BMJ Open

Immigration Status and Mortality: The Importance of Accounting for Loss to Follow-up

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-046377
Article Type:	Original research
Date Submitted by the Author:	28-Oct-2020
Complete List of Authors:	Vyas, Manav; University of Toronto, Division of Neurology Fang, Jiming; ICES Austin, Peter; Institute for Clinical Evaluative Sciences, Laupacis, Andreas; University of Toronto Cheung, Matthew; University of Toronto Silver, Frank ; University of Toronto Kapral, MK; University of Toronto
Keywords:	EPIDEMIOLOGY, SOCIAL MEDICINE, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Immigration Status and Mortality: The Importance of Accounting for Loss to Follow-up

Manav V. Vyas[1,2,3], Jiming Fang[3], Peter C. Austin[2,3], Andreas Laupacis[2,3,4] Matthew C. Cheung[2,3,5], Frank L. Silver[1,3] and Moira K. Kapral[2,3,4]

1. Division of Neurology, Department of Medicine, University of Toronto, Canada
2. Institute of Health Policy, Management and Evaluation, University of Toronto, Canada
3. ICES, Toronto, Canada
4. Division of General Internal Medicine, Department of Medicine, University of Toronto, Canada
5. Division of Haematology & Oncology, Department of Medicine, University of Toronto, Canada

Abstract count: 246 words

Running Title: Immigration, Emigration, and Mortality

Word count: 2,928 words (not including abstract, references, tables, and figures)

Character count of the title: 85 (with spaces)

Number of references: 39

Keywords: immigration, emigration, mortality, lost to follow-up, emigration.

Figures: 3

Tables: 2

Corresponding author:

Manav V. Vyas, MBBS MSc FRCPC

Division of Neurology, Department of Medicine, University of Toronto

2075 Bayview Avenue, Toronto, ON M4N 3M5

Email: manav.vyas@mail.utoronto.ca

ABSTRACT

Introduction: Previous studies report lower mortality rates in immigrants compared to host populations, although many immigrants return to their home regions closer to end of life. Little is known about the variation in the association between immigration status and all-cause mortality in different disease cohorts and about the impact of emigration on this association.

Methods: Using linked population-based data, we followed adults with first-ever diagnosis of ischemic stroke, cancer, or schizophrenia between 2002 and 2013 in Ontario, Canada from index event to death, lost to follow-up (emigration), or end of follow-up in 2018. For each disease cohort, we calculated adjusted hazard ratios (HR) of death in immigrants compared to long-term residents, adjusting for demographic characteristics and comorbidities, with and without censoring for those who were lost to follow-up. We calculated the ratio of two HRs and the respective confidence limits (CL) using bootstrapping methods.

Results: Immigrants were more likely to be lost to follow-up than long-term residents in all disease cohorts. Not accounting for this lost to follow-up moderately overestimated the variable magnitude of association between immigration status and mortality based on underlying disease: ischemic stroke (HR of death before vs. after accounting for censoring: 0.78 vs. 0.83, ratio = 0.95; 95% CL, 0.93-0.97), cancer (0.74 vs. 0.78, ratio = 0.96; 0.95-0.96), and schizophrenia (0.54 vs. 0.56, ratio = 0.97; 0.96-0.98).

Conclusions: The immigrant survival advantage varies by the disease studied, and it is modestly overestimated by not for the higher loss to follow-up in immigrants.

ARTICLE SUMMARY

Strengths and limitations of this study

- Immigrants to high-income countries are known to have lower mortality rates than host populations; however, the extent to which loss to follow-up (emigration) accounts for this immigrant health advantage is not well-understood.
- In this retrospective cohort study of patients with first ever diagnosis of ischemic stroke, cancer, or schizophrenia, we present evidence of an immigrant mortality advantage that varies by the disease studied.
- Immigrants had higher rates of lost to follow-up (emigration) compared to long-term residents, and not accounting for loss to follow-up modestly overestimated the immigrant health advantage in all disease cohorts.
- The administrative database definition of lost to follow-up used in this study does not definitively suggest return migration of immigrants.

INTRODUCTION

In 2016, approximately 7.5 million immigrants were living in Canada, with nearly 3.8 million residing in Ontario and accounting for about 30% of total population of the most populous province in Canada [1]. Studies from high-income countries, including Canada, have shown that immigrants have lower mortality compared to host populations [2–4]. This immigrant health advantage has been observed not only for all-cause mortality, but also for cardiovascular mortality [5] and cancer mortality [6], and for the incidence of non-communicable chronic conditions such as cardiovascular disease [7], cancer [8] and schizophrenia [9].

This phenomenon has been termed the *healthy immigrant effect*, and is partly attributed to a selection bias whereby only those who are healthy and have human capital are able to migrate [10,11]. However, another potential explanation for the observed immigrant health advantage is a phenomenon termed the *salmon effect,* whereby immigrants return to their home countries when they are gravely ill [12]. Thus, they are lost to follow-up and not accounted for in studies that rely on the mortality statistics alone [13]. Previous work in both observational studies and randomized controlled trials has shown that unbalanced loss to follow-up in two comparison groups can lead to biased estimates of association [14,15]. Further, the variation in the immigrant health advantage in different chronic disease cohorts has been seldom reported.

The aim of this study was to use linked population-based data from Ontario, Canada to evaluate the association between immigration status and all-cause mortality after a new diagnosis of ischemic stroke, cancer or schizophrenia, and to quantify the loss to follow-up in immigrants and long-term residents and its influence on the association between immigration status and all-cause mortality in each disease cohort.

METHODS

Setting and population

We created three separate incident disease cohorts of ischemic stroke, primary cancer and schizophrenia using validated case definitions for incidence during an ascertainment period of April 1, 2002 to March 31, 2012 (e-table 1). We focused on these three diseases because they are chronic conditions associated with high morbidity that require regular health care system contact following the diagnosis. We identified patients with ischemic stroke using the Ontario Stroke Registry which is a province-wide registry that includes data on a random sample of consecutive patients seen at over 150 hospitals in Ontario [16]. Data collection for the registry was performed by chart abstractors with neurological expertise, with the final diagnosis and other data elements obtained through review of clinical and neuroimaging data. We identified patients with diagnosis of a primary malignant cancer from the Ontario Cancer Registry, a population-based registry, which is created by combining information from discharge and day surgery summaries, pathology reports with any mention of cancer, or records of patients referred to specialized institutions treating cancer patients in Ontario [17]. It captures approximately 95% of all cancer diagnoses in the province [17]. We identified patients with schizophrenia based on a validated algorithm whereby a diagnosis of schizophrenia was made if the patient had one or more hospital admissions and/or three or more outpatient visits with a diagnosis of schizophrenia or schizoaffective disorder [18].

Within each disease cohort, we excluded prevalent cases if they had a diagnosis of the specific disease prior to April 1, 2002. If patients had multiple cohort-defining events during the ascertainment period, only information at the time of the first cohort-defining event was recorded. We excluded patients who were younger than 18 years or older than 104 years at the time of the index event, those who resided in long-term care homes at the time of the index event, and those who resided in rural areas (population < 10,000) because most immigrants (> 95%) reside in large urban areas.

Using unique identifiers, we linked these cohorts to population-based data held securely at ICES (formerly known as Institute for Clinical Evaluative Sciences), Toronto. ICES is a prescribed entity under the Ministry of Health and Long-Term Care where Ontario's public health services data sets are stored, linked and used for research. We obtained information on neighbourhood-level income (in quintiles) based on the postal-code files, and on history of previous diagnosis of hypertension [19], diabetes [20], chronic obstructive pulmonary disease (COPD) [21], congestive heart failure (CHF) [22] and atrial fibrillation [23] using case definitions (e-table 1).

Exposure and outcomes

Our exposure of interest was immigration status obtained from the Ministry of Immigration, Refugee and Citizenship (IRCC) which collected information on all immigrants arriving in Ontario after 1985 (IRCC Permanent Resident Database). We classified individuals born outside of Canada who arrived in Ontario after 1985 as *immigrants*, and those born in Canada or those who were born outside of Canada but arrived before 1985 as *long-term residents*.

Our primary outcome was death from any cause, which was obtained from the death registry along with the date of death. We set the end date of follow-up as March 31, 2018.

We determined each person's date of last contact with the health system by using administrative databases to identify any contact with health care system such as a visit to a doctor's office, refill of prescriptions (in those over 65 years), hospitalization or emergency visits, receipt of home care, or admission to a rehabilitation facility (e-table 2) until January 31, 2020. Those alive on March 31, 2018 (end date of follow-up) with last health system contact prior to this date were flagged as *lost to follow-up* at the date of last health system contact (e-Figure 1).

Statistical analyses

Analyses were conducted separately in each disease cohort. We compared baseline characteristics between immigrants and long-term residents within each disease cohort using

the Chi-squared test for categorical variables and the Wilcoxon rank sum test for continuous variables.

We used the time of the index diagnosis as time zero. We created unadjusted cumulative incidence curves of death and loss to follow-up in immigrants and long-term residents, separately. We developed multivariable cause-specific hazards models to estimate the adjusted hazard ratio (HR) of loss to follow-up (emigration) in immigrants compared to long-term residents accounting for death as a competing event, and adjusting for age, sex, neighbourhood-level income, hypertension, diabetes, COPD, CHF and atrial fibrillation.

We then fit two multivariable Cox proportional hazards models to estimate the adjusted HR of death in immigrants compared to long-term residents, adjusting for demographic information and chronic conditions as before. In the first model we censored individuals only on March 31, 2018 (end date of follow-up). In the second model, we censored individuals on March 31, 2018, or at the date of last health system contact (i.e. date of loss to follow-up).

We then calculated a ratio of the two adjusted HRs obtained from these two models and calculated 95% confidence limits around this ratio using percentile-based bootstrapping methods and 1000 bootstrap samples. If the confidence limits for the ratio include 1, it would suggest that there is no statistical difference between the adjusted HRs obtained with and without accounting for loss to follow-up. The direction and magnitude of the difference between two HRs can be inferred based on the ratio, with values under 1 suggesting overestimation of the association between immigration status and mortality when not accounting for loss to follow-up. We similarly obtained adjusted HRs of death for each covariate in the multivariable models using two separate models, with and without accounting for loss to follow-up. Using the methods described above, we also evaluated whether the association between other covariates and mortality changed after accounting for loss to follow-up. All analyses were conducted using SAS 9.4

 We did not involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

Sensitivity analyses

We re-defined the date of last health care system contact as the recorded date plus 180 days to account for patients who may not interact with health care system for up to 6 months. We re-calculated the adjusted hazard of death accounting for loss to follow-up for each disease cohort. We chose a *lag-time* of 6 months because all patients in this study had a chronic condition that would typically require follow-up within this time frame.

To evaluate how the association between immigration status and mortality would change if those lost to follow-up had returned to their country of origin when gravely ill (salmon effect) and had died, we re-calculated the adjusted hazard of death in immigrants compared to long-term residents in two hypothetical scenarios in which patients, irrespective of the immigration status, were considered to have died within 30 days and 1 year following their last recorded health system contact.

RESULTS

The study sample included 24,557 patients with ischemic stroke, 310,529 patients with primary cancer and 54,691 patients with schizophrenia (Figure 1). A greater proportion of patients with schizophrenia were immigrants (17.4%) compared to those with ischemic stroke (8.5%) or cancer (8.4%) (Table 1). Irrespective of the underlying diagnosis, immigrants were younger at the time of the diagnosis and more likely to reside in a low-income neighbourhood compared to long-term residents (Table 1). Other characteristics of the study cohorts are shown in Table 1.

