Race-based data collection among COVID-19 inpatients: A retrospective chart review

Clara Lu, MD[1],* ; Achieng Tago, MD[1],*, Oluwatobi Olaiya, MD, MSc[2],
and Madeleine Verhovsek, MD[1]

Abstract

Public health data have demonstrated disproportionate COVID-19 morbidity and mortality among racialized populations. However, limited hospital data may prevent research into racial disproportionality among inpatients. We conducted a retrospective cross-sectional study of patients admitted with or without COVID-19 to an Ontario tertiary hospital between March and October 2020 to determine the percentage of inpatients with a formal race or ethnicity assessment in their medical record. The COVID-19 group included inpatients with concurrent COVID-19 positivity; the reference group included a random sample of General Medicine inpatients without COVID-19. We reviewed 80 patients with COVID-19 and 80 patients without COVID-19. Formal ethnicity assessments were recorded among 44% of the COVID-19 group and 49% of the reference group. Race and ethnicity data collection was less than 50% among inpatients with and without COVID-19 in one Ontario hospital. Adequate data collection is necessary to study racial health disparities in the hospital setting.

Introduction

In healthcare systems, data is power. As illuminated by the COVID-19 pandemic, the quality of one's data will inform the quality, efficiency, and equity of public health initiatives and healthcare leadership decisions.

Once characterized as "the great equalizer," COVID-19 has instead unmasked complex and intersecting layers of health inequity.[1] Worldwide, evidence abounds for disproportionate rates of COVID-19 morbidity and mortality among racialized groups.[2-4] COVID-19 racial disproportionality has been attributed to increased exposure (for example, housing or employment conditions that limit one's ability to implement public health guidance) as well as increased susceptibility (for example, higher burdens of comorbid conditions), both of which are rooted in social determinants and systemic racism.[1,2,5-7]

Since the collection of race and ethnicity data has not historically been prioritized in the Canadian context,[3,7-11] evidence for racial disproportionality with COVID-19 in Canada has largely been limited to local community-level data rather than individual demographic data.[5,8-10,12-16] As such, the COVID-19 pandemic—along with increasing recognition of racism as a public health crisis—has renewed calls for high-quality race-based data in Canada.[3,7,10,11,17-21] By bringing racial health disparities to light, such data could prompt healthcare interventions extending far beyond COVID-19.

Responding to these calls, Public Health Units across Canada have acknowledged the importance of sociodemographic data[12,14-16,20,22] and some have initiated race-based data collection efforts during the pandemic.[10,12,16] However, Public Health Units also emphasize the limitations of missing race-based data among patients hospitalized with COVID-19.[12] In this study, we explore whether sufficient race and ethnicity data are collected at the hospital level to detect racial disproportionality among inpatients admitted with COVID-19, starting with one Ontario tertiary hospital.

Methods

Study design

We conducted a retrospective cross-sectional study using chart review data from a large tertiary academic hospital in Ontario, Canada. Variables in the dataset included patient demographics (age and sex), race or ethnicity data (reported vs. unknown), and outbreak at primary residence (yes vs. no). Data abstraction forms were piloted by two reviewers. The reviewers initially conducted a pilot review by extracting data from 10% of charts, which were subsequently reviewed to ensure inter-rater reliability using the kappa statistic.

Patients admitted between March 1, 2020 and October 31, 2020 were included in this study. We included all COVID-19 patients admitted to the General Internal Medicine Unit or the Intensive Care Unit between the study start and end dates. We defined COVID-19 patients as: (i) patients admitted for any reason who concurrently tested positive for COVID-19; (ii) patients admitted from a multi-residence housing facility (such as a nursing home) in the context of an outbreak with a positive COVID-19 test but without an acute indication for admission; or (iii) patients initially admitted to a unit outside of the COVID-19 unit who later tested positive for

[1] McMaster University, Hamilton, Ontario, Canada.
[2] McMaster University, Hamilton, Ontario, Canada.
* These authors contributed equally to this work.

