Where possible, data management, processing, and transformation techniques should focus on and/or accentuate the variance accounted for by education policies, procedures, and systems rather than on that which pertains to individuals in the data; insofar as the research question dictates.

Principles of Governance

We propose that medical education data sharing across institutions for the purposes of scholarship and academic activity be governed by the following principles:

1. An *a priori* establishment of clear purposes, rationale, and methodology for the data sharing. While these can be broad and cover various scholarly activities, they should be clearly articulated along with a consideration of the benefits and harms. It may not be possible to articulate the specific research questions for every program of work supported by data sharing. However, governance should ensure the types of associations/linkages and methodologies used - including the provisions for protection of confidentiality, privacy, and minimization of harm – are clearly articulated.
2. Transparency in communicating these benefits and harms, risks, and protections, forms the bedrock of informed consent. Informed consent on the part of those who generate data in education systems including applicants, learners, as well as teachers, staff, and organizations is a non-negotiable aspect of research. This informed consent should be mediated via processes of meta-consent (i.e., pertaining to different potential types of secondary use as determined by a relevant data sharing agreement), which should be 'time-limited' (i.e., given for a certain period for the consented uses) and then terminated, renewed, or renegotiated. This informed consent is especially important for data sharing that involves social identity characteristics. Time-limited consent provides opportunity to respect the autonomy of data generators/participants with new information about risks/benefits resulting from new data created through data linkage as well as to the confirm the continued validity of data in the system. This process should be a shared responsibility between researchers, data sharing facilities, and governing bodies responsible for the accountability of the data sharing with the aim of minimizing the burden of consent for participants.
3. Data sharing creates unique process and outcomes risks that requires special governance and accountability that must be sensitive to risks and benefits as well as knowledgeable of the research processes. These risks are not evenly distributed among all populations and community. Some communities, especially those with a history of marginalization, should be given special consideration in the contemplation of the risks.

4. In furtherance of principle 3, data uses should be co-governed by all relevant stakeholders. This means that governance structures should include expert or knowledgeable representation from institutions, communities, and other groups which hold meaningful sovereignty, stewardship, or other relationships with the data. This is particularly important for communities to who have historically been marginalized or oppressed in relation to social and scientific practices. Knowledgeable members of these communities should be especially invited to inform the creation of governance.
5. The governance structures for data sharing should be guided by respect for 'data sovereignty', i.e., the rules/policies/laws within the community, jurisdiction, or institution in which it is collected. The relevant representatives should be identified a priori to any data sharing and engaged in the creation and/or review and refinement of the data sharing agreement. In this regard, researchers should continually engage with the governance body throughout the entire research process, including during the interpretation of results and prior to the dissemination of outcomes, so as to ensure that the research has meaningful benefit and/or does not perpetuate harms for affected communities, jurisdictions, and institutions.
6. Governance structures should review proposed uses of data against the criteria of respect for the foundational values (i.e., autonomy, informed consent, privacy and anonymity, appropriate data collection, appropriate purpose, non-discrimination), scientific merit, and potential for beneficial impact. Regardless of the intended activity of the data sharing (for e.g., quality improvement, program evaluation, research, etc.), the risks are the same and still relevant. Thus, the onus should be on researchers to engage in the appropriate levels of ethical review regardless of the category of academic activity.
7. Governance structures should be proactive and vigilant in overseeing and ensuring the use of data in accordance with the principles, processes, and specific policies established in data sharing agreements and approved research practices. Governance should promote data management plans aligned with best practices including the current and future policies adopted by the Tri-Council on research data management.

APPENDICES

APPENDIX A: PRE-WORKSHOP SERIES CONSULTATIONS AND THEMES

In preparation for the SSHRC Connection series, we engaged in a series of consultations with data custodians, research scientists, and student advocacy groups from across Canada.