During a median follow-up of 7 years, 13 667 patients were lost to follow-up across three disease cohorts. A greater proportion of patients with schizophrenia were lost to follow-up than patients with ischemic stroke or cancer (Table 2). Immigrants were more likely than long-term

residents to be lost to follow up in all disease cohorts (Table 2 and Figure 2); however, the magnitude of association was greater in patients with ischemic stroke (HR 2.87; 95% CI 2.38-3.44) and cancer (HR 3.07; 95% CI 2.91-3.23) than schizophrenia (HR 1.54; 95% CI 1.44-1.64) (e-Table 3).

During 2.7 million person-years follow-up, 176 301 deaths were recorded across the three disease cohorts. The crude mortality rate was highest in patients with ischemic stroke (95.3 per 1000-person-years) followed by cancer (76.8 per 1000-person-year) and schizophrenia (13.7 per 1000-person years). In all three disease cohorts, the unadjusted hazard of mortality was lower in immigrants compared to long-term residents (Table 2 and Figure 2). This remained true even after accounting for baseline differences in age, socio-economic status and comorbidities, with an adjusted HR of death in immigrants compared to long-term residents of 0.78 [95% CI 0.73-0.84] in patients with ischemic stroke, a HR of 0.74 (95% CI 0.73-0.76) in patients with cancer, and a HR of 0.54 (95% CI 0.50-0.59) in patients with schizophrenia (Table 2). The magnitude of the mortality advantage in immigrants compared to long-term residents attenuated after accounting for loss to follow-up, with adjusted HR of death in immigrants compared to long-term residents of 0.83 (95% CI 0.77-0.89) for ischemic stroke, 0.78 (95% CI 0.76-0.79) for cancer, and 0.56 (95% CI 0.51-0.61) for schizophrenia (Table 2).

The ratio of two adjusted HRs obtained using models with and without accounting for loss to follow-up was 0.95 [95% confidence limits (CL), 0.93 to 0.97] for ischemic stroke, 0.96 (95% CL, 0.95 to 0.96) for cancer and 0.97 (95% CL, 0.96 to 0.98) for schizophrenia, suggesting that not accounting for loss to follow-up overestimated the mortality advantage in immigrants in all cohorts (Figure 3). The effect of not accounting for loss to follow-up on the association between other covariates and mortality is shown in Figure 3.

Sensitivity analyses

Using a lag-time of 6 months in determining the date of last health care system contact, to account for patients who may not interact with health care system, did not alter the association between immigration status and mortality for any disease cohort (e-table 3).

In hypothetical scenarios in which, irrespective of immigration status, patients lost to follow-up were considered to be dead at 30 days and 1 year, the healthy immigrant advantage was eliminated in patients with schizophrenia, and attenuated in patients with ischemic stroke and cancer (Table 2).

DISCUSSION

In this study using linked population-based data on over 380,000 patients with a new diagnosis of ischemic stroke, cancer or schizophrenia, we demonstrated that immigrants have a survival advantage compared to long-term residents, and that immigrants are more likely to be lost to follow-up than long-term residents. Not accounting for loss to follow-up resulted in statistically different, but a modest overestimation of the immigrant health advantage. Both the magnitude of the mortality advantage (or healthy immigrant effect) and the loss to follow up (or salmon effect) varied based on the disease studied.

Our finding of lower mortality in immigrants compared to long-term residents with stroke, cancer and schizophrenia is consistent with previous studies, and is partly explained by the younger age and the lower comorbidity in immigrants at the time of the index event [6,24]. However, the effect persisted even in the adjusted analyses. Of note, immigrants with schizophrenia had the greatest mortality advantage compared to those with ischemic stroke or cancer. While certain immigrant subgroups such as refugees or asylum seekers may be at increased risk of poor mental health outcomes [25] and mortality [26], the magnitude of mortality advantage in immigrants with schizophrenia observed in our study is consistent with previous reports of lower suicide rates in immigrants compared to long-term residents across different ethnic groups in the

US [27] and in Canadian youth [28]. The variation in the magnitude of this health advantage based on the disease is a novel contribution, and could be due to variation in incidence of these diseases between immigrants and long-term residents [7,29], or due to variation in disease-specific health care provision in immigrants compared to long-term residents [30,31]. Future studies will be required to thoroughly understand mechanisms for this observed heterogeneity.

We found a greater likelihood of loss to follow-up in immigrants compared to long-term residents in all disease cohorts, despite adjusting for known comorbidities and the competing risk of death. Because we were able to identify any health system contact of cohort participants using health administrative data, being lost to follow-up would mean that participant left the province. One explanation for this is that immigrants with chronic conditions with physical health care needs (especially ischemic stroke and cancer) may emigrate to their home countries to be closer to their family members [32]. Those with schizophrenia may be less likely to do so due to stigma related to mental health diagnoses in some countries of origin [33,34].

Because immigration status was directly related to the censoring event, out-migration, we found that accounting for loss to follow-up altered the magnitude of the association between immigration status and mortality. Thus, previous estimates of mortality advantage in immigrants that have relied on death statistics alone and did not account for lost to follow-up of immigrants may have overestimated the immigrant health advantage [4,13,35]. Accounting for loss to follow-up did not change the magnitude of association between mortality and other variables of interest included in the multivariable models, except for older age in the ischemic stroke and cancer cohorts. This suggests that studies using administrative health data to evaluate the association between other covariates (sex, income or comorbidities) and mortality could yield adequate results even if they fail to account for loss to follow-up.

Our study is strengthened by the use of comprehensive administrative databases that allowed us to identify loss of health system contact in three separate chronic disease cohorts. The

findings are likely to be generalizable to other jurisdictions with immigrant populations and to other disease conditions not included in this study, but the magnitude of bias may vary depending on the disease condition, health care jurisdiction, and immigrant-related variables (country of origin, time since immigration or immigration class).

Some limitations merit discussion. First, we did not have information on factors such as physical activity [36] and smoking [37], both of which are associated with mortality, and we did not have information specific to each disease condition such as disease severity, disability, response to treatment, or palliative care status, all of which could influence mortality. Second, we assumed that loss of health system contact equated to patients leaving health care jurisdiction rather than reflecting an excellent recovery negating the need for ongoing medical management. However, the misclassification introduced by this assumption should not vary based on immigration status. Third, we assumed that, at least in immigrants, loss to follow-up was likely to be due to emigration to their home countries rather than to other parts of Canada. A study from the IRCC found that only 9% of immigrants who landed in Ontario between 1991 and 2006 had moved to other provinces by 2006 [38]. Lastly, movement of individuals in and out of a health care jurisdiction is a dynamic process, and those who emigrate can return. If such individuals return after the end date of follow-up, they could be falsely censored at the date of their emigration.

This study demonstrated that inadequate handling of loss to follow-up can lead to biased estimation of immigrant health advantage as immigrant deaths may not be captured if immigrants return to their home region when gravely ill. Based on these findings, we recommend that future studies comparing mortality and other long-term outcomes in immigrants and non-immigrants carefully record loss to follow-up, quantify it, and account for it using appropriate methodology. When this information is not available, other measures could include use of updated postal code files during follow-up [39], measuring outcomes in short term or assuming specific rates of emigration based on previous reports. While the magnitude of the bias

associated with not accounting for lost to follow-up is small, this study highlights the heterogeneity in the healthy immigrant effect and salmon effect across different diseases, and supports the need of future studies to explain the reasons for the observed heterogeneity.

Data sharing statement– The data used in this study is held securely in coded form at ICES. Data sharing agreements prohibit ICES from making the dataset publicly available, but access may be granted to those who meet prespecified criteria for confidential access. Please contact corresponding author for details.

Code availability – Can be made available upon request to the corresponding author.

Ethics approval – This study was approved by Research Ethics Board at Sunnybrook Health Sciences Centre, Canada.

Patient and public involvement statement

Patients or the public WERE NOT involved in the design, or conduct, or reporting, or dissemination plans of our research

Disclaimer

The study was supported by ICES (formerly the Institute for Clinical Evaluative Sciences), which is supported by an annual grant from the Ontario Ministry of Health and Long-Term Care. Data were made available with help from the Immigration, Refugees and Citizenship Canada. Parts of this material are based on data and/or information compiled and provided by the Canadian Institute for Health Information. The analyses, opinions, results and conclusions reported in this article are those of the authors and are independent from the funding sources. No endorsement by ICES, the Ontario Ministry of Health and Long-Term Care, Canadian Institute for Health Information, or the Immigration, Refugees and Citizenship Canada is intended or should be inferred.

Contributorship statement

MVV, JF, and MKK were involved in the concept and design. MVV, PA, JF were involved in data acquisition, and analysis. FLS and MKK were involved in the primary data acquisition data for

Ontario Stroke Registry. All authors were involved in developing the project and interpreting the results. MVV was responsible for drafting the manuscript which was critically revised by everyone. MKK supervised the study and is the guarantor.

Funding

This study was funded from Heart and Stroke Foundation of Canada (HSFC GIA19-26333). MKK and PCA hold Mid-Career Investigator Awards from the Heart and Stroke Foundation of Canada. MKK holds the Lillian Love Chair in Women's Health from the University Health Network/University of Toronto. MVV holds a Fellowship Award from the Canadian Institutes of Health Research.

Competing interests – none.

References

1 Government of Canada SC. The Daily — Immigration and ethnocultural diversity: Key results from the 2016 Census. 2017.http://www.statcan.gc.ca/daily-quotidien/171025/dq171025b-eng.htm (accessed 26 Oct 2017).

2 Vang Z, Sigouin J, Flenon A, *et al.* The Healthy Immigrant Effect in Canada: A Systematic Review. *Population Change and Lifecourse Strategic Knowledge Cluster Discussion Paper Series/ Un Réseau stratégique de connaissances Changements de population et parcours de vie Document de travail* 2015;**3**.http://ir.lib.uwo.ca/pclc/vol3/iss1/4

3 Syse A, Dzamarija MT, Kumar BN, *et al.* An observational study of immigrant mortality differences in Norway by reason for migration, length of stay and characteristics of sending countries. *BMC Public Health* 2018;**18**:508. doi:10.1186/s12889-018-5435-4

4 Omariba DWR, Ng E, Vissandjée B. Differences between immigrants at various durations of residence and host population in all-cause mortality, Canada 1991-2006. *Population Studies* 2014;**68**:339–57. doi:10.1080/00324728.2014.915050

5 McKeigue PM, Marmot MG. Mortality from coronary heart disease in Asian communities in London. *BMJ* 1988;**297**:903. doi:10.1136/bmj.297.6653.903

6 Cheung MC, Earle CC, Fischer HD, *et al.* Impact of Immigration Status on Cancer Outcomes in Ontario, Canada. *J Oncol Pract* 2017;**13**:e602–12. doi:10.1200/JOP.2016.019497

7 Tu JV, Chu A, Donovan LR, *et al.* The Cardiovascular Health in Ambulatory Care Research Team (CANHEART). *Circulation: Cardiovascular Quality and Outcomes* 2015;**8**:204–12. doi:10.1161/CIRCOUTCOMES.114.001416

8 Paszat L, Sutradhar R, Liu Y, *et al.* Risk of colorectal cancer among immigrants to Ontario, Canada. *BMC Gastroenterol* 2017;**17**:85. doi:10.1186/s12876-017-0642-5

9 Anderson KK, Cheng J, Susser E, *et al.* Incidence of psychotic disorders among first-generation immigrants and refugees in Ontario. *CMAJ* 2015;**187**:E279–86. doi:10.1503/cmaj.141420

10 Marmot MG, Adelstein AM, Bulusu L. Lessons from the study of immigrant mortality. *Lancet* 1984;**1**:1455–7.

11 McDonald JT, Kennedy S. Insights into the 'healthy immigrant effect': health status and health service use of immigrants to Canada. *Social Science & Medicine (1982)* 2004;**59**:1613–27. doi:10.1016/j.socscimed.2004.02.004