Corresponding author:
Clara Lu, MD, McMaster University, Hamilton, Ontario, Canada.
E-mail: clara.lu@medportal.ca

COVID-19, with subsequent transfer to the COVID-19 unit. We matched COVID-19 patients with a random sample of reference patients. This reference group included patients admitted for any reason without concurrent COVID-19 infection to a General Medicine Unit. Pediatric patients (<18 years old) and adult patients who were assessed for admission by the Internal Medicine service, but discharged without admission, were excluded from the study.

Analysis

First, we summarized data descriptively as counts (percentages) in cross tabulations according to whether patients were COVID-19 positive or in the reference group. The primary outcome was any formal assessment of race or ethnicity. We defined race or ethnicity data as formally assessed when the "Ethnicity" section on the Electronic Medical Record (EMR) was completed. When race or ethnicity data were extracted from unstructured data in the patient's chart (eg, chart note mentioning patient's ethnicity), this was considered an informal assessment. Notably, the analysis was descriptive rather than comparative in nature, with the goal of estimating the amount of race or ethnicity data being collected at a large tertiary centre.

Second, we used logistic regression to determine the relationship between race or ethnicity data collection and COVID-19 status, age, and sex. In the COVID-19 group, we explored whether an association existed between race or ethnicity data collection and whether the patient was admitted in the context of an outbreak at their place of residence, without an acute medical indication for admission. We calculated and reported Area Under the Curve (AUC) as a summary measure of the model's ability to discriminate between cases (race or ethnicity data collection) and non-cases (unreported race or ethnicity data). We also reported crude Odds Ratio (OR), Adjusted Odds Ratios (aORs), corresponding 95% Confidence Intervals (CIs), and P values. The level of statistical significance was set at $\alpha = .05$.

Sensitivity analysis

In the COVID-19 sample, some patients were admitted to hospital for medical reasons while testing positive for COVID-19, while other patients were admitted due to a COVID-19 outbreak in their congregate care setting without an acute medical indication for admission. We conducted a sensitivity analysis excluding patients admitted due to a COVID-19 outbreak alone from the COVID-19 positive group to evaluate the robustness of our primary outcome.

Results

Eighty patients admitted with a COVID-19 infection between March and October 2020 were included in the COVID-19 group. A random sample of 80 patients admitted to the General Medicine Unit during the same period without COVID-19 comprised the reference group. The median patient age was 74 and 69 years in the COVID-19 and reference group, respectively, with balanced

sex distributions (Table 1). Our pilot chart review demonstrated good inter-rater reliability for the primary endpoint with a Cohen's kappa of 1.

For our primary endpoint, less than 50% of inpatients in each group had a formal assessment for race or ethnicity recorded in the "Demographics" section of the EMR (Table 1). Most of these assessments identified patients of "Caucasian" ethnicity, with less than 20% of formal assessments identifying patients from non-dominant groups.

Several categories available for selection on the hospital's EMR (Table 2) did not correspond to valid race categories as defined by the Government of Ontario Anti-Racism Data Standards and more recently proposed by the Canadian Institute of Health Information.[21,23] Consequently, three patients in each group (8.6% COVID-19 group and 7.7% reference group) had formal ethnicity assessments that could not be reassigned to a category recognized as valid.

Informal assessments for race or ethnicity were identified within provider chart documentation for 10% and 4% of inpatients in the COVID-19 and reference group, respectively. Among a handful of patients who had both formal and informal assessments in their records, only 40% had concordant assessments. Assessments among the remaining 60% were discordant or unclear—for example, if a patient was formally identified as "Caucasian" in the Demographics section but informally described as "South Asian" in chart notes.