- Custodians: Association of Faculties of Medicine of Canada, Medical Council of Canada, Federation of Medical Regulatory Authorities of Canada, College of Family Physicians of Canada, Royal College Physicians and Surgeons of Canada, Canadian Medical Protective Association
- Research Scientists: Christina St-Onge (Université de Sherbrooke), Jean-Michel Leduc (Université de Montreal), Suzanne Archie (McMaster University), Onye Nnorom (University of Toronto), LLana James (Queen’s University), Saad Chahine (Queen’s University), Douglas Archibald (University of Ottawa)
- Student Advocacy Groups: Black Medical Students Association of Canada, Indigenous Physicians Association of Canada, Canadian Federation of Medical Students

Our goal in these consultations was to identify key concerns pertaining to inter-institutional data-driven medical education research. Here, we highlight our synthesis of those conversations; distilling a series of themes that informed our conversations.

Theme 1: Balancing Social Accountability Mandates

Our consultants regularly pointed to their institutional/organizational social accountability mandates as a reason to participate or not participate in inter-institutional data sharing for research purposes. In doing, the group collectively highlights two tensions that need to be balanced when making decisions to participate in these initiatives.

a. The first tension pertains to the conflict that may exist within an organization’s own social missions. For instance, many groups recognize the power of data-driven research as an effective quality improvement tool to advance efforts towards better health workforce planning; improved education and outcomes of training; enhanced equity, diversity, and inclusivity in the medical profession; and/or improved societal and health outcomes. In this regard, there is a feeling that there is an imperative to make use of data for these ends. However, several of the same institutions highlight their social accountability to protecting their members/constituents from undue harms and respecting their privacy and confidentiality. In this regard, there is a parallel feeling that there is an obligation to not participate in data sharing for research purposes.

b. The second tension pertains to the conflicts that may exist across the various mandates of different institutions, and what this may mean for how research is conceptualized, data sets are used, and research findings are acted upon. For instance, medical schools are interested in data-driven education research to highlight areas for curricular or policy improvement. Regulators on the other hand have an interest in identifying potential risks to patient communities. In making this comparison, we can see that both groups may take different stances on which questions are posed, data are sought, and innovations are pursued.

Theme 2: The logistical or technical challenges of data sharing for research

Our consultants regularly pointed to the technical challenges of data sharing as a reason to participate or not participate in inter-institutional data-driven research. Many challenges were highlighted under this theme including those pertaining to the processes of harmonizing and standardizing data fields across organizations, building effective data sharing agreements between institutions, and conducting research activities within reasonable timelines of relevance.

Cross-cutting theme: The addition of identity-based data

Through our conversations, a singular topic emerged as relevant to both Theme 1 and Theme 2, which delineated two types of data: admissions and assessment data currently held and used by the various data stewards; and social identity data (e.g., pertaining to race, gender, ethnicity) that are not currently collected but that most stewards and organizations believe should be collected for the purposes of performing research that yields evidence that informs advances in social accountability missions

With respect to Theme 1, it was clear that our consultants recognized that the implications of data exchange are impacted when one shifts focus to include social identity data. These data were clearly seen as critical for the advance of effective equity, diversity, and inclusivity mandates. However, many noted that the simple collection and analysis of these data is insufficient. In this regard, the tension between use and misuse was very prevalent. There are concerns (present and historical) that social identity data can be used to harm, stigmatize, and not improve the lives of those represented. In this regard, our consultations highlighted the importance of understanding the context in which these data are situated and the need to acknowledge that the potential for pro-social research benefits to be offset by oppressive outcomes exists at all stages of the data-driven research continuum - from research question through to the interpretation of findings. With respect to Theme 2, our discussions also highlighted that the inclusion of social identity data has implications for technical aspects of data sharing. In particular, the inclusion of social identity data enhances the focus on the composition of the data governance structures, which mediate research approvals and data sharing agreements. These data also increase the difficulty in maintaining student anonymity within a dataset.