12 Markides KS, Eschbach K. Aging, migration, and mortality: current status of research on the Hispanic paradox. *J Gerontol B Psychol Sci Soc Sci* 2005;**60 Spec No 2**:68–75. doi:10.1093/geronb/60.special_issue_2.s68

13 Khan AM, Urquia M, Kornas K, *et al.* Socioeconomic gradients in all-cause, premature and avoidable mortality among immigrants and long-term residents using linked death records in Ontario, Canada. *J Epidemiol Community Health* 2017;**71**:625–32. doi:10.1136/jech-2016-208525

14 Akl EA, Briel M, You JJ, *et al.* Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ* 2012;**344**. doi:10.1136/bmj.e2809

15 Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol* 1977;**106**:184–7. doi:10.1093/oxfordjournals.aje.a112451

16 Kapral MK, Hall RE, Silver FL, *et al.* Registry of the Canadian Stroke Network. Report on the 2004/05 Ontario Stroke Audit. https://www.ices.on.ca/~/media/Files/Atlases-Reports/2009/RCSN-report-on-2004-05-Ontario-stroke-audit/Full-report.ashx

17 Hall S, Schulze K, Groome P, *et al.* Using cancer registry data for survival studies: the example of the Ontario Cancer Registry. *J Clin Epidemiol* 2006;**59**:67–76. doi:10.1016/j.jclinepi.2005.05.001

18 Kurdyak P, Lin E, Green D, *et al.* Validation of a Population-Based Algorithm to Detect Chronic Psychotic Illness. *Can J Psychiatry* 2015;**60**:362–8. doi:10.1177/070674371506000805

19 Tu K, Chen Z, Lipscombe LL. Prevalence and incidence of hypertension from 1995 to 2005: a population-based study. *CMAJ* 2008;**178**:1429–35. doi:10.1503/cmaj.071283

20 Hux JE, Ivis F, Flintoft V, *et al.* Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002;**25**:512–6.

21 Gershon AS, Wang C, Guan J, *et al.* Identifying individuals with physcian diagnosed COPD in health administrative databases. *COPD* 2009;**6**:388–94.

22 Schultz SE, Rothwell DM, Chen Z, *et al.* Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic Dis Inj Can* 2013;**33**:160–6.

23 Tu K, Nieuwlaat R, Cheng SY, *et al.* Identifying Patients With Atrial Fibrillation in Administrative Data. *Canadian Journal of Cardiology* 2016;**32**:1561–5. doi:10.1016/j.cjca.2016.06.006

24 Saposnik G, Redelmeier DA, Lu H, *et al.* Risk of premature stroke in recent immigrants (PRESARIO) Population-based matched cohort study. *Neurology* 2010;**74**:451–7. doi:10.1212/WNL.0b013e3181cf6e9e

25 Berthold SM, Kong S, Mollica RF, *et al.* Comorbid mental and physical health and health access in Cambodian refugees in the US. *J Community Health* 2014;**39**:1045–52. doi:10.1007/s10900-014-9861-7

26 Hollander A-C. Social inequalities in mental health and mortality among refugees and other immigrants to Sweden – epidemiological studies of register data. *Glob Health Action* 2013;**6**. doi:10.3402/gha.v6i0.21059

27 Borges G, Orozco R, Rafful C, *et al.* Suicidality, ethnicity and immigration in the USA. *Psychol Med* 2012;**42**:1175–84. doi:10.1017/S0033291711002340

28 Saunders NR, Lebenbaum M, Stukel TA, *et al.* Suicide and self-harm trends in recent immigrant youth in Ontario, 1996-2012: a population-based longitudinal cohort study. *BMJ Open* 2017;**7**:e014863. doi:10.1136/bmjopen-2016-014863

29 Singh GK, Hiatt RA. 072: Cancer Incidence, Mortality, and Health-Related Behaviors Among US Immigrants. *Am J Epidemiol* 2005;**161**:S18–S18. doi:10.1093/aje/161.Supplement_1.S18c

30 Yarnell CJ, Fu L, Manuel D, *et al.* Association Between Immigrant Status and End-of-Life Care in Ontario, Canada. *JAMA* 2017;**318**:1479–88. doi:10.1001/jama.2017.14418

31 Chen A, Kazanjian A. 516: Ethnicity, Time in Canada and Pap Testing among Recent Immigrants to British Columbia, Canada. *Am J Epidemiol* 2005;**161**:S129–S129. doi:10.1093/aje/161.Supplement_1.S129c

32 Giraudo M, Bena A, Costa G. Migrant workers in Italy: an analysis of injury risk taking into account occupational characteristics and job tenure. *BMC public health* 2017;**17**:351. doi:10.1186/s12889-017-4240-9

33 Saunders NR, Gill PJ, Holder L, *et al.* Use of the emergency department as a first point of contact for mental health care by immigrant youth in Canada: a population-based study. *CMAJ* 2018;**190**:E1183–91. doi:10.1503/cmaj.180277

34 Derr AS. Mental Health Service Use Among Immigrants in the United States: A Systematic Review. *Psychiatr Serv* 2016;**67**:265–74. doi:10.1176/appi.ps.201500004

35 Kennedy S, Kidd MP, McDonald JT, *et al.* The Healthy Immigrant Effect: Patterns and Evidence from Four Countries. *Int Migration & Integration* 2015;**16**:317–32. doi:10.1007/s12134-014-0340-x

36 Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet* 1997;**349**:1436–42. doi:10.1016/S0140-6736(96)07495-8

37 GBD 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. *Lancet* 2017;**389**:1885–906. doi:10.1016/S0140-6736(17)30819-X

38 Interprovincial Mobility of Immigrants in Canada 2006-2011 - Canada.ca. https://www.canada.ca/en/immigration-refugees-citizenship/corporate/reports-statistics/research/interprovincial-mobility-immigrants-canada-2006-2011.html (accessed 2 Mar 2020).

39 Youens D, Preen DB, Harris MN, *et al.* The importance of historical residential address information in longitudinal studies using administrative health data. *Int J Epidemiol* 2018;**47**:69–80. doi:10.1093/ije/dyx156

Figure 1. Cohort selection and follow-up.

Footnote: values in parenthesis represent proportion.

Table 1. Baseline characteristics and mortality rate in immigrants and long-term residents with a first-ever diagnosis of ischemic stroke, cancer or schizophrenia between 2002 and 2012 in Ontario, Canada.

	Ischemic stroke		Cancer		Schizophrenia	
	Immigrants	Long-term residents	Immigrants	Long-term residents	Immigrants	Long-term residents
	2078 (8.5)	22 479 (91.5)	26 084 (8.4)	284 445 (91.6)	9525 (17.4)	45 166 (82.6)
Female, n (%)	982 (47.3)	10 697 (47.6)	13 602 (52.1)	130 324 (45.8)	4346 (45.6)	19 943 (44.2)
Median age in years at index event	68 (55-78)	74 (63-82)	58 (48-70)	67 (58-76)	34 (25-45)	40 (26-53)
Neighbourhood-level income, n (%)						
Lowest quintile (1st)	668 (32.1)	5043 (22.4)	7041 (27.0)	50 044 (17.6)	3803 (39.9)	13 525 (29.9)
Highest quintile (5th)	201 (9.7)	4330 (19.3)	3326 (12.8)	62 667 (22.0)	734 (7.7)	6434 (14.2)
Hypertension, n (%)	1420 (68.3)	16 046 (71.4)	11 120 (42.6)	152 177 (53.5)	1165 (12.2)	8253 (18.3)
Diabetes, n (%)	727 (35.0)	6495 (28.9)	4850 (18.6)	53 444 (18.8)	737 (7.7)	4178 (9.3)
Congestive heart failure, n (%)	258 (12.4)	3728 (16.6)	878 (3.4)	20721 (7.3)	59 (0.6)	807 (1.8)
COPD, n (%)	111 (5.3)	2547 (11.3)	1023 (3.9)	31 745 (11.2)	60 (0.6)	1494 (3.3)
Atrial fibrillation, n (%)	243 (11.7)	3786 (16.8)	777 (3.0)	19 278 (6.8)	34 (0.4)	525 (1.2)

Parenthesis represent 1st and 3rd quartile for continuous variables; whereas it represents the proportion of total for count variables, COPD – chronic obstructive pulmonary disease.

Table 2. Lost to follow-up and mortality in immigrants and long-term residents with and without accounting for loss to follow-up.

	Ischemic stroke		Cancer		Schizophrenia	
	Immigrants	Long-term residents	Immigrants	Long-term residents	Immigrants	Long-term residents
	2078 (8.5)	22 479 (91.5)	26 084 (8.4)	284 445 (91.6)	9525 (17.4)	45 166 (82.6)
Loss to follow-up, n (%)	158 (7.6)	512 (2.3)	2016 (7.7)	5995 (2.1)	1238 (13.0)	3748 (8.3)
Adjusted HR of lost to follow-up (95% CI)[a] with competing risk of death	2.87 (2.38-3.44)	1.00	3.07 (2.91-3.23)	1.00	1.54 (1.44-1.64)	1.00
Death, n (%)	796 (35.4)	12 575 (55.9)	9014 (34.6)	146 723 (51.6)	546 (5.7)	6647 (14.7)
Unadjusted HR (95 %CI)	0.61 (0.56-0.65)	1.00	0.60 (0.59-0.62)	1.00	0.39 (0.35-0.42)	1.00
Adjusted HR[a] (95% CI) not accounting for loss to follow-up	0.78 (0.73-0.84)	1.00	0.74 (0.73-0.76)	1.00	0.54 (0.50-0.59)	1.00
Adjusted HR[a] (95% CI) accounting for loss to follow-up[b]	0.83 (0.77-0.89)	1.00	0.78 (0.76-0.79)	1.00	0.56 (0.51-0.61)	1.00
Sensitivity analyses[c]						

Death within 30 days of loss of follow-up						
Adjusted HR (95% CI)	0.93 (0.87-1.00)	1.00	0.90 (0.88-0.91)	1.00	1.00 (0.95-1.05)	1.00
Death within 1 year of loss of follow-up						
Adjusted HR (95% CI)	0.93 (0.87-0.99)	1.00	0.89 (0.87-0.91)	1.00	1.00 (0.95-1.06)	1.00

[a]Multivariable adjusted model adjusted for the following: age, sex, neighbourhood-level income, and comorbidities (known hypertension, diabetes, congestive heart failure, chronic obstructive pulmonary disease, and atrial fibrillation); [b]censoring those who were lost to follow-up, which was determined when date of last health system contact occurred before end of follow-up among those alive; [c]assigning date of death among those loss to follow-up and re-calculating adjusted hazard of death. Abbreviations: HR – hazard ratio, CI – confidence interval.

Figure 2. Cumulative incidence curves in immigrants (red) and long-term residents (blue) showing probabilty of death (left hand figures) and of loss to follow-up (right hand figures) in patients with ischemic stroke (top), primary cancer (middle) and schizophrenia (bottom).

Figure 3. Ratios of adjusted hazard ratios of death obtained using two multivariable cox-regression models with and without accounting for loss to follow-up. Each box represents point estimate of this ratio, and the error bars represent 95% confidence limits. Values less than 1 suggest overestimation of the magnitude of association when loss to follow-up is not accounted for.

Footnote: Immigrants are compared to long-term residents; age less than 55 years is the comparison group; and the 5th quintile of income represent the HR of death in the highest quintile compared to lowest quintile based on neighbourhood-level income.

