

Recognition of inherent biases in administrative data

Alan Katz MBChB MSc CCFP FCFP Amy Freier PhD

Administrative data are valuable tools for primary care research. Using routinely collected data avoids the burden of data collection for both researchers and research subjects, making many studies more feasible and pragmatic. Additionally, using administrative data can help avoid biases inherent in survey research, including recall bias, response bias, social acceptability bias, and sampling bias.¹ Administrative billing and dispensing data can be used to explore services provided in the health care system from both provider and user perspectives. Understanding the health of the population and how the population uses the health care system provides essential information for health system planning and policy development.

While whole population studies avoid some biases, they introduce others. This does not negate the value of administrative data research, but it is imperative that we understand these biases and implement appropriate responses that consider the importance of inclusion, diversity, equity, and accessibility in research.

Sources of bias in administrative data

Administrative data are usually collected routinely by hospitals, pharmacies, laboratories, and health authorities, which are products of institutional oppression and colonialism, and thus the data reflect the structural biases inherent in those systems. To understand these data fully, we need to understand their provenance. While one dictionary initially defines *provenance* as “source,” its second definition adds depth to our understanding and its use in this context: “the history of ownership of a valued object or work of art or literature.”² Using the word provenance suggests we need to know not just the source of data, but also the circumstances of their collection. In Canada, understanding the history of collection and ownership of data is particularly important considering past and ongoing colonial violence, which is itself embedded in the health care system.


Population-based analyses using health system data include in the numerator only those who have accessed the health care system. While financial barriers to care are less common in countries such as Canada that have universal health care coverage, these barriers still exist. There might be costs associated with transportation, child care, or lack of benefits such as sick leave at work, or even with the price of medication. Access is particularly challenging in many rural and remote communities owing to the distances people must travel for care. Accessibility challenges are not limited to physical

accessibility and include difficulties in accessing information and other resources. For example, language translation services are not routinely available at most health care sites. All of these factors influence the circumstances of data collection and embed systemic biases in data that are not initially apparent.

Further, Canadians who have had negative experiences when seeking care—such as racism, gender discrimination, ableism, and other demeaning attitudes at points of care—might not seek care as often or as quickly. These negative experiences and resultant mistrust of the health care system are of particular relevance to Indigenous peoples.^{3,4} Some biases are systemic owing to provider ignorance and attitude. When providers do not understand the cultural norms of patients, they might have communication challenges that lead to misunderstandings or failure to make correct diagnoses. We cannot assume that because a health service was accessed that the service was appropriate and will lead to the desired health outcome⁵; this disconnect is especially concerning with respect to mental illness, as both communication and cultural awareness are crucial components of the diagnostic process.

Implications for research using administrative data

Although some see these limitations as fatal flaws,⁶ research using administrative data still provides critical information. It is the responsibility of the research team to acknowledge and mitigate these limitations, just as they must do with other data sources. Seeking other data sources to confirm the findings and comparing the results with those from studies using other methods are appropriate mitigations to use where possible.^{7,8} Key differences may result from limitations in 1 or more of the studies being compared, and it is important to explore how those limitations might have affected results. Using within-group analysis or identifying different comparison groups might also help bring intersectional inequities to light.⁹

Mitigating the impact of biases in administrative data can mean including people with the greatest risk of marginalization and discrimination as members of the research team, training researchers to recognize inequities and their own potential biases, and creating data systems and processes that are inclusive, diverse, equitable, and accessible. The aphorism “the results speak for themselves” does not recognize the critical importance of ensuring a valid and culturally appropriate interpretation of results that accounts for biases. 

Dr Alan Katz is Professor in the Department of Family Medicine and the Department of Community Health Sciences at the University of Manitoba in Winnipeg. **Dr Amy Freier** is Research Associate at the Manitoba Centre for Health Policy at the University of Manitoba.

Competing interests

None declared

References

1. Choi BCK, Pak AWP. A catalog of biases in questionnaires. *Prev Chronic Dis* 2005;2(1):A13. Epub 2004 Dec 15.
2. Merriam-Webster.com Dictionary. *Provenance*. Springfield, MA: Merriam-Webster Inc. Available from: <https://www.merriam-webster.com/dictionary/provenance>. Accessed 2021 Dec 22.
3. Nelson SE, Wilson K. Understanding barriers to health care access through cultural safety and ethical space: Indigenous people's experiences in Prince George, Canada. *Soc Sci Med* 2018;218:21-7. Epub 2018 Sep 15.
4. McCallum MJL, Perry A. *Structures of indifference. An Indigenous life and death in a Canadian city*. Winnipeg, MB: University of Manitoba Press; 2018.
5. Lavoie JG, Kaufert J, Browne AJ, O'Neil JD. Managing matajoosh: determinants of First Nations' cancer care decisions. *BMC Health Serv Res* 2016;16(1):402.
6. McLemore MR. Reimagining methodological considerations for research studies using 'big' administrative data sets. *Paediatr Perinat Epidemiol* 2021;35(4):491-2.
7. Hayward A, Wodtke L, Craft A, Robin T, Smylie J, McConkey S, et al. Addressing the need for Indigenous and decolonized quantitative research methods in Canada. *SSM Popul Health* 2021;15:100899.
8. Hardeman RR, Karbeah J. Examining racism in health services research: a disciplinary self-critique. *Health Serv Res* 2020;55(Suppl 2):777-80.
9. Bauer GR, Scheim AI. Methods for analytic intercategory intersectionality in quantitative research: discrimination as a mediator of health inequalities. *Soc Sci Med* 2019;226:236-45. Epub 2019 Jan 21.

Can Fam Physician 2022;68:933-4. DOI: 10.46747/cfp.6812933

Hypothesis is a quarterly series in *Canadian Family Physician (CFP)*, coordinated by the Section of Researchers of the College of Family Physicians of Canada. The goal is to explore clinically relevant research concepts for all *CFP* readers. Submissions are invited from researchers and nonresearchers. Ideas or submissions can be submitted online at <https://mc.manuscriptcentral.com/cfp> or through the *CFP* website <https://www.cfp.ca> under "Authors and Reviewers."