Document généré le 7 mai 2024 19:27

Canadian Medical Education Journal
Revue canadienne de l'éducation médicale

Inter-institutional data-driven education research: Consensus values, principles, and recommendations to guide the ethical sharing of administrative education data in the Canadian medical education research context

Recherche interinstitutionnelle en éducation fondée sur des données : valeurs consensuelles, principes et recommandations pour guider le partage éthique des données administratives relatives à l'enseignement médical à des fins de recherche au Canada

Lawrence Grierson, Alice Cavanagh, Alaa Youssef, Rachelle Lee-Krueger, Kestrel McNeill, Brenton Button et Kulamakan Kulasegaram

Volume 14, numéro 5, 2023

URI : https://id.erudit.org/iderudit/1107924ar
DOI : https://doi.org/10.36834/cmej.75874

Aller au sommaire du numéro

Éditeur(s)

Canadian Medical Education Journal

ISSN

1923-1202 (numérique)

Découvrir la revue

Citer cet article

Grierson, L., Cavanagh, A., Youssef, A., Lee-Krueger, R., McNeill, K., Button, B. & Kulasegaram, K. (2023). Inter-institutional data-driven education research: Consensus values, principles, and recommendations to guide the ethical sharing of administrative education data in the Canadian medical education research context. *Canadian Medical Education Journal / Revue canadienne de l'éducation médicale*, *14*(5), 113–120. https://doi.org/10.36834/cmej.75874

Résumé de l'article

Contexte : Des données administratives sont générées dans le cadre de la formation des médecins, d'octroi de permis d'exercice et de réglementation des activités professionnelles, mais ces données sont rarement utilisées au-delà de leurs objectifs prédéfinis. Il convient de créer un système de supervision réactif et sensible aux risques pour permettre le partage de données relatives à l'enseignement médical entre établissements à des fins de recherche.

Méthode : Une initiative pancanadienne de recherche de consensus a été réalisée pour parvenir à un accord sur les objectifs, les avantages, les risques, les valeurs et les principes qui devraient sous-tendre la recherche interinstitutionnelle sur l'enseignement médical à l'aide des données existantes. Ce projet s'est appuyé sur une analyse de la littérature scientifique, sur des consultations avec diverses parties prenantes et sur cinq ateliers successifs de synthèse des connaissances. Des discussions ont été menées sur la base de propositions formulées préalablement jusqu'à la cristallisation d'un accord collectif.

Résultats : Un consensus s'est dégagé autour de six principes clés : la création a priori d'objectifs, d'une logique et d'une méthodologie clairs pour la recherche interinstitutionnelle fondée sur les données; l'obtention, sans exception, du consentement éclairé des personnes concernées par la collecte de données dans les systèmes d'éducation; la création d'un cadre de gouvernance visant spécifiquement le partage des données entre établissements; le respect, dans ce cadre, de la souveraineté des données; l'utilisation des données fondée sur un ensemble de valeurs partagées; et l'application des meilleures pratiques en matière de gestion des données de recherche.

© Lawrence Grierson, Alice Cavanagh, Alaa Youssef, Rachelle Lee-Krueger, Kestrel McNeill, Brenton Button, Kulamakan Kulasegaram, 2023

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Canadian Medical Education Journal

Commissioned Scientific Reports

Inter-institutional data-driven education research: consensus values, principles, and recommendations to guide the ethical sharing of administrative education data in the Canadian medical education research context

Recherche interinstitutionnelle en éducation fondée sur des données : valeurs consensuelles, principes et recommandations pour guider le partage éthique des données administratives relatives à l'enseignement médical à des fins de recherche au Canada

Lawrence Grierson,[1] Alice Cavanagh,[2] Alaa Youssef,[3] Rachelle Lee-Krueger,[4,5] Kestrel McNeill,[6] Brenton Button,[7,8] Kulamakan Kulasegaram[9]