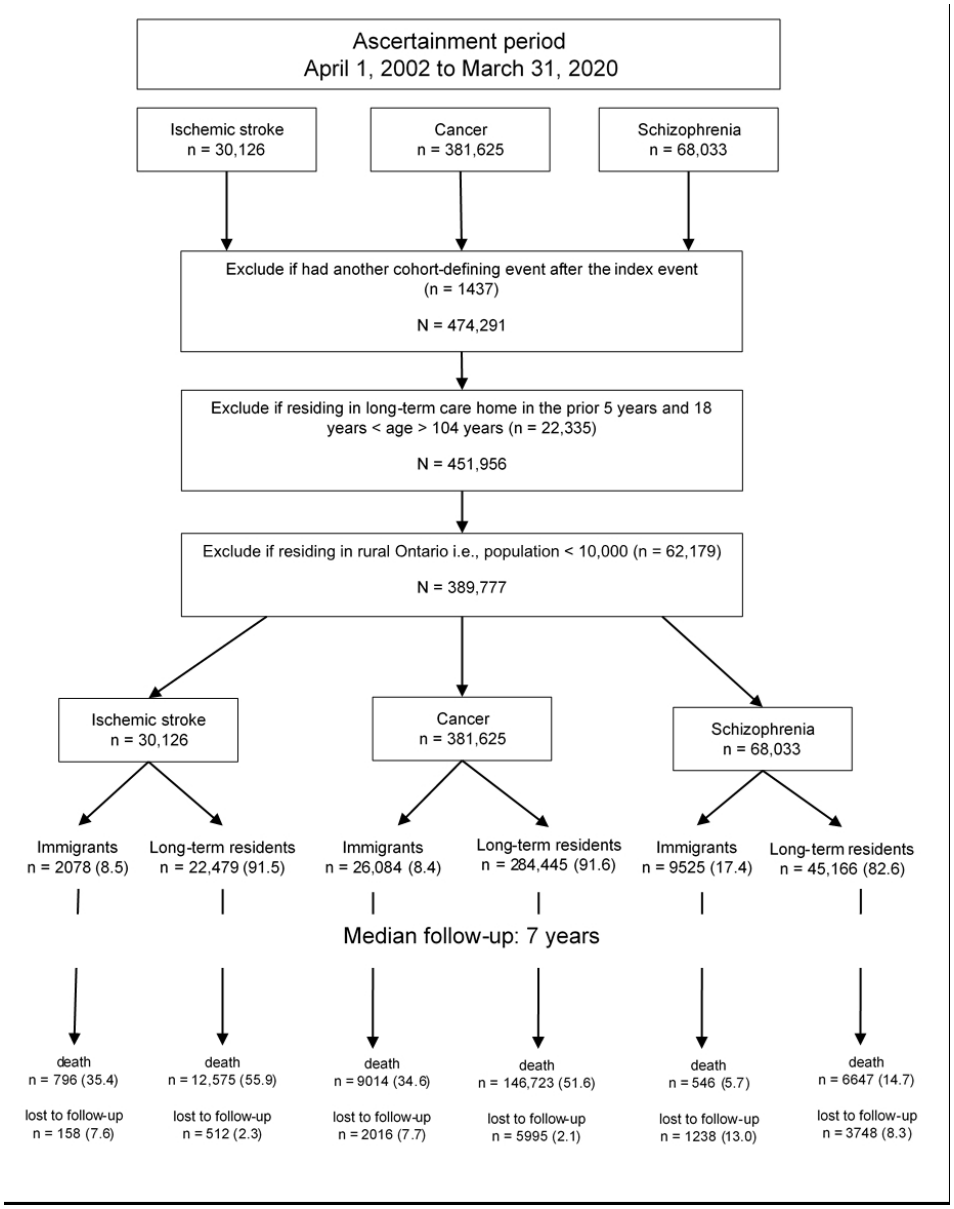


Figure 1

162x206mm (144 x 144 DPI)

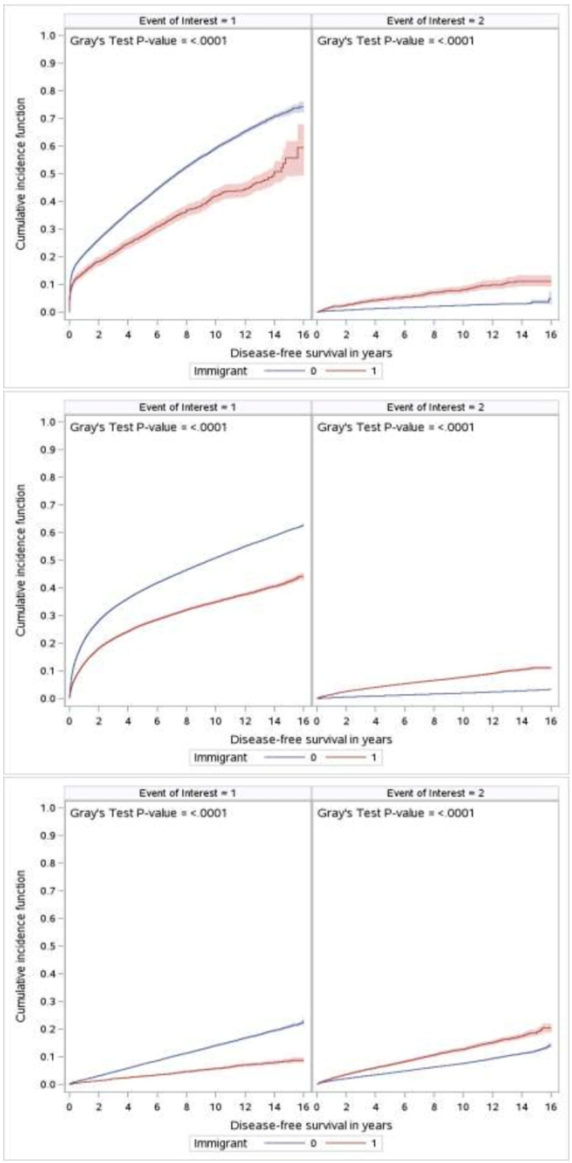

Event of Interest = 1
Event of Interest = 2
Gray's Test P-value = <.0001
Gray's Test P-value = <.0001
Cumulative incidence function
Disease-free survival in years
Immigrant 0 1

Figure 2

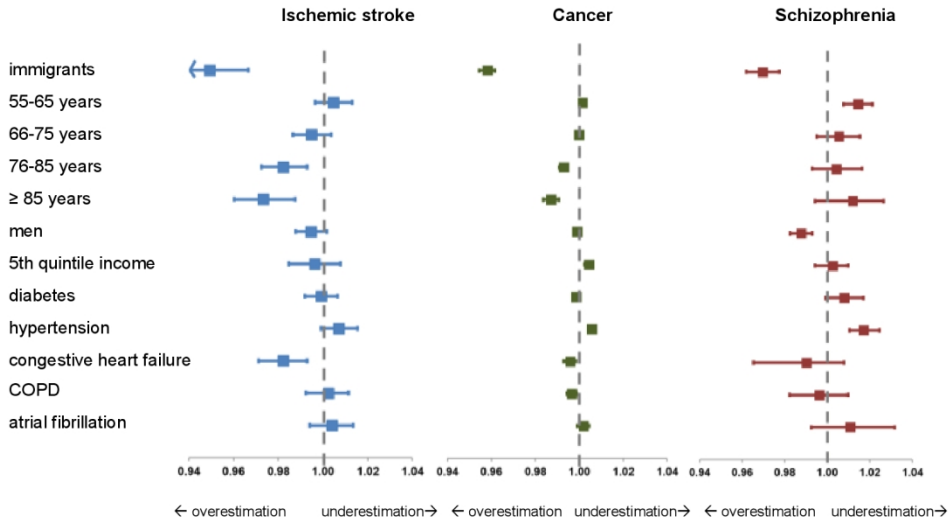

Ischemic stroke
Cancer
Schizophrenia
immigrants
55-65 years
66-75 years
76-85 years
≥ 85 years
men
5th quintile income
diabetes
hypertension
congestive heart failure
COPD
atrial fibrillation
0.94
0.96
0.98
1.00
1.02
1.04
← overestimation
underestimation→

Figure 3

Appendix

e-table 1. Definitions of variables included in the study.

e-table 2. Administrative databases used to determine date of last health system contact and statistics on contact with health care system in Ontario.

e-table 3. Results of sensitivity analyses using a lag-time of 6 months when determining the date of last health system contact.

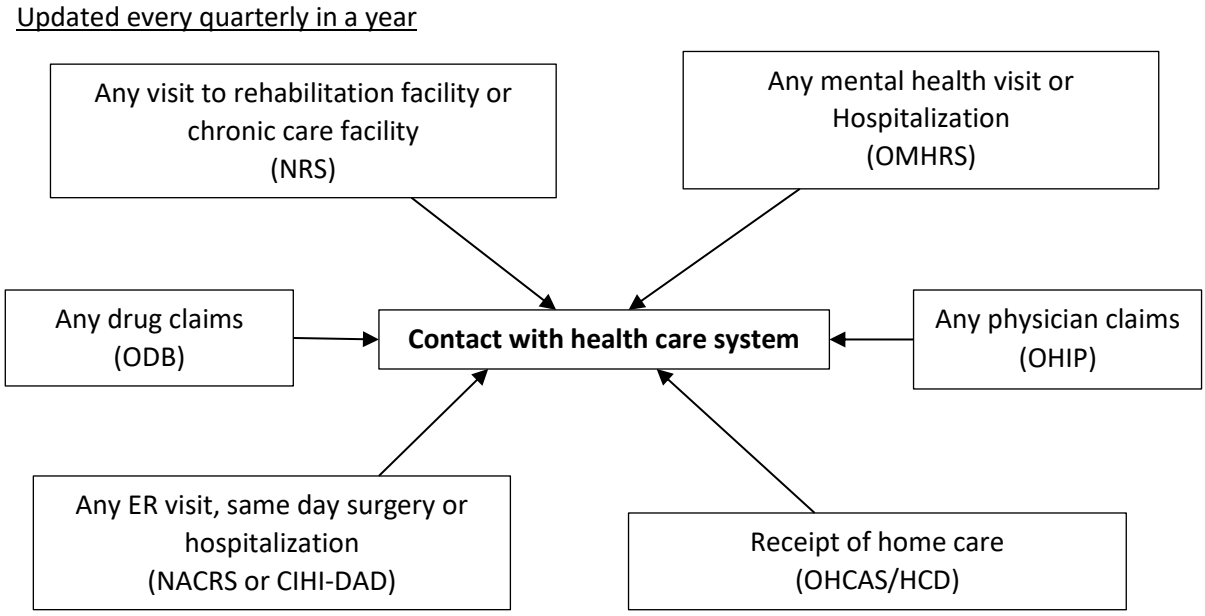
e-figure 1. Hypothetical cases to illustrate loss to follow-up using administrative database. Subject A was not to lost to follow-up, Subject B would be considered lost to follow-up, and Subject C had the event of interest (death) and so is not considered lost to follow-up.

e-table 1. Definitions of variables included in the study.