Logistic regression analyses of the primary endpoint revealed that age, sex, and COVID-19 status were not predictive of race or ethnicity data collection (Table 3). Thirty percent of patients in the COVID-19 group were admitted due to a COVID-19 outbreak at their place of residence with no other acute medical indication for hospitalization. The odds of formal race or ethnicity data collection were 3.2 times (95% confidence interval 1.1–9.3, P = .031) when patients were admitted from non-outbreak settings compared to outbreak settings. After adjusting for age and sex, the odds of data collection were even higher (aOR = 4.44, 95% CI 1.4–14.3, P = .012). A sensitivity analysis excluding patients admitted due to a COVID-19 outbreak alone, without another indication for admission, increased the primary endpoint (formal race or ethnicity assessment) from 43.8% to 51.7%.

Discussion

Race-based data is an important tool necessary to highlight health inequities that will inform healthcare leadership and public health interventions. We sought to identify whether an Ontario tertiary care hospital was collecting adequate race and ethnicity data to make meaningful conclusions about racial disproportionality among COVID-19 inpatients. Our results indicate an insufficient quantity of race and ethnicity data collection. Less than 50% of COVID-19 inpatients had a formal assessment of race or ethnicity, among which greater than 80% were identified as "Caucasian." Our study also demonstrates an inadequate quality of data with poor concordance between formal and informal assessments. Notably, formal selections

Table 1. Collected patient demographic data by COVID-19 status.

Variable	COVID-19 group (n = 80)	Reference group (n = 80)
Age on admission, years	Median 74 (20–94)	Median 69 (19–100)
Female sex	40/80 (50.0%)	41/80 (51.2%)
Admission due to outbreak	24/80 (30.0%)	N/A
Formal ethnicity or race assessment	35/80 (43.8%)	39/80 (48.8%)
"Caucasian"	28/35 (80.0%)	34/39 (87.2%)
"Hispanic or Latino"	0/35 (0%)	1/39 (2.6%)
"Other Hispanic, Latino/a, or Spanish origin"	3/35 (8.6%)	3/39 (7.7%)
"African"	2/35 (5.7%)	0/39 (0%)
"Asian—Chinese"	2/35 (5.7%)	1/39 (2.6%)
Formal assessments that correspond to valid racial categories[a]	32/35 (91.4%)	36/39 (92.3%)
Informal ethnicity or race assessment	8/80 (10.0%)	3/80 (3.8%)
Both formal and informal assessments	5/80 (6.2%)	0/80 (0%)
Concordant assessments	2/5 (40.0%)	N/A
Discordant assessments	2/5 (40.0%)	N/A
Unclear	1/5 (20.0%)	N/A

[a]As defined by the Canadian Institute of Health Information (21) and the Government of Ontario (23). N/A, not applicable.

Table 2. Available ethnicity/race category options for selection on the hospital electronic medical record between March and October 2020.

Aboriginal—Off Reserve
Aboriginal—On Reserve
African
Asian-Chinese
Asian-Indian
Caucasian
Cuban
Decline to answer
Hispanic or Latino
Mediterranean
Mexican, Mexican American, or Chicano/a
North American Indian
Not Hispanic, Latino/a, or Spanish origin
Other Hispanic, Latino/a, or Spanish origin
Puerto Rican
Unknown

were made among EMR categories that did not correspond to valid government categories as defined by the Ontario Anti-Racism Data Standards.[23]

In our study, the majority of formal assessments in both the COVID-19 group and the reference group identified a "Caucasian" ethnicity. It is difficult to speculate whether this reflects each group's true demographics or an increased comfort with collecting race and ethnicity data from the majority racial group. However, if the purpose of race- and ethnicity-based data collection is to advance health equity, data collected predominantly from the majority group will be less useful for identifying health disparities among minority groups. As well, among a handful of patients with both formal and informal assessments of race and/or ethnicity, the agreement between these assessments was mostly discordant or unclear. One source of these inaccuracies may have been the EMR's requirement for selection of formal assessments from a non-standardized list of racial and ethnic categories that do not reflect societal constructs. Government-standardized categories do exist within Canada and their use could help minimize inaccuracies while maximizing data quality.[21,23]