[1]Department of Family Medicine, and Scientist, McMaster Education Research, Innovation, and Theory (MERIT), Faculty of Health Sciences, McMaster University, Ontario, Canada; [2]Michael G. DeGroote School of Medicine, McMaster University and Health Policy PhD Program, Faculty of Health Sciences, McMaster University, Ontario, Canada; [3]Stanford School of Medicine, Department of Radiology, Stanford Center for Artificial Intelligence in Medicine and Imaging, California, USA; [4]Faculty of Education, University of Ottawa, Ontario, Canada, [5]Education Consultant, Office of Continuing Medical Education & Professional Development, Cumming School of Medicine, University of Calgary, Alberta, Canada; [6]Department of Psychology, Neuroscience & Behaviour, McMaster University, Ontario, Canada; [7]Northern Ontario School of Medicine University, Ontario, Canada; [8]Faculty of Education, University of Winnipeg, Manitoba, Canada; [9]Department of Community and Family Medicine, University of Toronto and Scientist at Wilson Centre, University of Toronto, Ontario, Canada

Correspondence to: Lawrence Grierson, PhD; Department of Family Medicine, McMaster University, 100 Main St. W., Hamilton, ON, L8P 1H6; phone: (905) 525-9140 x22738; email: griersle@mcmaster.ca

Edited by: Marcel D'Eon (editor-in-chief)

Published ahead of issue: Jun 21, 2023; published: Nov 8, 2023. CMEJ 2023, 14(5) Available at https://doi.org/10.36834/cmej.75874

© 2023 Grierson, Cavanagh, Youssef, Lee-Krueger, McNeill, Button, Kulasegaram; licensee Synergies Partners. This is an Open Journal Systems article distributed under the terms of the Creative Commons Attribution License. (https://creativecommons.org/licenses/by-nc-nd/4.0) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited.

Abstract

Background: Administrative data are generated when educating, licensing, and regulating future physicians but these data are rarely used beyond their pre-specified purposes. The capacity necessary for sensitive and responsive oversight that supports the sharing of administrative medical education data across institutions for research purposes needs to be developed.

Method: A pan-Canadian consensus-building project was undertaken to develop agreement on the goals, benefits, risks, values, and principles that should underpin inter-institutional data-driven medical education research in Canada. A survey of key literature, consultations with various stakeholders and five successive knowledge synthesis workshops informed this project. Propositions were developed, driving subsequent discussions until collective agreement was distilled.

Results: Consensus coalesced around six key principles: establishing clear purposes, rationale, and methodology for inter-institutional data-driven research *a priori*; informed consent from data generators in education systems is non-negotiable; multi-institutional data sharing requires special governance; data governance should be guided by data sovereignty; data use should be guided by an identified set of shared values; and best practices in research data-management should be applied.

Résumé

Contexte : Des données administratives sont générées dans le cadre de la formation des médecins, d'octroi de permis d'exercice et de réglementation des activités professionnelles, mais ces données sont rarement utilisées au-delà de leurs objectifs prédéfinis. Il convient de créer un système de supervision réactif et sensible aux risques pour permettre le partage de données relatives à l'enseignement médical entre établissements à des fins de recherche.

Méthode : Une initiative pancanadienne de recherche de consensus a été réalisée pour parvenir à un accord sur les objectifs, les avantages, les risques, les valeurs et les principes qui devraient sous-tendre la recherche interinstitutionnelle sur l'enseignement médical à l'aide des données existantes. Ce projet s'est appuyé sur une analyse de la littérature scientifique, sur des consultations avec diverses parties prenantes et sur cinq ateliers successifs de synthèse des connaissances. Des discussions ont été menées sur la base de propositions formulées préalablement jusqu'à la cristallisation d'un accord collectif.

Résultats : Un consensus s'est dégagé autour de six principes clés : la création a priori d'objectifs, d'une logique et d'une méthodologie clairs pour la recherche interinstitutionnelle fondée sur les données; l'obtention, sans exception, du consentement éclairé des personnes concernées par la collecte de données dans les systèmes d'éducation; la création d'un cadre de gouvernance visant spécifiquement le partage des données entre établissements; le respect, dans ce cadre, de la souveraineté des données; l'utilisation des données fondée sur un ensemble de valeurs partagées; et l'application des meilleures pratiques en matière de gestion des données de recherche.

