On the Importance of Centralized Learning in Health Research

Michael Wolfson[1]

[1]School of Epidemiology and Public Health, Faculty of Medicine, and Centre for Health Law, Policy and Ethics. Faculty of Law, University of Ottawa, Ottawa, Ontario, Canada

Abstract

Canada has potentially some of the best data in the world for a wide range of health research. Unfortunately, though, much of it is confined to data silos. Perhaps the most important barrier is the inability to analyze already computerized individual-level health care encounter data from multiple provincial jurisdictions, in part to respect privacy, but also related to concerns about loss of control over data use and publication of sensitive or controversial findings. A first level of response is the use of meta-analysis (MA), as exemplified by the work of the Canadian Network for Observational Drug Effect Studies (CNODES). This kind of MA involves essentially one iteration where separate statistical analyses are done node by node, and then the non-confidential results are aggregated to produce overall estimates. A more sophisticated response is federated learning (FL) building on the DataSHIELD idea of iterations between a central server with data remaining distributed across nodes. The iterations in FL are intended to transmit only non-confidential data. Finally, the "oracle" or gold standard is centralized learning (CL) where the analysis uses all the individual-level personal health information pooled from all the nodes. There has been a great deal of research devoted to improving FL methods, but relatively little on the challenges of MA and FL approaches compared to CL. In this paper, some of these challenges are explored using a Statistics Canada health survey. In conclusion, an option for increasing the feasibility of CL in Canada is described.

Keywords

health research, federated learning, meta-analysis, privacy barriers

1. Introduction

With the dramatically growing power of various AI methods in health research, a crucial impediment is access to data. This is especially important for individual-level data including those which have been linked longitudinally across different health care encounters. But this personal health information (PHI) is sensitive and raises major privacy concerns. As a result, important health research is often stymied by inability to pool PHI from different data holding organizations.

A notable example in Canada has been the Canadian Network for Observational Drug Effect Studies [1]. In this case, the objective is to use routinely collected health care encounter administrative data to search for adverse drug reactions (ADRs). Because the provinces would not allow their PHI to cross provincial boundaries, the "second best" method of meta-analysis (MA) was used. Researchers agreed on a specific drug to study, and the specification of a logistic regression drawing on regressors in each province that were as harmonized as possible. High-dimensional propensity score (hdPS) matching [2, 3] was used to create the controls for a case-control analysis. However, these statistical analyses had to be done separately within each participating province, and only the final non-confidential node-specific results could be shared outside each province. These results were then aggregated using standard MA methods.

An important concern in this context is that compared to a centralized learning (CL) approach, where the data from all the participating provinces are first pooled into a single data base and then the search for the ADRs undertaken, the MA approach could miss or inappropriately find an ADR. Given

Joint Proceedings of the AIME 2026 Workshops: 1st International Workshop on Multicentric and Privacy-preserving Learning in Healthcare, Foundation Models for Public Health and Epidemiology: From Promise to Practice, and First International Workshop on Knowledge Graphs for Health, July 10, 2026, Ottawa, Canada

mwolfson@uottawa.ca (M. Wolfson)

0000-0003-1941-955X (M. Wolfson)

widespread evidence of over-prescribing [4], stronger and more focused evidence on ADRs would be most relevant.

Unfortunately, comparisons of CNODES type MA with CL are infeasible in Canada due to current provincial restrictions, motivated in part by privacy concerns, which prevent pooling of the required PHI subsets of variables from each of the provinces. Instead, to assess the concern that CNODES-style MA may produce incorrect or biased results, this analysis draws on an already pooled data set. The question is whether it is possible to find a plausible example where CNODES-style MA results differ materially from the CL results. If so, it would constitute a "proof by construction" of concerns that MA methods can yield misleading or incorrect results.

Leaving aside MA, there are federated learning (FL) methods such as DataSHIELD [5, 6] that do not require PHI to cross provincial data boundaries, and can in many cases exactly reproduce the statistical results that would be found using CL. However, even to the extent that such FL methods can work in principle, FL still raises other concerns regarding its practical implementation, as discussed further below.