Calls for higher quantity and quality of race and ethnicity data reporting have resounded throughout the COVID-19 pandemic.[10,13,24] Public health equity initiatives—for example, targeted vaccination programs for racialized individuals—are driven by health disparity data which in turn depend on adequate demographic data collection. McKenzie describes how proper collection of race-based data can be applied to decrease COVID-19 racial inequities.[18] For example, data gathered from June 2020 pointed to the Latino population in Toronto, Ontario, Canada, as the minority group with the highest rate of new COVID-19 infections. After implementation of targeted public health interventions based upon this statistic, subsequent monthly data analyses demonstrated a decrease in the number of cases among Latino people.[18] Here, the importance of adequate demographic data is clear: healthcare leadership cannot initiate such programs without first identifying that disparities exist and where they lie.[2,12,16,25]

Despite this, our findings align with recent studies underscoring how Canada lags behind other countries, including the United States, in the collection of race and ethnicity data.[7] A summary of COVID-19 health equity data reporting was undertaken by Blair et al. using a scorecard approach ranging from A+ (excellent) to F (very poor). Although Ontario achieved a high rating for population-level data collection such as age, sex, and occupational sector, there was a notable

Table 3. Factors associated with reporting ethnicity data in the electronic medical record.

Variable	Univariate model		Multivariable model	
	OR (95% CI)	P	aOR (95% CI)	p
Age	1.0 (.99–1.03)	.271	1.0 (.99–1.03)	.251
Sex				
Male	1 (reference group)		1 (reference group)	
Female	.95 (.51–1.76)	.865	.97 (.52–1.80)	.915
COVID status				
No	1 (reference group)		1 (reference group)	
Yes	.82 (.44–1.52)	.526	.79 (.42–1.49)	.470

dearth of equity-based data collection including categories such as race/ethnicity, income, and education. Similar findings were found across Canada with no single province achieving greater than an "F" score in COVID-19 equity data collection.[26] Even in the United States, where race and ethnicity data collection within COVID-19 studies is more robust, ethnic minorities are underrepresented in clinical trials used to set standards of care for COVID-19 infections, despite being disproportionately affected by COVID-19 morbidity and mortality.[24] This limits the generalizability of COVID-19 interventions, further negatively impacting marginalized groups.[7,23]

Our study has several strengths. Few studies in Canada have assessed the adequacy of race- and ethnicity-based data collection at the hospital level, and we hope that our work will spur further literature in this area and inform health leaders who endeavour to improve demographic data collection at their respective institutions. Logistic regression analyses showed that race/ethnicity data collection was not influenced by age, sex, or COVID-19 status. The COVID-19 pandemic in Canada was characterized by outbreaks among residents of long-term care, some of whom were hospitalized for logistical reasons rather than an acute medical indication for admission. However, a sensitivity analysis did not demonstrate a meaningful difference in our primary endpoint when excluding COVID-19 patients admitted due to outbreak alone. Finally, our study timeline spanned two waves of the COVID-19 pandemic, thereby ensuring that our sample was representative.

Our study also has certain limitations. During the study period, the available EMR employed outdated categories, did not allow for multiple selections and did not differentiate between race and ethnicity. In its Data Standards for the Identification and Monitoring of Systemic Racism, the Government of Ontario outlines valid race categories that should be used by Ontario public sector organizations including hospitals.[23] Therefore, a lack of appropriate categories may have contributed to inaccurate and inadequate race and ethnicity data collection. For example, the EMR used categories such as Puerto Rican (more consistent with American demographics) and the term "Caucasian" (originating from outdated 18th-century classifications). Importantly, more recent data collection standards differentiate between race and ethnicity: they emphasize using race-based data to highlight health inequities and using ethnicity data to

implement tailored public health services.[21] Last, our analysis was not stratified by predetermined time intervals. The statistics reported summarize data collection practices over a large cross-section of time. The accuracy and completion of race-based data collection may have varied throughout the pandemic, with more missing data occurring at peaks of disease incidence.