Conclusion: We recommend establishing a representative governance body, engaging trusted data facility, and adherence to extant data management policies when sharing administrative medical education data for research purposes in Canada.

Conclusion : En vue du partage des données administratives relatives à l'enseignement médical à des fins de recherche au Canada, nous recommandons la création d'une instance de gouvernance représentative ainsi que l'utilisation d'infrastructures fiables et le respect des politiques existantes régissant la gestion des données.

Introduction

Throughout training, a vast amount of data on future physicians is generated - admissions information, assessments, attendance records, professionalism reports, logs of clinical activities, certification results, and much, much more – which are held by numerous institutions, including medical schools, licensing bodies, regulatory authorities, and health human resource agencies. While regularly collected, these data are rarely used beyond their pre-specified purposes (e.g., to assess competency). Recently, however, there have been calls to make better use of these data to create new insights for educational practice.[1] These calls are further amplified with widespread curricular reforms currently underway.[2,3] Previous stand-alone investigations using administrative education data have been effective at evaluating training efficacy in relation to educational,[4] professional,[5] and patient outcomes.[6] Accordingly, the need for a structural foundation that makes inter-institutional data-driven research in medical education more feasible and sustainable is recognized.

The challenge here is not in methodological or technological expertise. Rather, it relates to the need for building the governance conditions that are pre-requisites for this type of research: Scholars of higher education have highlighted the ethical, legal, privacy, and autonomy-related issues inherent to this type of work.[7-9] Without similar consideration for these issues in the context of medical education, advances in data-driven research in our field is fraught with risks. Institutions must understand when it is safe to share data; learners should recognize the potential benefit or harm data-driven research poses to them; and researchers must have clarity on how to engage in this research in a respectful and sustainable manner. In this regard, the capacity for sensitive and responsive oversight needs to be developed. Without this, our field will lag and increase the potential for unreflective and harmful work. To this end, we recently engaged a comprehensive group of Canadian medical education stakeholders and data stewards to advance collective thinking and build agreement. Our goal was to generate consensus recommendations and principles for governance that can support the collection, sharing, and use of administrative medical education data for the purposes of research.

Method

Our process of building consensus recommendations and principles for governance was informed by 1) key literature concerned with data ethics for education, 2) a set of pre-workshop consultations, 3) a series of five successive 3-hour knowledge synthesis workshops held online between March 25th and April 21st, 2021, and 4) a set of post-workshop consultations. At each stage, we developed a set of propositions concerning the risks, benefits, values, and principles that should underpin the ethical conduct of inter-institutional data-driven medical education research in Canada. Throughout the consensus-building process, we solicited ideas and feedback on our propositions from representatives of the Association of Faculties of Medicine of Canada (AFMC), Black Medical Students Association of Canada (BMSAC), Canadian Federation of Medical Students (CFMS), Canadian Medical Protective Association (CMPA), College of Family Physicians of Canada (CFPC), Federation of Medical Regulatory Authorities of Canada (FMRAC), Indigenous Physicians Association of Canada (IPAC), Medical Council of Canada (MCC), Resident Doctors of Canada (RDOCs), Royal College Physicians and Surgeons of Canada (RCPSC), and the 17 Canadian medical schools, compiling descriptive field notes pertaining to salient discussions as we progressed. Invitations to participate were extended directly from the consensus team to these organizations as well as to individuals in our professional network who are publicly engaged in relevant data science research. Across the consultations and 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, we 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 the series a shared set of goals, benefits, risks, values, recommendations, and principles for governance were distilled.

Literature foundations

We relied on key literature describing the ethical and governance issues in general education and data science.[7,10,11] Regan and Jesse's (2018) framework for contemplating technology and big data in general education was notably relevant. This framework describes ethical concerns associated with information privacy anonymity, surveillance, autonomy, non-discrimination, and ownership of education information[8]. We oriented ourselves to each of these concepts at the outset, using them as a starting point for our pre-workshop consultations.

Pre-workshop consultations

Prior to our series of five pan-Canadian workshops, we engaged in consultations with data custodians, research scientists, and learner advocacy groups from across Canada. This involved independent discussions with representatives of the AFMC, BMSAC, CFMS, CMPA, CFPC, FMRAC, IPAC, MCC, RCPSC, and prominent Canadian data science scholars from multiple disciplines within and beyond medical education.