APPENDIX B: LIST OF PARTICIPANTS

Doug Archibald (University of Ottawa)

Cassandra Barber (Maastricht University)

Geoffrey Barnum (CAPER)

Ariel Barrientos (BMSAC, University of Toronto)

Ilona Bartman (MCC)

Brenton Button (NOSM)

Alice Cavanagh (McMaster University)

Erin Cameron (NOSM)

Saad Chahine (Queen's University)

Shen Chu (McMaster University)

Daniel Corral (University of Toronto)

Annabelle Cumyn (Universite de Sherbrooke)

Gurm Dhugga (UBC)

Holly Fleming (NOSM)

Jeffrey Gardner (Portage Network, UBC)

Qi Guo (MCC)

Mark Hanson (University of Toronto)

Deena Hamza (University of Alberta)

Tanya Horsley (RCPSC)

Rebekah Jacques (Western University)

Christopher Jones (University of Toronto)

Eleni Katsoulas (Queen's University)

Jonathan Kimball (AFMC)

Dragan Kljujic (CFPC)

Jean-Michel Leduc (Universite de Montreal)

Hana Lee (University of Toronto)

Rachelle Lee-Krueger (University of Ottawa)

Fleur-Ange Lefebvre (FMRAC)

Henry Li (CFMS, University of Manitoba)

Cathy MacLean (University of Saskatchewan)

Laura McEwen (Queen's University)

Victoria McKinnon (CFMS, OMSA, McMaster University)

Kestrel McNeill (McMaster University)

Jeffrey Moat (Pallium Canada)

Susan Moffatt-Bruce (RCPSC)

Sandra Monteiro (McMaster University)

Ivy Oandasan (CFPC, University of Toronto)

Gbolahan Olarewaju (BMSAC, UBC)

Ghislaine Pilot-Attema (NOSM)

Nick Rochlin (Portage Network, UBC)

Shelley Ross (CAME; University of Alberta)

Lisa Schwartz (McMaster University)

Bolade Shipeolu (BMSAC, University of Toronto)

Steve Slade (CFPC)

Christina St-Onge (Universite de Sherbrooke)

Brent Thoma (University of Saskatchewan)

Kimberley Thomas (BMSAC, UBC)

Elizabeth Wenghofer (Laurentian University)

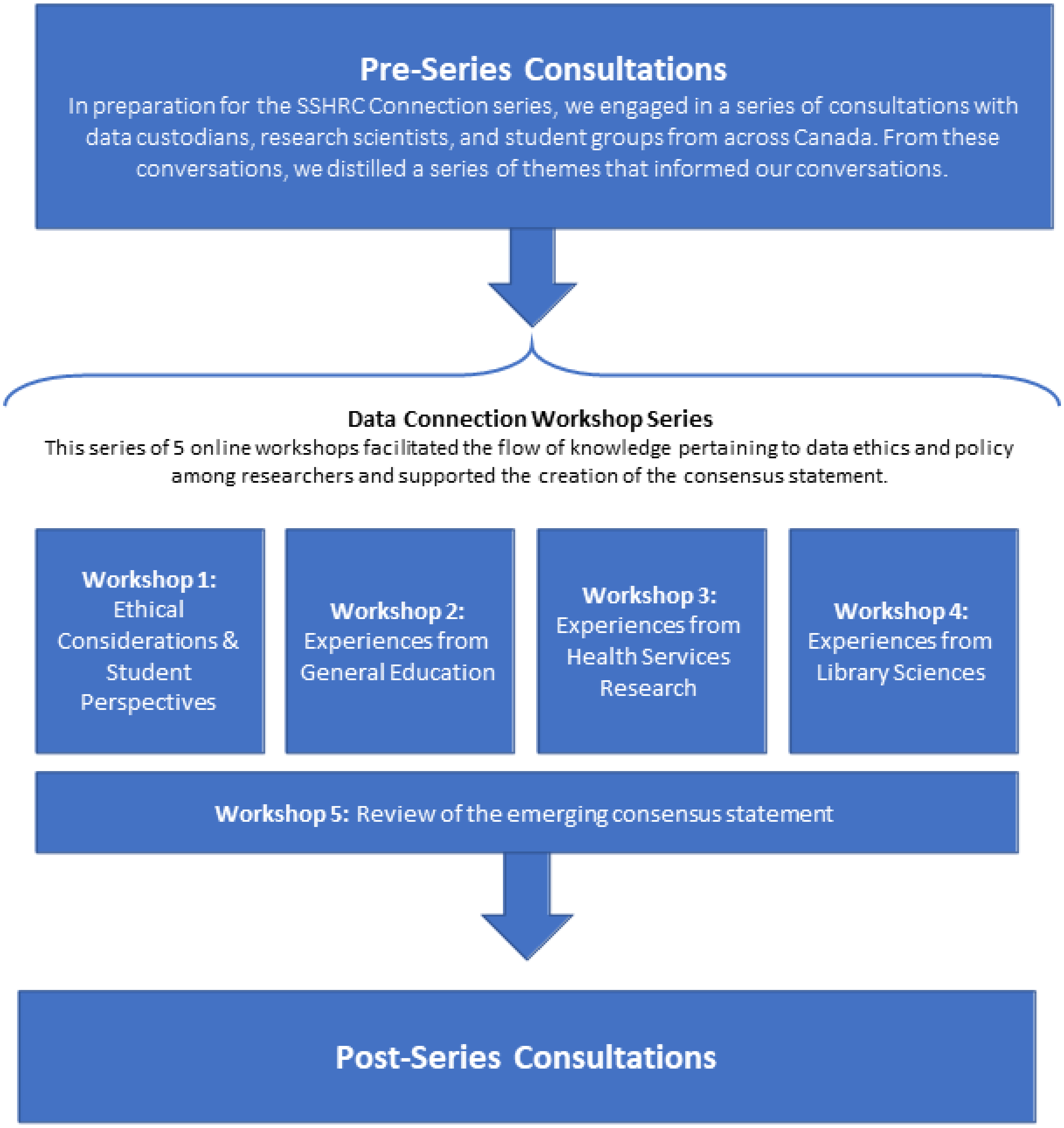
Brianne Wood (NOSM)