2. Data

To construct an appropriate example, Statistics Canada's 2008/2009 Canadian Community Health Survey (CCHS) public use microdata file on Healthy Ageing has been used [7]. To create the analogue of the different provincial data silos encountered in the CNODES studies, this CCHS household survey sample was artificially partitioned horizontally into 12 sub-samples = 6 regions (combining the four Atlantic provinces and the two prairie provinces) and whether or not the survey respondent lived in a census metropolitan area (CMA), resulting in 12 disjoint geographic regions (reg-s). While straightforward, this partition was chosen to ensure reasonable sample sizes within each region, and to increase the likelihood of significant heterogeneity across the reg-s.

3. Random Forest

A well-known major concern with both MA and FL is heterogeneity. One kind is where the data elements in each province, or more generally data silo or node, are not defined or measured in exactly the same way. This kind of heterogeneity is avoided here with the use of the CCHS, as all the data elements are identical across the artificially partitioned geographies. Another kind of heterogeneity is where the "shapes" of the various multivariate joint distributions of key parameters (e.g. correlations, skewness) vary across nodes. This is possible both in real world analyses and in the CCHS data used here.

As an initial exploration of such heterogeneity, a random forest was estimated from the CCHS data both for the pooled data and then for each of the 12 reg-s. The dependent variable (DV) is whether the individual was taking medicine for pain daily. The independent variables (IVs) included the 12 reg-s, 5 age groups, 8 levels of life satisfaction, and binary variables for race, having chronic diseases, mobility impairments, being able to work, emotional problems, and sleep problems.

This specific set of dependent and independent variables is not intended to constitute a full exploration of the factors most closely associated with taking pain medications daily. Rather it is sufficient for this analysis for them to be plausibly illustrative.

Figure 1 shows the results of 13 random forests, one for each of the 12 reg-s individually, and then one for all the sample data when pooled. The spreads of the rank orders of these predictors clearly show the heterogeneity, not only in the range of importance rankings across the 12 reg-s, but also by their asymmetry compared to the pooled importance rank.

These random forest results provide an important caution for subsequent regression analyses. For some variables there are important urban-rural differences, and the evident asymmetries in ranks indicate that the predictive roles of the covariates vary substantially by geographic region. It is therefore an important diagnostic showing that predictor importance is not stable across the artificial CCHS regions.