The workshop series

The workshops were attended by more than 50 individuals including representatives of the AFMC, BMSAC, CFMS, CFPC, FMRAC, MCC, RCPSC, CAPER, and members the education scholarship communities associated with Canada's 17 medical training institutions. Workshops were each three hours in length. The series began by prompting consideration for the ethical concerns and sensitizing considerations determined through the literature and pre-workshop consultations. The first four workshops were augmented by invited presentations from experts 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 had made considerable progress in data sharing within their own field. In each workshop, participants engaged in facilitated activities geared towards the development and refinement of consensus agreement. These activities involved reflective breakout group exercises and large group dialogue. As we progressed, discussions aimed to foster ideation and promote understanding of the perspectives of all stakeholders. Participants were provided orienting concepts and relevant readings that would prepare them for each workshop. Notably, each session was attended by an observer group, composed of representatives of relevant physician trainee advocacy organizations. This group closed each session by reflecting on the workshop's progress and sharing their perspectives during a dedicated roundtable activity. This reflection ensured that a clear orientation pertaining to the trainee perspective were integrated into the final agreement.

Post-workshop consultations

Prior to its final presentation, as a comprehensive form of member checking, draft iterations of this document were reviewed by representatives of the AFMC, BMSAC, CFMS, CFPC, FMRAC, RCPSC, Canadian Association for Medical Education (CAME), Society of Rural Physicians of Canada (SRPC), and Resident Doctors of Canada (R-DOCS). We engaged in a final round synthesis to reconcile and incorporate the final articulation of consensus in response to this feedback.

Outcomes

Sensitizing considerations

Through the pre-workshop consultations, we identified three salient themes regarding inter-institutional data-driven medical education research in Canada:

1. The challenge of balancing institutional mandates. Consultants highlighted tensions that may exist within an organization's own missions (e.g., promoting quality improvement versus protecting constituents from harm), and across the mandates of different institutions (e.g., curricular improvement versus identifying potential risks to patients).
2. The logistical obstacles associated with sharing data across institutions. Consultants pointed to the importance of establishing processes for harmonizing data fields across organizations, building effective data sharing agreements, and conducting research activities within reasonable timelines.
3. The need to differentiate our considerations for admissions and assessment data currently held and used by data stewards from those pertaining to socio-demographic data (e.g., race, gender, ethnicity) that are not currently collected within the context of Canadian medical education. Consultants emphasized strongly that the implications of data exchange shift when socio-demographic data are included. Although critical for the advancement of effective equity, diversity, and inclusivity mandates, many noted that

> the collection and analysis of these data is challenged with the tensions regarding the ethical use and misuse of data.

These initial considerations informed the structure and nature of the workshop series, which gave way to following findings that were ultimately member-checked through post-workshop consultation.

Shared goals and perspectives

Stakeholders and data stewards represented in the consensus-building project included organizations involved in physician training, physician licensure, physician certification, physician regulation, training program accreditation, and advocacy for physicians and physicians-in-training. While each has a separate function, a shared goal was recognized among them:

> *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.*

They also agreed that medical education plays an important role in meeting this goal; that education data has the potential to support knowledge generation that can enhance education practice and policy; and that there exists a potential for data-driven research to contribute to the adoption of education policy and practice that promotes harm and accentuates social inequity.

Benefits of inter-institutional data-driven education research

It was agreed that the benefits of inter-institutional data-driven education include positive impacts on education program evaluation efforts, social accountability missions, and education research quality.

Risks of inter-institutional data-driven education research

It was agreed that the pursuit of the presumed benefits may see data used in ways that expose individuals and institutions to risks. The group agreed upon a set of procedural risks, which include the contravention of institutional commitments to privacy and confidentiality, informed consent, and promoting data sovereignty. The group also agreed upon a set of outcome risks, which include the use of data-driven research to discriminate physicians-in-training and physicians on the basis of personal or social identity characteristics, promote a culture of surveillance, draw harmful misinterpretations of findings, and proliferate inter-institutional comparisons that negatively impact institutions with respect to public perception. These risks were identified as relevant regardless of the intent of the scholarship.