Meredith Young (McGill University)

Alaa Youssef (University of Toronto)

Note: Not all participants attended all sessions.

APPENDIX C: OVERVIEW OF EDUCATION DATA CONNECTION SERIES



Pre-Series Consultations

In preparation for the SSHRC Connection series, we engaged in a series of consultations with data custodians, research scientists, and student groups from across Canada. From these conversations, we distilled a series of themes that informed our conversations.

Data Connection Workshop Series

This series of 5 online workshops facilitated the flow of knowledge pertaining to data ethics and policy among researchers and supported the creation of the consensus statement.

Workshop 1: Ethical Considerations & Student Perspectives

Workshop 2: Experiences from General Education

Workshop 3: Experiences from Health Services Research

Workshop 4: Experiences from Library Sciences

Workshop 5: Review of the emerging consensus statement

Post-Series Consultations

APPENDIX D: CONSIDERING THE COLLECTION AND SHARING OF SOCIAL IDENTITY DATA

The conversations over the Connection Series brought to the fore one particularly compelling suggestion concerning education research that requires the collection and sharing of social identity data; specifically: *that the risks of inter-institutional data-sharing are too high and that we should not engage in this type of education research*. This is an idea that has been positioned as maintaining the status quo.

In this regard, we recognize that recent calls for the collection and analysis of social identity data frame this type of activity in support of inquiry that may yield evidence that informs advances in theories of social accountability and the outcomes of equity, diversity, and inclusivity missions. While many institutions and advocacy groups engage in some manner of social demographic data collection, usually in the form of anonymized surveys that give way to de-identified aggregated datasets, these data are not currently collected in a standardized way that links these characteristics directly to physicians-in-training and physicians.

The Connection event has highlighted that the tension between the potential use and misuse of these data is very prevalent. There are present and historical concerns that social identity data can be used to harm, stigmatize, and not improve the lives of those represented by the data. Accordingly, it was agreed that we must acknowledge the potential that pro-social research benefits may be offset by oppressive outcomes at all stages of the data-driven research continuum - from research question through to the interpretation of findings, and that the inclusion of social identity data increases the likelihood that an individual may be identified within a dataset.

In this regard, we recognize that the values mentioned above should be especially contemplated in the conduct of any education research that involves the collection and/or sharing of social identity data. It is imperative that this type of data use supports research that addresses and does not perpetuate social inequity. It must acknowledge that these data increase the likelihood that individuals may be identified within data sets and that the interpretations of predictive analyses may serve to decrease learner autonomy. In the case that *a priori* expressed research purposes, rationales, and methods are unable to account for these potential risks, then it is our recommendation that social identity data pertaining to physicians-in-training and physicians are not collected, shared, or analyzed.