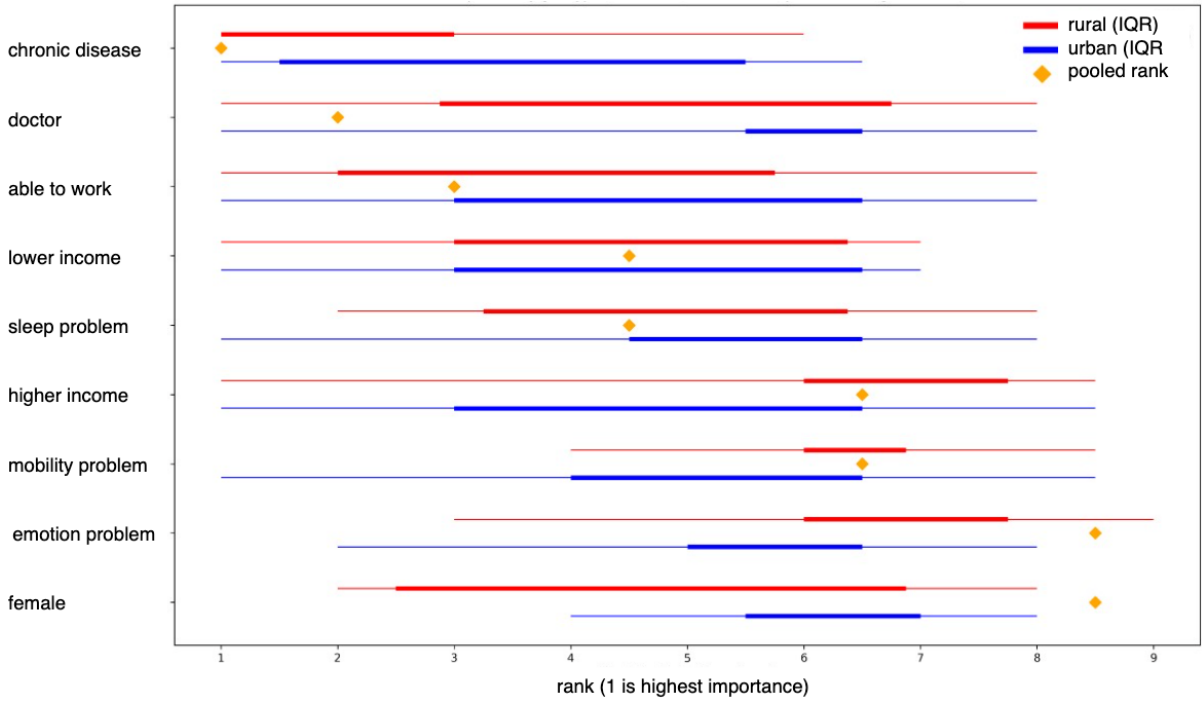

Figure 1: Illustration of geographic heterogeneity showing ranges and inter-quartile ranges (IQRs) for selected CCHS variables from random forest analyses for each of 12 data nodes horizontally partitioned from the CCHS sample, plus a CL random forest using the fully pooled sample. In all cases, daily pain medication use was the dependent variable.

This instability motivates regression analyses where region fixed effects, random intercepts, and random slopes are considered to represent (in this case) geographic heterogeneity.

4. Regressions

We turn now to an exploration of the same CCHS data but now using regression. Recall that our objective is not finding a definitive result for a given statistical question. Rather it is to demonstrate, proof by construction, that there can be cases where MA and FL provide misleading or incorrect results compared to CL.

Studies with objectives like CNODES typically use logistic regressions. As an analogue for this kind of analysis, the CCHS unweighted sample data have here been used to explore the prediction of the daily use of at least two different medications (unweighted 38% of the age 45+ population covered by the survey). The potentially predictive independent variables (IVs) included: whether living alone, 5-year age groups, having a regular doctor, race, and sex. Again, these variables have been chosen to be plausibly illustrative to construct the counterexample, not necessarily to be definitive.

For MA, there is a variety of methods of increasing complexity. A critical aspect is just what information from the individual regressions in each node needs to be communicated to the central analytical group. The simplest is the set of regression coefficients and their standard errors estimated separately at each data node.

More complex MA analyses require each node additionally to compute matrices of coefficients that enable the central analysis group to take one step of a Newton–Raphson iteration. These extra data from each node enable a generally closer approximation to what would have been found had all highly sensitive individual-level data from across the nodes first been pooled and then used to estimate a single regression, i.e. a CL approach.

However, beyond a certain point, the extra data output from each node in more complex MA methods raises non-trivial privacy concerns. As a result, the focus here initially is on a straightforward logistic

regression (M1) which by construction is feasibly estimated with the common simpler versions of MA, hence raises no privacy concerns. Since M1 can be exactly reproduced by DataSHIELD and similar FL methods, even though it is estimated on the pooled data, i.e. in a CL manner, it is treated for purposes of discussion as if it were estimated by FL.

Still, the M1 regression specification may not adequately address this risk of significant heterogeneity across the nodes. As a result, two further regressions have been estimated, both using GLMM fitted using JASP by maximum likelihood with a Bernoulli family and logit link, and with the identical DV and IVs as in M1. M2 is estimated with random intercepts only, and M3 with both random intercepts and random slopes for race.

The regression results for these three models are shown in Tables 1 and 2.

Table 1
Harmonized fixed effects: odds ratios (with 95% confidence intervals) and p values, as well as fit statistics from three unweighted regression models; dependent variable is daily use of two or more medications; region effects in M1 are fixed effects compared to the reference; region effects in M2 and M3 are random intercept deviations for intercepts; and in M3 also with random slopes for race. N = 30,458, 12 regions with 407 excluded observations due to missing data.