In summary, data from one Ontario tertiary care hospital demonstrate that less than 50% of patients hospitalized with or without COVID-19 between March and October of 2020 had a formal selection for race and/or ethnicity, with only 10% having an informal assessment. This has many implications for future work, which could expand to other hospital systems, diseases, and demographics. First, studies assessing the adequacy of race- and ethnicity-based data collection across hospital and provincial health systems are needed to identify gaps in demographic data collection and call on healthcare leadership to better adhere to current government standards for race and ethnicity data collection.[21,23] While this study focused on COVID-19 inpatients, data on race and ethnicity are likely to be lacking across a broad range of other diseases that may disproportionately affect minority groups. Finally, future studies on Canadian healthcare systems should explore health equity data collection among other marginalized groups such as the LGBTQ2IA2P community, rural and low-income communities, immigrants and refugees, and people with disabilities.

Racial health disparities highlighted by the COVID-19 pandemic—along with a widespread reckoning with racism both within and outside of healthcare—have prompted healthcare leadership across Canada to emphasize their commitment to antiracism and health equity. The long arc toward health equity begins with identifying areas of health inequity, for which collecting adequate demographic data is both a rudimentary and essential step.

ORCID iD

Clara Lu ⬤ https://orcid.org/0000-0003-0488-415X

References

1. Ali S, Asaria M, Stranges S. COVID-19 and inequality: Are we all in this together? *Can J Public Health*. 2020 Jun 23; 111(3): 415-416.

2. Mackey K, Ayers CK, Kondo KK, et al. Racial and ethnic disparities in COVID-19-related infections, hospitalizations, and deaths. *Ann Intern Med*. 2021 Mar; 174(3): 362-373.

3. Yaya S, Yeboah H, Charles CH, Otu A, Labonte R. Ethnic and racial disparities in COVID-19-related deaths: Counting the trees, hiding the forest. *BMJ Glob Health*. 2020 Jun; 5(6): e002913.

4. Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and mortality among black patients and white patients with Covid-19. *N Engl J Med*. 2020 Jun 25; 382(26): 2534-2543.

5. Tuyisenge G, Goldenberg SM. COVID-19, structural racism, and migrant health in Canada. *Lancet*. 2021 Feb 20; 397(10275): 650-652.

6. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. *Annu Rev Public Health*. 2006; 27: 1-28.

7. McKenzie K. *Race and ethnicity data collection during COVID-19 in Canada: if you are not counted you cannot count on the pandemic response*. [Internet]. 2020 Nov[cited 2021 Jul 10]. https://rsc-src.ca/en/race-and-ethnicity-data-collection-during-covid-19-in-canada-if-you-are-not-counted-you-cannot-count.

8. Denice P, Choi K, Haan M, Zajacova A. Visualizing the geographic and demographic distribution of COVID-19. *Socius Sociological Research for a Dynamic World*. 2020; 6: 213–216.

9. Choi KH, Denice P, Haan M, Zajacova A. Studying the social determinants of COVID-19 in a data vacuum. *Can Rev Sociol*. 2021 May; 58(2): 146-164.

10. McKenzie K, Dube S, Petersen S. *Tracking COVID-19 Through Race-Based Data [Internet]*. Wellesley Intitute & Ontario Health; 2021 Aug. p. 25.https://www.wellesleyinstitute.com/wp-content/uploads/2021/08/Tracking-COVID-19-Through-Race-Based-Data_eng.pdf

11. Devakumar D, Selvarajah S, Shannon G, et al. Racism, the public health crisis we can no longer ignore. *Lancet*. 2020; 395(10242): e112.