Table 1. Descriptions of the shared values for inter-institutional data-driven education research in Canada

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.

Shared values

The consensus-building process highlighted that a key goal for all participants is to establish trust between stakeholders and researchers. Values that the participating medical education stakeholders and data custodians agreed were fundamental to building this trust include commitments to informed consent, appropriate data collection, appropriate research purposes, privacy and anonymity, autonomy, and non-discrimination (Table 1).

Principles for governance

With the shared goal, perspectives, benefits, risks, and values agreed upon and articulated, we reached consensus on a set of principles for governance when engaging in the collection, sharing, and use of administrative medical education data across institutions for the purposes of research.

1) Researchers should establish clear *a priori* purposes, rationale, and methodology for data sharing. While these can be broad and cover various scholarly activities, the types of associations, linkages, methodologies, and the provisions for protection of confidentiality, privacy, and minimization of harm should be articulated before research commences.

2) Informed consent on the part of those who generate data in education systems including learners, faculty, staff, and organizations is a non-negotiable aspect of research. In the case of meta-informed consent (i.e., pertaining to a range of different potential types of secondary use as determined by a relevant data sharing agreement), this should be time-limited. Time-limited meta-consent provides opportunity to respect the autonomy of data generators with information about risks and benefits resulting from new data created through data linkage as well as to the confirm the continued validity of data in the system.

3) Research oversight bodies should include knowledgeable representation from institutions, communities, and groups which hold meaningful relationships with the data. Notably, the noted risks are not evenly distributed among all populations. Relevant communities with a history of marginalization should be given special consideration in the contemplation of the risks.

4) Governance should be guided by respect for data sovereignty (i.e., the rules/policies/laws of the community from which data are collected). In the Canadian context, this is particularly relevant to notions of Indigenous autonomy from post-colonial states. Representatives with meaningful sovereignty relationships to the data should be engaged in the creation, review, and refinement of data sharing agreements prior to any data sharing. This involvement should be maintained throughout the entire research process, including during the interpretation of results and dissemination of outcomes, ensuring that the research has meaningful benefit and/or does not perpetuate harms for affected communities. This recommendation aligns with those outlined in the Tri-Agency Research Data Management Policy,[12] which contemplates Chapter 9 of the Tri-Council Policy Statement 2 (2022) on Research Involving First nations, Inuit, and Metis Peoples of Canada[13] and the First Nations Principles of Ownership, Control, Access, and Possession.[14]

5) 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 (e.g., program evaluation, research), the risks are the same and still relevant.

6) Governance structures should be vigilant in ensuring data are used 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 established policies for research data management (e.g., Tri-Agency Research Data Management Policy).[12]

Recommendations

On the basis of this consensus, three recommendations for any proposed data-driven research study are offered:

First, a governance body should be established to oversee data sharing agreements and activity. This body should include representation of the perspectives of learners, 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 shared and the research questions pursued. Additionally, this body may provide additional ethical review 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 data sharing that aligns with the rules, policies, and laws of the sovereign jurisdictions in which the data have been collected.

Secondly, a trusted data management facility should be engaged in the technical procedures of data linking, applying the processes 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 should be responsible for quality assurance of the data, reporting to the governance body on safety of data and adherence to protocols. This facility should 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.

Third and lastly, inter-institutional education data should be managed according to best practice standards.[12-14] Data shared for the purposes of analysis should minimize the use of individual and institutional identifiers. 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.

Implications

Inter-institutional data-driven research directed towards the improvement of medical education requires a strong structural foundation. Data sharing creates new opportunities, but also amplifies risks. Given the myriad potential uses of administrative education data and institutional actors potentially involved, a single policy may not be feasible to protect and promote ethical research. Thus, we present here principle-based recommendations that can support data-driven education research in Canada, offering a feasible approach to mitigating the identified risks. While each Canadian data steward and medical education stakeholder has 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, this consensus presents a set of values that can guide institutions and researchers when they work together.