	M1 GLM, region fixed effects		M2 GLMM, random intercept		M3 GLMM, uncorrelated random intercept + random race slope	
Predictor comparison	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p
age 45–49	reference		reference		reference	
age 50–54 vs ref	1.58 (1.38–1.80)	< 0.001	1.58 (1.38–1.80)	< 0.001	1.58 (1.39–1.81)	< 0.001
age 55–59 vs ref	2.12 (1.88–2.38)	< 0.001	2.12 (1.88–2.38)	< 0.001	2.13 (1.89–2.39)	< 0.001
age 60–64 vs ref	2.88 (2.56–3.24)	< 0.001	2.88 (2.56–3.24)	< 0.001	2.88 (2.56–3.25)	< 0.001
age 65–69 vs ref	3.91 (3.47–4.41)	< 0.001	3.91 (3.46–4.40)	< 0.001	3.93 (3.48–4.43)	< 0.001
age 70–74 vs ref	4.90 (4.32–5.56)	< 0.001	4.89 (4.32–5.55)	< 0.001	4.93 (4.34–5.59)	< 0.001
age 75–79 vs ref	5.77 (5.08–6.55)	< 0.001	5.77 (5.08–6.55)	< 0.001	5.82 (5.12–6.61)	< 0.001
age 80–84 vs ref	6.34 (5.55–7.25)	< 0.001	6.33 (5.54–7.24)	< 0.001	6.36 (5.56–7.27)	< 0.001
age 85+ vs ref	6.37 (5.64–7.21)	< 0.001	6.36 (5.63–7.20)	< 0.001	6.42 (5.67–7.26)	< 0.001
lives alone vs not	1.04 (0.98–1.09)	0.191	1.04 (0.98–1.09)	0.194	1.03 (0.98–1.09)	0.237
female vs male	1.21 (1.16–1.27)	< 0.001	1.21 (1.16–1.27)	< 0.001	1.21 (1.16–1.27)	< 0.001
regular doctor vs not	3.32 (2.95–3.74)	< 0.001	3.32 (2.95–3.74)	< 0.001	3.35 (2.97–3.77)	< 0.001
white vs non-white	1.36 (1.25–1.48)	< 0.001	1.37 (1.26–1.49)	< 0.001	**1.15 (0.96–1.38)**	**0.133**
AIC	38,732		≈ 38,760*		≈ 38,740*	
BIC	38,932		≈ 38,880*		≈ 38,860*	
regional random-intercept SD, log-odds scale	—		0.188		0.391	
regional random race-slope SD, log-odds scale	—		—		0.260	

*JASP output was rounded to three digits.

By AIC and BIC, M3 fits better than M2, though the fit is not clearly better in M2 compared to M1, or M3 compared to M1. Most notably, in M1 and M2, race is highly significant. However, in M3, the average fixed race coefficient is no longer statistically distinguishable from zero once regional variation in the race slope is allowed. This set of regressions provides evidence against imposing a single common race coefficient across regions. The relationship between race and use of medications varies importantly across geographic regions.

Recall that the intent here is to be illustrative; it is not to produce the best inferences about the main factors affecting daily medicine use, the DV. Still, this kind of result could be very seriously misleading in clinical or health research. An effect is present in the simpler analysis but absent or substantially more nuanced in a CL "oracle" or gold standard regression.

M1 was estimated here using pooled data, but because it is a standard common-coefficient GLM, it is also the kind of model that DataSHIELD-style iterative FL can reproduce without pooling each node's

Table 2
Common and region-specific race associations. M1 and M2 impose a common white vs non-white OR across regions. M3 combines the fixed race coefficient with the region-specific random race-slope deviation to produce region-specific race ORs. Region labels combine province or region abbreviations with "c" for CMA and "n" for non-CMA. M1 regional effects are expressed relative to the reference region abc. M2 and M3 regional effects are partially pooled deviations from the model-wide mean, so no observed region serves as the reference category in the GLMMs.