Variable	Definition
Incident ischemic stroke	Hospitalized or non-hospitalized adult patients (i.e., outpatient/ER) with confirmed acute ischemic stroke between April 1, 2002 and March 31, 2013 seen at 150 participating hospitals – regional and non-regional stroke centres. Information gathered using chart abstractors with neurological expertise, with the final diagnosis and other data elements obtained through review of clinical and neuroimaging data.
Incident primary cancer	A diagnosis of cancer either in hospitalized or non-hospitalized adult patients obtained from 4 different sources: hospital or ER visit using appropriate ICD codes, pathology reports with a diagnosis of cancer, regional cancer centres where patients with cancer are seen, and death certificates.
Incident schizophrenia	primary diagnosis of schizophrenia or schizoaffective disorder from a general hospital bed (prior to 2002, ICD9 - 295; as of 2002 ICD10 - F20 or F25) **OR** primary diagnosis of schizophrenia from a psychiatric hospital bed (DSM-IV – 295.x) **OR** three outpatient visits with a diagnosis of schizophrenia (295 or F20/F25) from outpatient physician billings within a 3-year period. 93.1% Sensitivity - 58.7% Specificity
Hypertension	≥ 1 Hospitalization **OR**

	≥ 2 physician claims in a two-year period **OR** 1 physician claim followed by another physician claim or hospitalization within two years. 72% Sensitivity - 95% Specificity - 87% PPV - 88% NPV
Diabetes	≥ 3 physician diagnostic code (250) in a one-year period 79.9% Sensitivity - 99.1% Specificity - 91.4% PPV
CHF (congestive heart failure)	≥ 1 Hospitalization **OR** 1 physician claim in ER or clinic, followed by ≥ 1 Hospitalization, ER visit, or physician claim within one year. 84.8% Sensitivity - 97.0% Specificity - 55.6% PPV
Atrial fibrillation	1 hospitalization (CIHI-DAD) or 1 emergency room visit (NACRS/SDS), ICD-10 (2002 onwards) – I48; ICD-9 (pre-2002) – 427.31 or 427.32 **OR** cardioversion (without physician billing codes) – using Billing code Z437
COPD (chronic obstructive pulmonary disease)	≥1 Hospitalization for COPD OR ≥ 3 physician claims in a two-year period 57.5% Sensitivity - 95.4% Specificity

e-table 2. Administrative databases used to determine date of last health system contact and statistics on contact with health care system in Ontario.



Abbreviations:

NRS – National Rehabilitation Reporting System; ODB – Ontario Drug Benefit; NACRS – National Ambulatory Care Reporting System; CIHI-DAD – Canadian Institute for Health Information-Discharge Abstract Database; OHCAS – Ontario Home Care Administration System; HCD – Home Care Database; OHIP – Ontario Health Insurance Plan Claims Database; OMHRS – Ontario Mental Health Reporting System

% of eligible people with some health care contact in 2015

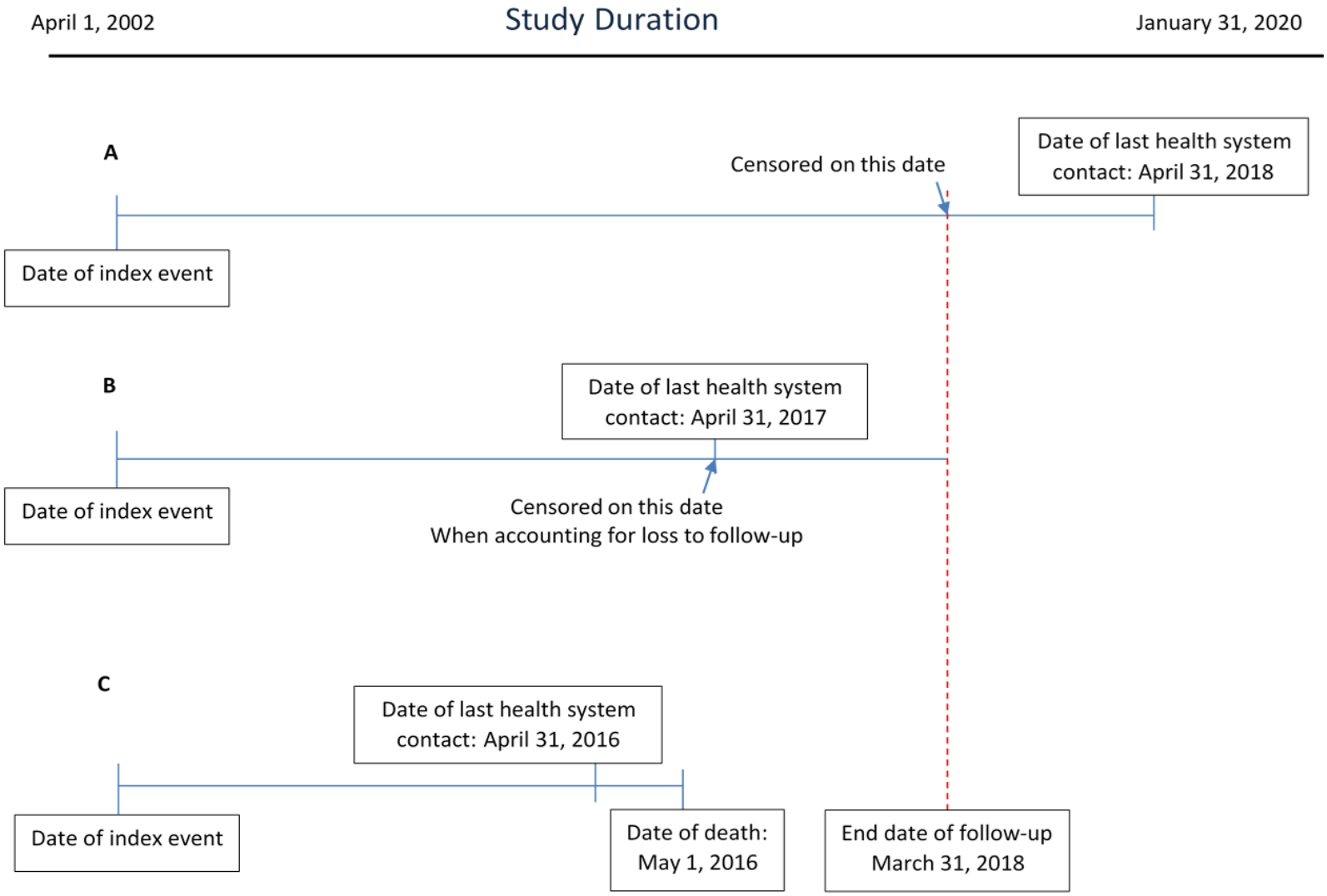
	Age in years (grouped)				
Sex	0-19	20-35	36-65	66-84	≥ 85
Male	85%	69%	80%	95%	94%
Female	86%	85%	88%	96%	96%

e-table 3. Results of sensitivity analyses using a lag-time of 6 months when determining the date of last health system contact.

		Ischemic stroke	Cancer	Schizophrenia
Immigrants	loss of follow-up, n (%)	145 (7.0)	1895 (7.3)	1120 (11.8)
Long-term residents		472 (2.1)	5472 (1.9)	3333 (7.4)
Immigrants	death, n (%)	796 (35.4)	9014 (34.6)	546 (5.7)
Long-term residents		12,575 (55.9)	146,723 (51.6)	6647 (14.7)
Adjusted HR of death (95% CI)^	Immigrants vs. long-term residents	0.82 (0.77-0.89)	0.77 (0.76-0.79)	0.56 (0.51-0.61)

^accounting for loss of follow-up by censoring them.

e-figure 1. Hypothetical cases to illustrate loss to follow-up using administrative database. Subject A was not to lost to follow-up, Subject B would be considered lost to follow-up, and Subject C had the event of interest (death) and so is not considered lost to follow-up.



STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Location
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 4
Methods			
Study design	4	Present key elements of study design early in the paper	Page 5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 5 and e-table 2
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	n.a.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 5 and e-table 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Page 5 and e-table 1
Bias	9	Describe any efforts to address potential sources of bias	Page 6
Study size	10	Explain how the study size was arrived at	Figure 1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Table 1
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Page 6-7
		(*b*) Describe any methods used to examine subgroups and interactions	Page 6-7
		(*c*) Explain how missing data were addressed	None present
		(*d*) If applicable, explain how loss to follow-up was addressed	e-table 2 and page 7
		(*e*) Describe any sensitivity analyses	Page 8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Figure 1
		(b) Give reasons for non-participation at each stage	Figure 1

		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	None with missing data
		(c) Summarise follow-up time (eg, average and total amount)	Page 8
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 9 and Table 2
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Page 9, and table 2
		(*b*) Report category boundaries when continuous variables were categorized	Table 2
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n.a.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Figure 3
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Page 12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Page 10
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

Immigration Status and Mortality: The Importance of Accounting for Loss to Follow-up

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-046377.R1
Article Type:	Original research
Date Submitted by the Author:	11-May-2021
Complete List of Authors:	Vyas, Manav; University of Toronto, Division of Neurology Fang, Jiming; ICES Austin, Peter; ICES Laupacis, Andreas; University of Toronto, Department of Medicine Cheung, Matthew; University of Toronto, Division of Haematology & Oncology, Department of Medicine Silver, Frank ; University of Toronto, Division of Neurology, Department of Medicine Kapral, MK; University of Toronto, Division of General Internal Medicine, Department of Medicine
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Health services research
Keywords:	EPIDEMIOLOGY, SOCIAL MEDICINE, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Immigration Status and Mortality: The Importance of Accounting for Loss to Follow-up

Manav V. Vyas[1,2,3], Jiming Fang[3], Peter C. Austin[2,3], Andreas Laupacis[2,3,4] Matthew C. Cheung[2,3,5], Frank L. Silver[1,3] and Moira K. Kapral[2,3,4]

1. Division of Neurology, Department of Medicine, University of Toronto, Canada
2. Institute of Health Policy, Management and Evaluation, University of Toronto, Canada
3. ICES, Toronto, Canada
4. Division of General Internal Medicine, Department of Medicine, University of Toronto, Canada
5. Division of Haematology & Oncology, Department of Medicine, University of Toronto, Canada

Abstract count: 248 words

Running Title: Immigration, Emigration, and Mortality

Word count: 3,360 words (not including abstract, references, tables, and figures)

Character count of the title: 85 (with spaces)

Number of references: 39

Keywords: immigration, emigration, mortality, lost to follow-up, emigration.

Figures: 3

Tables: 2

Corresponding author:

Dr Manav V. Vyas, MBBS MSc PhD FRCPC

Staff Neurologist, St. Michael's Hospital

Assistant Professor, Department of Medicine, University of Toronto

Office: Li Ka Shing Knowledge Institute, Rm 628, 250 Yonge Street, Toronto, ON M5B 2L7

Email: manav.vyas@mail.utoronto.ca | Phone: 416-864-5056 | Fax: 416-864-5712

ABSTRACT

Objectives: To evaluate the association between immigration status and all-cause mortality in different disease cohorts, and the impact of loss to follow-up on the observed associations.

Design: Population-based retrospective cohort study using linked administrative health data in Ontario, Canada.

Setting: We followed adults with a first-ever diagnosis of ischemic stroke, cancer, or schizophrenia between 2002 and 2013 from index event to death, loss to follow-up, or end of follow-up in 2018.

Primary and secondary outcome measures: Our outcomes of interest were all-cause mortality and loss to follow-up. For each disease cohort, we calculated adjusted hazard ratios (HR) of death in immigrants compared to long-term residents, adjusting for demographic characteristics and comorbidities, with and without censoring for those who were lost to follow-up. We calculated the ratio of two the HRs and the respective confidence limits (CL) using bootstrapping methods.

Results: Immigrants were more likely to be lost to follow-up than long-term residents in all disease cohorts. Not accounting for this loss to follow-up overestimated the magnitude of the association between immigration status and mortality in those with ischemic stroke (HR of death before vs. after accounting for censoring: 0.78 vs. 0.83, ratio = 0.95; 95% CL, 0.93-0.97), cancer (0.74 vs. 0.78, ratio = 0.96; 0.95-0.96), and schizophrenia (0.54 vs. 0.56, ratio = 0.97; 0.96-0.98).

Conclusions: Immigrants to Canada have a survival advantage that varies by the disease studied. The magnitude of this advantage is overestimated by not accounting for the higher loss to follow-up in immigrants.