Operationalization at a pan-Canadian level or even in smaller collaborations will be challenging. That is indeed the point of this exercise – data sharing is not an easy task. It is necessary to consider and navigate the challenges that have meaningful impact on stakeholders. This consensus statement represents a starting point from which this type of work may be pursued. We recognize that the context of inquiry will have considerable influence on the way in which the principles and recommendations 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 the core values as they embark on their inquiry. We believe that firm adherence to the principles can guide implementation in each unique context.

One salient idea that emerged during this consensus project is that the risks of inter-institutional data-sharing involving socio-demographic data in the Canadian context are very high. This concern is not surprising given the numerous examples in other contexts but also historical misuses of research data to perpetuate inequity.[15-17] Participants highlighted the importance of these data to address institutional mandates for social accountability and inquiry that advances diversity, inclusivity, and equity. Learner and advocacy group stakeholders noted the continuing potential for harm, stigma, and inquiry that may not benefit those represented by the data. 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. One potential resolution for this tension is to focus inquiry on the impact of education policies and procedures represented in the data rather than on the individuals. Our consensus coalesced on the importance of centering the system.[8]

Conclusion

While this consensus originates in a Canadian context, it has relevance to other medical education communities. The number of medical education data sharing collaborations continue to increase with prominent consortiums and groups in the US, UK, and Canada.[4,18-20] However, formal knowledge on how to develop these collaborations remains limited. A consensus approach that engages all concerned stakeholders is one way to identify the necessary contextual knowledge for collaboration. While some of our recommendations may not be relevant to all contexts, we believe the presented values and principles resonate with good practice for any ethical and impactful inquiry. It is our hope that other collaborations can use this consensus to guide their own research.