Region	M1 GLM, region fixed effects, common race OR	M2 GLMM, random intercept, common race OR	M3 GLMM, uncorrelated random intercept + random race slope, region-specific race OR
	1.36	1.37	—
abc	—	—	0.808
abn	—	—	0.995
atc	—	—	1.318
atn	—	—	1.529
bcc	—	—	0.456
bcn	—	—	0.895
onc	—	—	0.598
onn	—	—	1.654
prc	—	—	0.995
prn	—	—	1.343
quc	—	—	0.924
qun	—	—	1.250

individual-level records. M3, in contrast, requires a model class that allows the race association itself to vary across regions. That model is not available to simple CNODES-style MA based only on node-specific coefficients and standard errors, nor to standard GLM-only distributed methods. It requires either CL or, potentially, a more specialized federated GLMM algorithm [8]. Thus, the contrast between M1/M2 and M3 is best interpreted as an empirical counterexample to the adequacy of simple common-effect distributed analyses. Importantly, it illustrates a critical limitation of treating FL reproducibility of a simpler CL model as sufficient. An FL method may exactly reproduce a common-effect GLM while still failing to reproduce the richer CL analysis needed to capture substantively important heterogeneity.

5. Discussion

Tremendous efforts are being devoted to the development of algorithms and statistical methods to surmount the problems of so much valuable health data residing in data silos [9]. This is understandable on the one hand because of the major expected benefits from health research and analysis[1] that can access ever larger and more richly detailed individual-level health data. On the other hand, pervasive "privacy chill" has been stymying the assembly of these data in centralized data repositories for "centralized learning" (CL).

The decentralized non-CL methods for statistical regression analyses can be divided into two broad groups: those where there is only one iteration of data flow from the decentralized nodes = meta-analysis (MA), and those where some sort of iterative process is required, with (intended to be) non-confidential data (e.g. coefficients, covariance matrices, Hessians) flowing back and forth between the nodes and the central analytical server = federated learning (FL). Unfortunately, both the MA and FL approaches are problematic [10], and for a wide variety of possible reasons [11, 12].

As shown with the illustrative regressions in the previous section, exact FL reproduction of a pooled GLM is not enough if the pooled GLM is mis-specified, in this case by imposing an empirically incorrect common-effect structure. The common race coefficient in M1/M2 is significant, while the average fixed race coefficient in M3 is attenuated and non-significant once regional race-slope heterogeneity

[1]While a distinction is often drawn between research and quality assurance, including in some provinces' privacy legislation, this is unfortunate. From a statistical perspective, the analyses are often identical. Therefore, in this discussion, we consider analyses related to health care quality as included in health research and analysis.

is allowed. Distributed methods should not only be judged by whether they reproduce a chosen CL model, but also by whether they can reproduce the right kind of CL model for the given data.

Further, even though there is evidence that highly sophisticated versions of FL may be able to produce results equivalent to those with CL [8], an open question is whether analogously misleading results are possible with FL, depending on the ability of the specific FL method to handle regression specifications that appropriately capture data heterogeneity. The regression results above show that exact reproduction of simpler GLM and GLMM specifications may still yield a substantively misleading result if the relevant heterogeneity requires a richer model class.

Beyond the regressions shown above, there is interest in a broader range of federated analyses. It is unclear whether the hdPS matching step used in many distributed drug-safety studies can be implemented in a privacy-preserving FL workflow without additional disclosure risks or substantial protocol-specific engineering. Advanced, problem-specific FL can reduce the gap between MA and CL, but that only reinforces the point that simple MA is not a general substitute for CL and that the validity of FL depends on the specific algorithm and estimand [13, 3].

Similar questions exist for recursive partitioning methods. Federated random-forest approaches have been proposed and can perform comparably to centralized RFs in some settings, but this does not imply that they reproduce the same pooled-data RF, split structure, or variable-importance rankings [14].

Similar questions arise for distributed neural-network methods, especially in medical imaging, where FL is now widely studied but remains sensitive to site heterogeneity, communications burden, privacy leakage, and validation constraints [15, 16].

In addition to these technical and statistical issues, FL suffers from a different problem. For the foreseeable future, it appears unlikely that Canadian provinces will each host a server that (a) holds a massive amount of linked highly sensitive albeit anonymized[2] individual-level health care encounter data, (b) has installed software that can read, analyze, and produce non-private statistical results (i.e. intended-to-be non-disclosive statistical summaries) from these sensitive data, and (c) can "talk" to a central analytical server over the internet by iteratively sending and receiving intermediate non-private statistical results.