ARTICLE SUMMARY

Strengths and limitations of this study

- Immigrants to high-income countries have lower mortality rates than host populations; however, the extent to which loss to follow-up (often due to emigration) accounts for this immigrant health advantage is not well-understood.
- In this retrospective cohort study of patients with a first ever diagnosis of ischemic stroke, cancer, or schizophrenia, we present evidence of an immigrant mortality advantage that varies by the disease studied.
- Loss to follow-up was higher in immigrants than long-term residents. Failure to account for this overestimated the immigrant health advantage in all disease cohorts.
- Our findings support the need to report, quantify, and account for loss to follow-up when evaluating long-term outcomes for a variety of health conditions.

INTRODUCTION

Studies from high-income countries, including Canada, have shown that immigrants have a lower mortality compared to host populations [1–3]. This immigrant health advantage has been observed not only for all-cause mortality, but also for cardiovascular [4] and cancer mortality [5], and for the incidence of non-communicable chronic conditions such as cardiovascular disease [6], cancer [7] and schizophrenia [8].

This phenomenon has been termed the *healthy immigrant effect*, and is partly attributed to a selection bias whereby only those who are healthy and have human capital are able to migrate [9,10]. However, another potential explanation for the observed immigrant health advantage is a phenomenon termed the *salmon effect,* whereby immigrants return to their home countries when they are gravely ill [11]. Thus, they are lost to follow-up and not accounted for in studies that rely on mortality statistics that do not record emigration [12]. Such lack of complete follow-up was identified in a large-scale meta-analysis on immigrant mortality in which none of the included 96 studies accounted for loss to follow-up and only 29 (28%) studies identified loss to follow-up as an issue [13]. Previous work in both observational studies and randomized controlled trials has shown that unbalanced loss to follow-up in two comparison groups can lead to biased estimates of association [14,15].

The aims of this study were to use linked population-based data from Ontario, Canada to evaluate the association between immigration status and all-cause mortality after a new diagnosis of ischemic stroke, cancer or schizophrenia, to quantify loss to follow-up in immigrants compared to long-term residents, and to determine how accounting for loss to follow-up influences the association between immigration status and all-cause mortality in each disease cohort.

METHODS

Setting and population

Ontario is Canada's most populous province, with an estimated population of 14 million [16]. In 2016, approximately 3.8 million immigrants were living in Ontario, accounting for about 30% of the total population [17]. All residents of Ontario (except for undocumented migrants and those visiting) are covered by the provincial health plan that includes physician services, hospital and emergency care, and investigations ordered by physicians. The linked administrative databases in Ontario capture all health care system contacts of insured Ontario residents.

We created three separate incident disease cohorts of ischemic stroke, primary cancer and schizophrenia using validated case definitions for incidence during an ascertainment period of April 1, 2002 to March 31, 2012 (e-table 1). We focused on these three diseases because they are chronic conditions associated with high morbidity that require regular health care system contact (which is captured in administrative databases) following the diagnosis, and because management of these conditions may benefit from family and social supports, which can be a factor in emigration. We identified patients with ischemic stroke using the Ontario Stroke Registry which is a province-wide registry that includes data on a random sample of consecutive patients seen at over 150 hospitals in Ontario [18]. Data collection for the registry was performed by chart abstractors with neurological expertise, with the final diagnosis and other data elements obtained through review of clinical and neuroimaging data. We identified patients with a diagnosis of a primary malignant cancer from the Ontario Cancer Registry, a population-based registry, which is created by combining information from discharge and day surgery summaries, pathology reports with any mention of cancer, or records of patients referred to specialized institutions treating cancer patients in Ontario [19]. It captures approximately 95% of all cancer diagnoses in the province [19]. We identified patients with schizophrenia based on a validated algorithm whereby a diagnosis of schizophrenia was made if the patient had one or more

hospital admissions and/or three or more outpatient visits with a diagnosis of schizophrenia or schizoaffective disorder [20].

Within each disease cohort, we excluded prevalent cases if they had a diagnosis of the specific disease prior to April 1, 2002. If patients had multiple cohort-defining events during the ascertainment period, only information at the time of the first cohort-defining event was recorded. We excluded patients who were younger than 18 years or older than 104 years at the time of the index event, those who resided in long-term care homes at the time of the index event, and those who resided in rural areas (population < 10,000) because most immigrants (> 95%) reside in large urban areas.

Using unique identifiers, we linked these cohorts to population-based data held securely at ICES (formerly known as the Institute for Clinical Evaluative Sciences), Toronto. ICES is a prescribed entity under the Ministry of Health and Long-Term Care where Ontario's public health services data sets are stored, linked and used for research. We obtained information on neighbourhood-level income (in quintiles) based on the postal-code files, and on previous diagnoses of hypertension [21], diabetes [22], chronic obstructive pulmonary disease (COPD) [23], congestive heart failure (CHF) [24] and atrial fibrillation [25] using validated case definitions (e-table 1).

Exposure and outcomes

Our exposure of interest was immigration status obtained from the Ministry of Immigration, Refugee and Citizenship (IRCC) Permanent Resident Database which collected information on all immigrants who arrived in Ontario after 1985. As information on immigration status was only available after 1985, we classified individuals born outside of Canada who arrived in Ontario after 1985 as *immigrants*, and those born in Canada or those who were born outside of Canada but arrived before 1985 as *long-term residents*.

Our primary outcome was death from any cause, which was obtained from the death registry along with the date of death. We set the end date of follow-up as March 31, 2018.

We determined each person's date of last contact with the health system by using administrative databases to identify any contact with health care system such as a visit to a doctor's office, refill of prescriptions (in those over 65 years), hospitalization or emergency visits, receipt of home care, or admission to a rehabilitation facility (e-table 2) until January 31, 2020, the latest date for which information from the administrative databases was available. The health care system contact could be for any reason, and not pertaining to the index diagnosis alone. Those who were not recorded as dying prior to March 31, 2018 (end date of follow-up), and who had their last health system contact prior to this date were flagged as *lost to follow-up* at the date of last health system contact (e-Figure 1).

Statistical analyses

Analyses were conducted separately in each disease cohort. We compared baseline characteristics between immigrants and long-term residents within each disease cohort using the Chi-squared test for categorical variables and the Wilcoxon rank sum test for continuous variables.

We used the time of the index diagnosis as time zero. We estimated unadjusted cumulative incidence functions for death and loss to follow-up in immigrants and long-term residents, separately. We developed multivariable cause-specific hazards models to estimate the adjusted hazard ratio (HR) of loss to follow-up in immigrants compared to long-term residents accounting for death as a competing event, and adjusting for age, sex, neighbourhood-level income, hypertension, diabetes, COPD, CHF and atrial fibrillation.

We then fit two multivariable Cox proportional hazards models to estimate the adjusted HR of death in immigrants compared to long-term residents, adjusting for demographic information and

chronic conditions as before. In the first model, which did not account for loss to follow-up, we censored individuals only on March 31, 2018. In the second model, which accounted for loss to follow-up, we censored individuals on the first of either March 31, 2018, or the date of last health system contact.

We then calculated a ratio of the two adjusted HRs obtained from these two models and calculated 95% confidence limits around this ratio using percentile-based bootstrapping methods and 1000 bootstrap samples. If the confidence limits for the ratio included 1, it would suggest that there is no statistical difference between the adjusted HRs obtained with and without accounting for loss to follow-up. The direction and magnitude of the difference between two HRs can be inferred based on the ratio, with values under 1 suggesting overestimation of the association between immigration status and mortality when not accounting for loss to follow-up. We similarly obtained adjusted HRs of death for each covariate in the multivariable models using two separate models, with and without accounting for loss to follow-up. Using the methods described above, we also evaluated whether the association between other covariates and mortality changed after accounting for loss to follow-up. All analyses were conducted using SAS 9.4 Copyright © 2002-2012 by SAS Institute Inc., Cary, NC, USA. We did not involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

Sensitivity analyses

We re-defined the date of last health care system contact as the recorded date plus 180 days to account for patients who may not interact with health care system for up to 6 months (e-figure 2) and then re-calculated the adjusted hazard of death accounting for loss to follow-up for each disease cohort. We chose a *lag-time* of 6 months because all patients in this study had a chronic condition that would typically require follow-up within this time frame.

To evaluate how the association between immigration status and mortality would change if those lost to follow-up had died, we re-calculated the adjusted hazard of death in immigrants compared to long-term residents in two hypothetical scenarios in which patients, irrespective of their immigration status, were considered to have died at 30 days or 1 year following their last recorded health system contact.

RESULTS

The total study sample included 389,777 people (9.7% immigrants). Of these 24,557 had ischemic stroke, 310,529 had cancer and 54,691 had schizophrenia (Figure 1). A greater proportion of patients with schizophrenia were immigrants (17.4%) compared to those with ischemic stroke (8.5%) or cancer (8.4%) (Table 1). Irrespective of the underlying diagnosis, immigrants were younger at the time of the diagnosis and more likely to reside in a low-income neighbourhood compared to long-term residents (Table 1). Other characteristics of the study cohorts are shown in Table 1 and e-table 3.

During a median follow-up of 7 years, 13,667 people (3.5%) were lost to follow-up across the three disease cohorts. A greater proportion of patients with schizophrenia were lost to follow-up (9.1%) than patients with ischemic stroke (2.7%) or cancer (2.6%) (Table 2). Immigrants were more likely than long-term residents to be lost to follow up in all disease cohorts (Table 2 and Figure 2); however, the magnitude of association between immigration status and loss to follow-up was greater in patients with ischemic stroke (HR 2.87; 95% CI 2.38-3.44) and cancer (HR 3.07; 95% CI 2.91-3.23) than schizophrenia (HR 1.54; 95% CI 1.44-1.64) (Table 2).

During 2.7 million person-years of follow-up, 176,301 deaths were recorded across the three disease cohorts. The crude mortality rate was highest in patients with ischemic stroke (95.3 per 1000-person-years) followed by cancer (76.8 per 1000-person-year) and schizophrenia (13.7 per 1000-person years). In all three disease cohorts, the unadjusted hazard of mortality was

lower in immigrants compared to long-term residents (Table 2 and Figure 2). This remained true even after adjusting for baseline differences in age, comorbidity, and area-level socio-economic status, with an adjusted HR of death in immigrants compared to long-term residents of 0.78 (95% CI 0.73-0.84) in patients with ischemic stroke, a HR of 0.74 (95% CI 0.73-0.76) in patients with cancer, and a HR of 0.54 (95% CI 0.50-0.59) in patients with schizophrenia (Table 2). The magnitude of the mortality advantage in immigrants compared to long-term residents attenuated after accounting for loss to follow-up, with adjusted HR of death in immigrants compared to long-term residents of 0.83 (95% CI 0.77-0.89) for ischemic stroke, 0.78 (95% CI 0.76-0.79) for cancer, and 0.56 (95% CI 0.51-0.61) for schizophrenia (Table 2).

The ratio of the two adjusted HRs obtained using models with and without accounting for loss to follow-up was 0.95 [95% confidence limits (CL), 0.93 to 0.97] for ischemic stroke, 0.96 (95% CL, 0.95 to 0.96) for cancer and 0.97 (95% CL, 0.96 to 0.98) for schizophrenia, suggesting that not accounting for loss to follow-up overestimated the mortality advantage in immigrants in all cohorts (Figure 3). The effect of not accounting for loss to follow-up on the association between other covariates and mortality is shown in Figure 3.

Sensitivity analyses

Using a lag-time of 6 months in determining the date of last health care system contact, to account for patients who have less frequent contact with the health care system, did not alter the association between immigration status and mortality for any disease cohort (e-table 4).

In hypothetical scenarios in which, irrespective of immigration status, patients lost to follow-up were considered to be dead at 30 days and 1 year after loss to follow-up, the healthy immigrant advantage was eliminated in patients with schizophrenia, and attenuated in patients with ischemic stroke and cancer (Table 2).