12. Toronto Public Health. *COVID 19: Ethno-Racial Identity & Income [Internet] City of Toronto*. City of Toronto; 2021. https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-pandemic-data/covid-19-ethno-racial-group-income-infection-data/

13. Subedi R, Greenberg L, Turcotte M. *COVID-19 mortality rates in Canada's ethno-cultural neighbourhoods*. [Internet]. Statistics Canada, Government of Canada; 2020 Oct. https://www150.statcan.gc.ca/n1/pub/45-28-0001/2020001/article/00079-eng.htm

14. Public Health Ontario. *Enhanced Epidemiological Summary: COVID-19 in Ontario – A Focus on Diversity*. [Internet]. 2020 May [cited 2021 Jul 10]. https://www.publichealthontario.ca/-/media/documents/ncov/epi/2020/06/covid-19-epi-diversity.pdf?la=en.

15. *Santé Montreal. Inequality during the pandemic: Racialized populations* [Internet]. 2021 [cited 2021 Jul 10]. https://santemontreal.qc.ca/en/public/coronavirus-covid-19/situation-of-the-coronavirus-covid-19-in-montreal/survey-of-the-health/racialized-populations/

16. City of Hamilton C of. *Hamilton's COVID-19 vaccination program prioritizes Black and other racialized populations*. /people of colour ages 18+ in priority neighbourhoods [Internet]. 2021 Apr[cited 2021 Jul 10]. https://www.hamilton.ca/government-information/news-centre/news-releases/hamiltons-covid-19-vaccination-program-prioritizes

17. Toronto Board of Health. *Addressing Anti-Black Racism as a Public Health Crisis in the City of Toronto*. [Internet]. 2020 [cited 2021 Jul 12]. http://app.toronto.ca/tmmis/viewAgendaItemHistory.do?item=2020.HL17.9

18. McKenzie K. Socio-demographic data collection and equity in covid-19 in Toronto. *EClinicalMedicine*. 2021 Apr 7; 34: 100812.

19. Ahmed R, Jamal O, Ishak W, Nabi K, Mustafa N. Racial equity in the fight against COVID-19: A qualitative study examining the importance of collecting race-based data in the Canadian context. *Tropical Diseases, Travel Medicine and Vaccines*. 2021 Jun 10; 7: 15.

20. Chief Public Health Officer of Canada. *From risk to resilience: An equity approach to COVID-19*. [Internet]. Government of Canada; 2020 Oct. [cited 2021 Jul 10]. https://www.canada.ca/en/public-health/corporate/publications/chief-public-health-officer-reports-state-public-health-canada/from-risk-resilience-equity-approach-covid-19.html

21. Canadian Institute for Health Information *Proposed Standards for Race-Based and Indigenous Identity Data Collection and Health Reporting in Canada*, 32; 2020.

22. Vancouver Social PolicyProjects ResearchData Team. *Populations Disproportionately Impacted by COVID-19: Current State Assessment*; 2021 Jan: 73.

23. Government of Ontario. *Data Standards for the Identification and Monitoring of Systemic Racism* [Internet]. 2018. [cited 2021 Jul 10]. https://www.ontario.ca/document/data-standards-identification-and-monitoring-systemic-racism https://www.ontario.ca/document/data-standards-identification-and-monitoring-systemic-racism

24. Public Health Ontario. *Environmental Scan: Addressing health inequities within the COVID19 public health response*. [Internet]. 2020 Dec. https://www.publichealthontario.ca/-/media/documents/ncov/he/2020/12/covid-19-environmental-scan-addressing-health-inequities.pdf?la=en

25. Hooper MW, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. *JAMA*. 2020 Jun 23; 323(24): 2466-2467.

26. Blair A, Warsame K, Naik H, Byrne W, Parnia A, Siddiqi A. Identifying gaps in COVID-19 health equity data reporting in Canada using a scorecard approach. *Can J Public Health*. 2021 Jun; 112(3): 352-362.