In Canada, research granting councils have provided over $200 million to two major longitudinal surveys, one on aging [17] and one on cancer [18]. In both surveys, an overwhelming majority of respondents have provided explicit consent to link their survey responses, as well as physical measures and bio-specimen data, to their provincial health care administrative data. While these linkages have been done within provinces, no province has allowed their linked administrative data to leave their jurisdiction, even though all the other respondents' data have been pooled for research and statistical analysis purposes.

These provincial restrictions have stymied fully CL approaches. In lieu, the Canadian Institutes of Health Research and partners have invested substantially in the Health Data Research Network (HDRN Canada) and the Strategic Patient-Oriented Research (SPOR) Canadian Data Platform, which support multi-jurisdictional data access, distributed analysis, and related infrastructure [19, 20]. Tens of millions have been invested which, among other initiatives and in the face of continuing provincial resistance to any kind of data pooling, have been promoting FL approaches for the analysis of these and other very rich PHI data Very important progress has been made in terms of prototypes.

More recently, the Canadian government, in its AI Strategy [21] has announced $100 million for CIHI in partnership "to launch the Health Sector Data Space, to link secure, private and standardized datasets to strengthen clinical trials, health services research and performance measurements. Canada will also invest $100 million to expand VITAL[3] in five additional provinces, supporting its ability to

[2]Note that anonymized versions of PHI may still be identifiable. Rendering PHI data non-identifiable would usually make it useless for more in-depth statistical analyses such as CNODES.

[3]"Canada's universal healthcare system is one of the richest sources of clinical data in the world. Every hospital encounter generates hundreds of thousands of data points—imaging, medication records, patient vitals, clinical outcomes—that could fuel breakthroughs in how we predict, prevent, and treat disease. But that data has been locked away in institutional silos, inaccessible to the researchers and entrepreneurs who could put it to work for Canadians. VITAL is building the infrastructure to change that. As a pan-Canadian health data platform designed to connect clinical data from hospitals across multiple

leverage clinical data from hospitals and enable health data innovation that reduces mortality rates and accelerates critical care for Canadians." [21, p. 35].

However, for both of these major investments, no CL is envisaged. Rather the intent, to the extent that the details have been published, will only involve MA and FL, with the phrase "federated access" as the watchword.

Of course, these investments signal major progress. When operational, they will enable tremendous progress in generating value from Canada's largely untapped wealth of PHI. However, as shown above, being limited only to MA and FL methods risks generating biased or incorrect results. It is therefore essential for there to be a venue where, from time to time, important FL analytical results can be checked against the gold standard of CL results.

A major possible Canadian venue for CL is Statistics Canada. In addition to the population census, cancer registry, and individual income tax records, Statistics Canada has decades of experience collecting highly sensitive individual-level health data along with linkable relevant microdata. It has both constitutional and legislative authority. Statistics Canada further has a very strong cultural and statutory framework for confidentiality, as well as world class expertise in secure data linkage and statistical methodology.

However, Statistics Canada is not yet able to support adequately the kinds and scales of health research and analysis needed from a CL perspective. It does not have the systems for enabling authorized researchers from outside the organization to access and analyze directly these microdata, though their virtual data lab (vDL) software and policies have the needed potential. There are also important funding and cost-sharing questions involving not least relationships with Canada's research granting councils.[4]

There is further a major challenge in federal-provincial relationships. The federal Statistics Act gives Statistics Canada broad authority to access and collect existing records held by governments and organizations anywhere in Canada for statistical purposes, subject to the Act's mandate and confidentiality framework. However, the Constitution grants much of the jurisdiction for health care to the provinces. The provinces guard their jurisdiction jealously. They are also afraid of losing control over how their data might be used, especially the possibility of embarrassing results becoming public. In this area, it is essential for a climate of mutual trust to be developed. It is beyond the scope of this paper to discuss the possible institutional mechanisms that could serve this objective.