DISCUSSION

In this study using linked population-based data on over 380,000 patients with a new diagnosis of ischemic stroke, cancer or schizophrenia, we demonstrated that immigrants have a survival advantage but are also more likely to be lost to follow-up compared to long-term residents, with variations in the magnitude of both the mortality advantage and the proportion lost to follow-up across the disease groups studied. Not accounting for loss to follow-up overestimated the immigrant health advantage.

Our finding of lower mortality in immigrants compared to long-term residents with stroke, cancer and schizophrenia is consistent with previous studies, including a large-scale meta-analysis of over 15.2 million immigrants across 92 countries. [13]. Potential explanations for lower mortality in immigrants include self-selection of immigrants based on health prior to migration [26], a healthier lifestyle in immigrants [27], and return migration [28]. We found that immigrants with schizophrenia had the greatest mortality advantage compared to those with ischemic stroke or cancer. Possible explanations include the relatively younger age of immigrants and long-term residents with schizophrenia, variations in disease-specific health care provision in immigrants compared to long-term residents [29,30], or other unmeasured confounders. While certain immigrant subgroups such as refugees or asylum seekers may be at increased risk of poor mental health outcomes [31] and mortality [32], the magnitude of the mortality advantage in immigrants with schizophrenia observed in our study is consistent with previous reports of lower suicide rates in immigrants compared to long-term residents across different ethnic groups in the US [33] and in Canadian youth [34].

In our study, loss to follow-up could be explained by either emigration from the province or by failure to access the health care system while remaining in the province. Since the medical conditions included in our cohorts typically require ongoing care, it is likely that emigration rather than failure to access the health care system accounts for the majority of the loss to follow-up in our study. Although our study does not provide information on the ultimate destination of those

emigrating, return to a home country and family supports at the end of life (the so-called salmon effect) has been described in immigrants with chronic conditions with physical health care needs such as ischemic stroke or cancer [35,36]. In contrast, those with schizophrenia may have less contact with the healthcare system because of their relatively young age or because of challenges in access related to mental illness, and may be less likely to return to their home country because of stigma related to mental health diagnoses in some countries of origin [37,38]. Our study did not allow us to determine whether loss to follow-up varied with disease severity, and previous studies have yielded inconsistent findings. For example, higher comorbidity in Denmark was associated with lower rates of emigration in immigrants whereas self-reported poor health in the US was associated with higher rates of emigration in Mexican immigrants [39,40].

Because immigration status was directly related to the censoring event, loss to follow-up, we found that accounting for loss to follow-up altered the magnitude of the association between immigration status and mortality. Thus, previous estimates of the mortality advantage in immigrants that have relied on death statistics alone and did not account for loss to follow-up may have overestimated the immigrant health advantage [3,12,41]. Consistent with this, a study from England and Wales found that although there was an immigrant mortality advantage, the magnitude of the association between immigration status and mortality was lower in all three hypothetical scenarios of immigrants' exits out of the country [42]. We found that accounting for loss to follow-up did not change the magnitude of the association between mortality and other variables of interest included in the multivariable models, except for older age in the ischemic stroke and cancer cohorts. This suggests that studies using administrative health data to evaluate the association between other covariates (sex, income or comorbidities) and mortality could yield adequate results even if they fail to account for loss to follow-up.

Our study is strengthened by the use of comprehensive administrative databases that allowed us to identify loss of health system contact in three separate chronic disease cohorts. The findings are likely to be generalizable to other jurisdictions with immigrant populations and to other disease conditions not included in this study, but the magnitude of bias may vary depending on the disease condition, health care jurisdiction, and immigrant-related variables (country of origin, time since immigration or immigration class).

Some limitations merit discussion. We were only able to define people as immigrants based on their immigration records, and because these were collected systematically only after 1985, immigrants who arrived prior to 1985 had to be classified as long-term residents. We did not have information on factors such as physical activity [43] and smoking [44], or other chronic conditions that may be associated with mortality, and we did not have information specific to each disease condition such as disease severity, disability, response to treatment, or palliative care status, all of which could influence mortality. Because we only included people with a known medical condition, we are unable to comment on patterns of loss to follow-up in healthy immigrants and long-term residents. We used area-level income as a proxy for socioeconomic status, and recognize that this may not reflect individual level income or other measures of socioeconomic status such as wealth, education, or occupation. We also assumed that loss of health system contact equated to patients leaving the health care jurisdiction rather than reflecting an excellent recovery negating the need for ongoing medical management. However, the misclassification introduced by this assumption should not vary based on immigration status. In addition, we assumed that, at least in immigrants, loss to follow-up was likely to be due to emigration to their home countries rather than to other parts of Canada or onwards to other regions of the world. A study from the IRCC found that only 9% of immigrants who landed in Ontario between 1991 and 2006 had moved to other provinces by 2006 [45]. Lastly, movement of individuals in and out of a health care jurisdiction is a dynamic process, and those who

emigrate can return. If such individuals return after the end date of follow-up, they could be falsely censored at the date of their emigration.

This study highlights the lower mortality in immigrants compared to long-term residents previously observed in other studies, but also demonstrates that inadequate handling of loss to follow-up can lead to biased estimates of the immigrant health advantage, as immigrant deaths may not be captured if immigrants return to their home region when gravely ill. Based on these findings, we recommend that future studies comparing mortality and other long-term outcomes in immigrants and non-immigrants carefully record loss to follow-up in both groups, quantify it, and account for it using appropriate methodology. When this information is not available, other measures could include use of updated postal code files during follow-up [46], measuring outcomes in the short term, or assuming specific rates of emigration based on previous reports. Future research could evaluate reasons for the variation in the magnitude of the association between immigration status and mortality based on the disease cohort, and evaluate the association between immigration-specific (immigration class, country of origin and time since immigration) and disease-specific (severity, palliative status and disease-related disability) factors and loss to follow-up.

Data sharing statement– The data used in this study is held securely in coded form at ICES. Data sharing agreements prohibit ICES from making the dataset publicly available, but access may be granted to those who meet prespecified criteria for confidential access. Please contact corresponding author for details.

Code availability – Can be made available upon request to the corresponding author.

Ethics approval – This study was approved by Research Ethics Board at Sunnybrook Health Sciences Centre, Canada (ID: 158-2017).

Patient and public involvement statement

Patients or the public WERE NOT involved in the design, or conduct, or reporting, or dissemination plans of our research

Disclaimer

The study was supported by ICES (formerly the Institute for Clinical Evaluative Sciences), which is supported by an annual grant from the Ontario Ministry of Health and Long-Term Care. Data were made available with help from the Immigration, Refugees and Citizenship Canada. Parts of this material are based on data and/or information compiled and provided by the Canadian Institute for Health Information. The analyses, opinions, results and conclusions reported in this article are those of the authors and are independent from the funding sources. No endorsement by ICES, the Ontario Ministry of Health and Long-Term Care, Canadian Institute for Health Information, or the Immigration, Refugees and Citizenship Canada is intended or should be inferred.

Contributorship statement

MVV, JF, and MKK were involved in the concept and design. MVV, PCA, JF were involved in data acquisition, and analysis. FLS and MKK were involved in the primary data acquisition data

for Ontario Stroke Registry. MVV, JF, PA, MCC, AL, FLS and MKK were involved in developing the project and interpreting the results. MVV was responsible for drafting the manuscript which was critically revised by everyone. MKK supervised the study and is the guarantor.

Funding

This study was funded from Heart and Stroke Foundation of Canada (HSFC GIA19-26333). MKK and PCA hold Mid-Career Investigator Awards from the Heart and Stroke Foundation of Canada. MKK holds the Lillian Love Chair in Women's Health from the University Health Network/University of Toronto. MVV holds a Fellowship Award from the Canadian Institutes of Health Research.

Competing interests – none.

References

1 Vang Z, Sigouin J, Flenon A, *et al.* The Healthy Immigrant Effect in Canada: A Systematic Review. *Population Change and Lifecourse Strategic Knowledge Cluster Discussion Paper Series/ Un Réseau stratégique de connaissances Changements de population et parcours de vie Document de travail* 2015;**3**.http://ir.lib.uwo.ca/pclc/vol3/iss1/4

2 Syse A, Dzamarija MT, Kumar BN, *et al.* An observational study of immigrant mortality differences in Norway by reason for migration, length of stay and characteristics of sending countries. *BMC Public Health* 2018;**18**:508. doi:10.1186/s12889-018-5435-4

3 Omariba DWR, Ng E, Vissandjée B. Differences between immigrants at various durations of residence and host population in all-cause mortality, Canada 1991-2006. *Population Studies* 2014;**68**:339–57. doi:10.1080/00324728.2014.915050

4 McKeigue PM, Marmot MG. Mortality from coronary heart disease in Asian communities in London. *BMJ* 1988;**297**:903. doi:10.1136/bmj.297.6653.903

5 Cheung MC, Earle CC, Fischer HD, *et al.* Impact of Immigration Status on Cancer Outcomes in Ontario, Canada. *J Oncol Pract* 2017;**13**:e602–12. doi:10.1200/JOP.2016.019497

6 Tu JV, Chu A, Donovan LR, *et al.* The Cardiovascular Health in Ambulatory Care Research Team (CANHEART). *Circulation: Cardiovascular Quality and Outcomes* 2015;**8**:204–12. doi:10.1161/CIRCOUTCOMES.114.001416

7 Paszat L, Sutradhar R, Liu Y, *et al.* Risk of colorectal cancer among immigrants to Ontario, Canada. *BMC Gastroenterol* 2017;**17**:85. doi:10.1186/s12876-017-0642-5

8 Anderson KK, Cheng J, Susser E, *et al.* Incidence of psychotic disorders among first-generation immigrants and refugees in Ontario. *CMAJ* 2015;**187**:E279–86. doi:10.1503/cmaj.141420

9 Marmot MG, Adelstein AM, Bulusu L. Lessons from the study of immigrant mortality. *Lancet* 1984;**1**:1455–7.

10 McDonald JT, Kennedy S. Insights into the 'healthy immigrant effect': health status and health service use of immigrants to Canada. *Social Science & Medicine (1982)* 2004;**59**:1613–27. doi:10.1016/j.socscimed.2004.02.004

11 Markides KS, Eschbach K. Aging, migration, and mortality: current status of research on the Hispanic paradox. *J Gerontol B Psychol Sci Soc Sci* 2005;**60 Spec No 2**:68–75. doi:10.1093/geronb/60.special_issue_2.s68

12 Khan AM, Urquia M, Kornas K, *et al.* Socioeconomic gradients in all-cause, premature and avoidable mortality among immigrants and long-term residents using linked death records in Ontario, Canada. *J Epidemiol Community Health* 2017;**71**:625–32. doi:10.1136/jech-2016-208525

13 Aldridge RW, Nellums LB, Bartlett S, *et al.* Global patterns of mortality in international migrants: a systematic review and meta-analysis. *The Lancet* 2018;**392**:2553–66. doi:10.1016/S0140-6736(18)32781-8

14 Akl EA, Briel M, You JJ, *et al.* Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ* 2012;**344**. doi:10.1136/bmj.e2809

15 Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol* 1977;**106**:184–7. doi:10.1093/oxfordjournals.aje.a112451

16 Ontario Demographic Quarterly: Highlights of first quarter 2020 | Ontario.ca. https://www.ontario.ca/page/ontario-demographic-quarterly-highlights-first-quarter-2020 (accessed 26 Apr 2021).

17 Government of Canada SC. The Daily — Immigration and ethnocultural diversity: Key results from the 2016 Census. 2017.http://www.statcan.gc.ca/daily-quotidien/171025/dq171025b-eng.htm (accessed 26 Oct 2017).