Acknowledgements: The authors would like to acknowledge the valuable contributions of our participants, consultants, Tasnia Khan, Muhammadasan Nasser, Saad Chahine, Erin Cameron, Douglas Archibald, Timothy Wood, Peter Wang, Amrit Kirpalani, Claire Touchie, Ilona Bartman, Eleni Katsoulos, Cassandra Barber.
Funding: This work draws on research supported by the Social Sciences and Humanities Research Council; the Education Investigator Award from the Department of Family & Community Medicine and the Temerty Chair in Learner Assessment in the Temerty Faculty of Medicine at the University of Toronto; Department of Family Medicine at McMaster University.
Conflict of Interest: None.
Data Sharing Statement: Information regarding the consensus building process including agendas, workbooks, presentations, and supporting documents are openly available at www.dataconnection.ca in both English and French versions.
Contributors: LG and KK designed and supervised all aspects of the project. BB, RLK, KM, and AY acted as facilitators for the workshop. AC also acted as a facilitator for the workshop series, with special attention to the inputs of the observer group of trainee stakeholders. All authors (LG, KK, BB, RLK, KM, AY, AC) participated in synthesis and interpretation of information gathered through the workshop series, contributed to the critical revision of the paper, approved the final manuscript for publication, and have agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Chahine S, Kulasegaram KM, Wright S, et al. A call to investigate the relationship between education and health outcomes using big data. *Acad Med*. 2018 Jun 1;93(6):829-32. https://doi.org/10.1097/acm.0000000000002217
2. Iobst WF, Sherbino J, Cate OT, et al. Competency-based medical education in postgraduate medical education. *Med Teach.* 2010 Aug 1;32(8):651-6. https://doi.org/10.3109/0142159x.2010.500709
3. Tannebaum D, Kerr J, Konkin J, et al. Triple C Competency-based Curriculum: report on the working group on postgraduate curriculum review – part 1. The College of Family Physicians of Canada. 2011 Mar. Available from https://www.researchgate.net/profile/Allyn-Walsh/publication/316428237_Triple_C_Competency_Based_Curriculum_Report_of_the_Working_Group_on_Postgraduate_Curriculum_Review_Part_1/links/58fd4d670f7e9ba3ba55df79/Triple-C-Competency-Based-Curriculum-Report-of-the-Working-Group-on-Postgraduate-Curriculum-Review-Part-1.pdf. [Accessed May 13, 2022].
4. Grierson LE, Mercuri M, Brailovsky C, et al. Admission factors associated with international medical graduate certification success: a collaborative retrospective review of postgraduate medical education programs in Ontario. *CMAJO.* 2017 Nov 24;5(4):E785-90. https://doi.org/10.9778/cmajo.20170073
5. Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *Jama*. 2007 Sep 5;298(9):993-1001. https://doi.org/10.1001/jama.298.9.993
6. Asch DA, Nicholson S, Srinivas SK, Herrin J, Epstein AJ. How do you deliver a good obstetrician? Outcome-based evaluation of medical education. *Acad Med.* 2014 Jan 1;89(1):24-6. https://doi.org/10.1097/acm.0000000000000067
7. Chen X, Liu CY. Big data ethics in education: connecting practices and ethical awareness. *JETDE*. 2015;8(2):5. https://doi.org/10.18785/jetde.0802.05
8. Regan PM, Jesse J. Ethical challenges of edtech, big data and personalized learning: twenty-first century student sorting and tracking. *Ethics Info Tech*. 2019 Sep;21(3):167-79. https://doi.org/10.1007/s10676-018-9492-2
9. Ellaway RH, Mackay MP, Lee S, et al. The impact of a national competency-based medical education initiative in family medicine. *Acad Med* 2018; 93(12), 1850-1857.
10. Ben-Porath S, Ben Shahar TH. Introduction: big data and education: ethical and moral challenges. *Theory Res Educ*. 2017 Nov;15(3):243-8. https://doi.org/10.1177/1477878517737201
11. Florea D, Florea S. Big Data and the ethical implications of data privacy in higher education research. *Sustainability*. 2020 Jan;12(20):8744. https://doi.org/10.3390/su12208744
12. Government of Canada*. Tri-agency research data management policy.* 2022. Available from https://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html [Accessed on Apr 13, 2022].
13. Government of Canada. *Tri-Council policy statement 2. chapter 9: research involving the First Nations, Inuit and Métis peoples of Canada.* 2022. Available from https://ethics.gc.ca/eng/tcps2-eptc2_2018_chapter9-chapitre9.html. [Accessed on Mar 14, 2023].
14. The First Nations Information Governance Centre. *Ownership, Control, Access and Possession (OCAP™): the path to First Nations information governance.* Ottawa: The First Nations Information Governance Centre, May 2014. 2014 May. Available from https://achh.ca/wp-content/uploads/2018/07/OCAP_FNIGC.pdf [Accessed on May 13, 2022].
15. Braunack-Mayer AJ, Street JM, Tooher R, Feng X, Scharling-Gamba K. Student and staff perspectives on the use of big data in the tertiary education sector: a scoping review and reflection on the ethical issues. *Rev Educ Res*. 2020 Dec;90(6):788-823. https://doi.org/10.3102/0034654320960213
16. Molldrem S, Smith AK. Reassessing the ethics of molecular HIV surveillance in the era of cluster detection and response: Toward HIV data justice*. Am J Bioeth*. 2020 Oct 2;20(10):10-23. https://doi.org/10.1080/15265161.2020.1806373.
17. Wesson P, Hswen Y, Valdes G, Stojanovski K, Handley MA. Risks and Opportunities to Ensure Equity in the Application of Big Data Research in Public Health. *Annu Rev Public Health*. 2022

Apr 5;43:59-78. https://doi.org/10.1146/annurev-publhealth-051920-110928

18. Dowell J, Cleland J, Fitzpatrick S, et al. The UK medical education database (UKMED) what is it? Why and how might you use it?. *BMC Med Educ.* 2018 Dec;18(1):1-8. https://doi.org/10.1186/s12909-017-1115-9
19. Gale TC, Lambe PJ, Roberts MJ. Factors associated with junior doctors' decisions to apply for general practice training programmes in the UK: secondary analysis of data from the UKMED project. *BMC Med*. 2017 Dec;15(1):1-4. https://doi.org/10.1186/s12916-017-0982-6
20. Jerant A, Fancher T, Fenton JJ, et al. How medical school applicant race, ethnicity, and socioeconomic status relate to multiple mini-interview–based admissions outcomes: findings from one medical school. *Acad Med*. 2015 Dec 1;90(12):1667-74. https://doi.org/10.1097/acm.0000000000000766.