6. Conclusion

This paper has focused on health research and analysis drawing on sensitive individual-level health data that are dispersed across separate data silos. Three broad kinds of statistical analysis have been briefly explored: meta-analysis (MA), federated learning (FL), and centralized learning (CL). The analysis has demonstrated serious limitations with CNODES-style MA, particularly when important between-node heterogeneity is present, though this is the only distributed inter-provincial method in actual use. To provide confidence in CNODES-style analyses, it is critical that at least some are replicated and explored using CL.

It is also argued that for pan-Canadian research and analysis utilizing provincial PHI, even though FL has tremendous potential, its practical implementation is unlikely for the foreseeable future. The CL gold standard, in contrast, is well-advanced at Statistics Canada, and has great potential. However, CL in Canada cannot be fully realized without significant changes in funding and institutional arrangements.

provinces, it is enabling AI-driven discovery: supporting research that can generate leads for novel therapies, attracting new clinical trials, and identifying genetic risk factors with greater precision." [21, p. 36]

[4]The Canadian Institute for Health Information (CIHI) is another possible venue. It has the advantage, compared to Statistics Canada, of board governance that includes senior provincial members. However, it has weaker privacy legislative protection than Statistics Canada, and is unable to link its PHI to an important range of other data including the population census (especially socio-economic characteristics) and various Statistics Canada surveys including on health and disability.

Acknowledgments

All the statistical analysis was done by the author using the JASP statistical package. Extensive use was made of ChatGPT in discussing statistical concepts and exploring the published literature. The papers cited in the references were all read by the author.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] S. Suissa, D. Henry, P. Caetano, C. R. Dormuth, P. Ernst, B. R. Hemmelgarn, J. LeLorier, A. R. Levy, P. J. Martens, J. M. Paterson, R. W. Platt, I. S. Sketris, G. F. Teare, Canadian Network for Observational Drug Effect Studies, CNODES: The canadian network for observational drug effect studies, Open Medicine 6 (2012) e134–e140.

[2] S. Schneeweiss, J. A. Rassen, R. J. Glynn, J. Avorn, H. Mogun, M. A. Brookhart, High-dimensional propensity score adjustment in studies of treatment effects using health care claims data, Epidemiology 20 (2009) 512–522. doi:10.1097/EDE.0b013e3181a663cc.

[3] J. A. Rassen, S. Schneeweiss, Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system, Pharmacoepidemiology and Drug Safety 21 (2012) 41–49. doi:10.1002/pds.2328.

[4] S. G. Morgan, J. Hunt, J. Rioux, J. Proulx, D. Weymann, C. Tannenbaum, Frequency and cost of potentially inappropriate prescribing for older adults: A cross-sectional study, CMAJ Open 4 (2016) E346–E351. doi:10.9778/cmajo.20150131.

[5] A. Gaye, Y. Marcon, J. Isaeva, P. LaFlamme, A. Turner, E. M. Jones, J. Minion, A. W. Boyd, C. J. Newby, M. L. Nuotio, R. Wilson, et al., DataSHIELD: Taking the analysis to the data, not the data to the analysis, International Journal of Epidemiology 43 (2014) 1929–1944. doi:10.1093/ije/dyu188.

[6] M. Wolfson, S. E. Wallace, N. Masca, G. Rowe, N. A. Sheehan, V. Ferretti, P. LaFlamme, M. D. Tobin, J. Macleod, J. Little, I. Fortier, DataSHIELD: Resolving a conflict in contemporary bioscience—performing a pooled analysis of individual-level data without sharing the data, International Journal of Epidemiology 39 (2010) 1372–1382. doi:10.1093/ije/dyq111.

[7] Statistics Canada, Canadian community health survey, 2008/2009: Public use microdata file, Statistics Canada, Ottawa, 2009. URL: https://www150.statcan.gc.ca/n1/en/catalogue/82M0015X, accessed February 17, 2026.

[8] Z. Yan, K. S. Zachrison, L. H. Schwamm, J. J. Estrada, R. Duan, A privacy-preserving and computation-efficient federated algorithm for generalized linear mixed models to analyze correlated electronic health records data, PLoS ONE 18 (2023) e0280192. doi:10.1371/journal.pone.0280192.

[9] F. Camirand Lemyre, S. Lévesque, M. P. Domingue, K. Herrmann, J. F. Ethier, Distributed statistical analyses: A scoping review and examples of operational frameworks adapted to health analytics, JMIR Medical Informatics 12 (2024) e53622. doi:10.2196/53622.

[10] P. S. W. Boedhoe, M. W. Heymans, L. Schmaal, Y. Abe, P. Alonso, S. H. Ameis, et al., An empirical comparison of meta- and mega-analysis with data from the ENIGMA obsessive-compulsive disorder working group, Frontiers in Neuroinformatics 12 (2019) 102. doi:10.3389/fninf.2018.00102.

[11] D. L. Burke, J. Ensor, R. D. Riley, Meta-analysis using individual participant data: One-stage and two-stage approaches, and why they may differ, Statistics in Medicine 36 (2017) 855–875. doi:10.1002/sim.7141.

[12] J. Bhanbhro, S. Nisticò, L. Palopoli, Issues in federated learning: Some experiments and preliminary results, Scientific Reports 14 (2024) 29881. doi:10.1038/s41598-024-81732-0.

[13] H. Li, C. Zang, Z. Xu, W. Pan, S. Rajendran, Y. Chen, F. Wang, Federated target trial emulation

using distributed observational data for treatment effect estimation, npj Digital Medicine 8 (2025) 387. doi:10.1038/s41746-025-01803-y.

[14] A. C. Hauschild, M. Lemanczyk, J. Matschinske, T. Frisch, O. Zolotareva, A. Holzinger, J. Baumbach, D. Heider, Federated random forests can improve local performance of predictive models for various healthcare applications, Bioinformatics 38 (2022) 2278–2286. doi:10.1093/bioinformatics/btac065.

[15] H. Guan, P.-T. Yap, A. Bozoki, M. Liu, Federated learning for medical image analysis: A survey, Pattern Recognition 151 (2024) 110424. doi:10.1016/j.patcog.2024.110424.

[16] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, M. J. Cardoso, The future of digital health with federated learning, npj Digital Medicine 3 (2020) 119. doi:10.1038/s41746-020-00323-1.

[17] P. Raina, C. Wolfson, S. Kirkland, L. E. Griffith, C. Balion, B. Cossette, I. Dionne, S. Hofer, D. B. Hogan, E. R. van den Heuvel, T. Liu-Ambrose, V. Menec, G. Mugford, C. Patterson, H. Payette, B. Richards, H. Shannon, D. Sheets, V. Taler, M. Thompson, H. Tuokko, A. Wister, L. Young, Cohort profile: The canadian longitudinal study on aging (CLSA), International Journal of Epidemiology 48 (2019) 1752–1753j. doi:10.1093/ije/dyz173.

[18] T. J. B. Dummer, P. Awadalla, C. Boileau, C. Craig, I. Fortier, V. Goel, J. M. T. Hicks, S. Jacquemont, B. M. Knoppers, N. Le, T. McDonald, J. McLaughlin, A.-M. Mes-Masson, L. J. Palmer, L. Parker, M. Purdue, P. J. Robson, J. J. Spinelli, C. Syme, E. White, M. Borugian, The canadian partnership for tomorrow project: A pan-canadian platform for research on chronic disease prevention, CMAJ 190 (2018) E710–E717. doi:10.1503/cmaj.170292.

[19] L. T. Dahl, K. McGrail, X. Wang, M. R. Law, D. Buckeridge, A. Sweetman, T. McDonald, J. F. Ethier, Y. Fortin, A. Guttmann, The SPOR-canadian data platform: A national initiative to facilitate data-rich multi-jurisdictional research, International Journal of Population Data Science 5 (2020) 1374. doi:10.23889/ijpds.v5i5.1374.

[20] Health Data Research Network Canada, SPOR canadian data platform, HDRN Canada website, 2026. URL: https://www.hdrn.ca/en/initiatives/spor-cdp/, accessed July 31, 2026.

[21] Innovation, Science and Economic Development Canada, Canada's national artificial intelligence strategy: AI for all, Government of Canada, 2026. URL: https://www.ised-isde.canada.ca/site/ised/en/canadas-national-artificial-intelligence-strategy-ai-all, accessed July 31, 2026.