18 Kapral MK, Hall RE, Silver FL, *et al.* Registry of the Canadian Stroke Network. Report on the 2004/05 Ontario Stroke Audit. https://www.ices.on.ca/~/media/Files/Atlases-Reports/2009/RCSN-report-on-2004-05-Ontario-stroke-audit/Full-report.ashx

19 Hall S, Schulze K, Groome P, *et al.* Using cancer registry data for survival studies: the example of the Ontario Cancer Registry. *J Clin Epidemiol* 2006;**59**:67–76. doi:10.1016/j.jclinepi.2005.05.001

20 Kurdyak P, Lin E, Green D, *et al.* Validation of a Population-Based Algorithm to Detect Chronic Psychotic Illness. *Can J Psychiatry* 2015;**60**:362–8. doi:10.1177/070674371506000805

21 Tu K, Chen Z, Lipscombe LL. Prevalence and incidence of hypertension from 1995 to 2005: a population-based study. *CMAJ* 2008;**178**:1429–35. doi:10.1503/cmaj.071283

22 Hux JE, Ivis F, Flintoft V, *et al.* Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002;**25**:512–6.

23 Gershon AS, Wang C, Guan J, *et al.* Identifying individuals with physcian diagnosed COPD in health administrative databases. *COPD* 2009;**6**:388–94.

24 Schultz SE, Rothwell DM, Chen Z, *et al.* Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic Dis Inj Can* 2013;**33**:160–6.

25 Tu K, Nieuwlaat R, Cheng SY, *et al.* Identifying Patients With Atrial Fibrillation in Administrative Data. *Canadian Journal of Cardiology* 2016;**32**:1561–5. doi:10.1016/j.cjca.2016.06.006

26 Osypuk TL, Alonso A, Bates LM. Understanding the Healthy Immigrant Effect and Cardiovascular Disease: Looking to Big Data and Beyond. *Circulation* 2015;**132**:1522–4. doi:10.1161/CIRCULATIONAHA.115.018707

27 Yi SS, Thorpe LE, Zanowiak JM, *et al.* Clinical Characteristics and Lifestyle Behaviors in a Population-Based Sample of Chinese and South Asian Immigrants With Hypertension. *Am J Hypertens* 2016;**29**:941–7. doi:10.1093/ajh/hpw014

28 Abraído-Lanza AF, Dohrenwend BP, Ng-Mak DS, *et al.* The Latino mortality paradox: a test of the 'salmon bias' and healthy migrant hypotheses. *Am J Public Health* 1999;**89**:1543–8. doi:10.2105/ajph.89.10.1543

29 Yarnell CJ, Fu L, Manuel D, *et al.* Association Between Immigrant Status and End-of-Life Care in Ontario, Canada. *JAMA* 2017;**318**:1479–88. doi:10.1001/jama.2017.14418

30 Chen A, Kazanjian A. 516: Ethnicity, Time in Canada and Pap Testing among Recent Immigrants to British Columbia, Canada. *Am J Epidemiol* 2005;**161**:S129–S129. doi:10.1093/aje/161.Supplement_1.S129c

31 Berthold SM, Kong S, Mollica RF, *et al.* Comorbid mental and physical health and health access in Cambodian refugees in the US. *J Community Health* 2014;**39**:1045–52. doi:10.1007/s10900-014-9861-7

32 Hollander A-C. Social inequalities in mental health and mortality among refugees and other immigrants to Sweden – epidemiological studies of register data. *Glob Health Action* 2013;**6**. doi:10.3402/gha.v6i0.21059

33 Borges G, Orozco R, Rafful C, *et al.* Suicidality, ethnicity and immigration in the USA. *Psychol Med* 2012;**42**:1175–84. doi:10.1017/S0033291711002340

34 Saunders NR, Lebenbaum M, Stukel TA, *et al.* Suicide and self-harm trends in recent immigrant youth in Ontario, 1996-2012: a population-based longitudinal cohort study. *BMJ Open* 2017;**7**:e014863. doi:10.1136/bmjopen-2016-014863

35 Bonifazi C, Paparusso A. Remain or return home: The migration intentions of first-generation migrants in Italy. *Population, Space and Place* 2019;**25**:e2174. doi:10.1002/psp.2174

36 Diaz CJ, Koning SM, Martinez-Donate AP. Moving Beyond Salmon Bias: Mexican Return Migration and Health Selection. *Demography* 2016;**53**:2005–30. doi:10.1007/s13524-016-0526-2

37 Saunders NR, Gill PJ, Holder L, *et al.* Use of the emergency department as a first point of contact for mental health care by immigrant youth in Canada: a population-based study. *CMAJ* 2018;**190**:E1183–91. doi:10.1503/cmaj.180277

38 Derr AS. Mental Health Service Use Among Immigrants in the United States: A Systematic Review. *Psychiatr Serv* 2016;**67**:265–74. doi:10.1176/appi.ps.201500004

39 Norredam M, Hansen OH, Petersen JH, *et al.* Remigration of migrants with severe disease: myth or reality?--a register-based cohort study. *Eur J Public Health* 2015;**25**:84–9. doi:10.1093/eurpub/cku138

40 Arenas E, Goldman N, Pebley AR, *et al.* Return Migration to Mexico: Does Health Matter? *Demography* 2015;**52**:1853–68. doi:10.1007/s13524-015-0429-7

41 Kennedy S, Kidd MP, McDonald JT, *et al.* The Healthy Immigrant Effect: Patterns and Evidence from Four Countries. *Int Migration & Integration* 2015;**16**:317–32. doi:10.1007/s12134-014-0340-x

42 Wallace M, Kulu H. Low immigrant mortality in England and Wales: a data artefact? *Soc Sci Med* 2014;**120**:100–9. doi:10.1016/j.socscimed.2014.08.032

43 Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet* 1997;**349**:1436–42. doi:10.1016/S0140-6736(96)07495-8

44 GBD 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. *Lancet* 2017;**389**:1885–906. doi:10.1016/S0140-6736(17)30819-X

45 Interprovincial Mobility of Immigrants in Canada 2006-2011 - Canada.ca. https://www.canada.ca/en/immigration-refugees-citizenship/corporate/reports-statistics/research/interprovincial-mobility-immigrants-canada-2006-2011.html (accessed 2 Mar 2020).

46 Youens D, Preen DB, Harris MN, *et al.* The importance of historical residential address information in longitudinal studies using administrative health data. *Int J Epidemiol* 2018;**47**:69–80. doi:10.1093/ije/dyx156

Figure 1. Cohort selection and follow-up.

Footnote: values in parenthesis represent proportion.

Table 1. Baseline characteristics in immigrants and long-term residents with a first-ever diagnosis of ischemic stroke, cancer or schizophrenia between 2002 and 2012 in Ontario, Canada.

	Ischemic stroke		Cancer		Schizophrenia	
	Immigrants	Long-term residents	Immigrants	Long-term residents	Immigrants	Long-term residents
	2078 (8.5)	22,479 (91.5)	26,084 (8.4)	284,445 (91.6)	9525 (17.4)	45,166 (82.6)
Female, n (%)	982 (47.3)	10,697 (47.6)	13,602 (52.1)	130,324 (45.8)	4346 (45.6)	19,943 (44.2)
Median age in years at index event (Q1-Q3)	68 (55-78)	74 (63-82)	58 (48-70)	67 (58-76)	34 (25-45)	40 (26-53)
Neighbourhood-level income, n (%)						
Lowest quintile (1st)	668 (32.1)	5043 (22.4)	7041 (27.0)	50,044 (17.6)	3803 (39.9)	13,525 (29.9)
Highest quintile (5th)	201 (9.7)	4330 (19.3)	3326 (12.8)	62,667 (22.0)	734 (7.7)	6434 (14.2)
Hypertension, n (%)	1420 (68.3)	16,046 (71.4)	11,120 (42.6)	152,177 (53.5)	1165 (12.2)	8253 (18.3)
Diabetes, n (%)	727 (35.0)	6495 (28.9)	4850 (18.6)	53,444 (18.8)	737 (7.7)	4178 (9.3)
Congestive heart failure, n (%)	258 (12.4)	3728 (16.6)	878 (3.4)	20,721 (7.3)	59 (0.6)	807 (1.8)
COPD, n (%)	111 (5.3)	2547 (11.3)	1023 (3.9)	31,745 (11.2)	60 (0.6)	1494 (3.3)
Atrial fibrillation, n (%)	243 (11.7)	3786 (16.8)	777 (3.0)	19,278 (6.8)	34 (0.4)	525 (1.2)

COPD – chronic obstructive pulmonary disease.

Table 2. Loss to follow-up and mortality in immigrants and long-term residents in Ontario, Canada.

	Ischemic stroke		Cancer		Schizophrenia	
	Immigrants	**Long-term residents**	**Immigrants**	**Long-term residents**	**Immigrants**	**Long-term residents**
N (%)	2078 (8.5)	22,479 (91.5)	26,084 (8.4)	284,445 (91.6)	9525 (17.4)	45,166 (82.6)
Loss to follow-up, n (%)	158 (7.6)	512 (2.3)	2016 (7.7)	5995 (2.1)	1238 (13.0)	3748 (8.3)
Adjusted HR of loss to follow-up (95% CI)[a] accounting for the competing risk of death	2.87 (2.38-3.44)	1.00	3.07 (2.91-3.23)	1.00	1.54 (1.44-1.64)	1.00
Death, n (%)	796 (35.4)	12,575 (55.9)	9014 (34.6)	146,723 (51.6)	546 (5.7)	6647 (14.7)
Unadjusted HR of death (95 %CI)	0.61 (0.56-0.65)	1.00	0.60 (0.59-0.62)	1.00	0.39 (0.35-0.42)	1.00
Adjusted HR[a] (95% CI) not accounting for loss to follow-up	0.78 (0.73-0.84)	1.00	0.74 (0.73-0.76)	1.00	0.54 (0.50-0.59)	1.00
Adjusted HR[a] (95% CI) accounting for loss to follow-up[b]	0.83 (0.77-0.89)	1.00	0.78 (0.76-0.79)	1.00	0.56 (0.51-0.61)	1.00
Sensitivity analyses[c]						
Death within 30 days of loss to follow-up						
Adjusted HR (95% CI)	0.93 (0.87-1.00)	1.00	0.90 (0.88-0.91)	1.00	1.00 (0.95-1.05)	1.00
Death within 1 year of loss to follow-up						
Adjusted HR (95% CI)	0.93 (0.87-0.99)	1.00	0.89 (0.87-0.91)	1.00	1.00 (0.95-1.06)	1.00

[a]Multivariable model adjusting for the following: age, sex, neighbourhood-level income, and comorbidities (known hypertension, diabetes, congestive heart failure, chronic obstructive pulmonary disease, and atrial fibrillation); [b]censoring those who were lost to follow-up, which was determined when date of last health system contact occurred before end of follow-up among those alive; [c]assigning date of death among those lost to follow-up and re-calculating adjusted hazard of death.
Abbreviations: HR – hazard ratio, CI – confidence interval.

Figure 2. Unadjusted cumulative incidence functions in immigrants (red) and long-term residents (blue) showing probabilty of death and of loss to follow-up in patients with ischemic stroke (top), primary cancer (middle) and schizophrenia (bottom).

Figure 3. Ratios of adjusted hazard ratios of death obtained using two multivariable cox-regression models with and without accounting for loss to follow-up. Each box represents the point estimate of this ratio, and the error bars represent 95% confidence limits. Values less than 1 suggest overestimation of the magnitude of association when loss to follow-up is not accounted for.

Footnote: Immigrants are compared to long-term residents; age less than 55 years is the comparison group; and the 5th quintile of income represent the HR of death in the highest quintile compared to lowest quintile based on neighbourhood-level income.

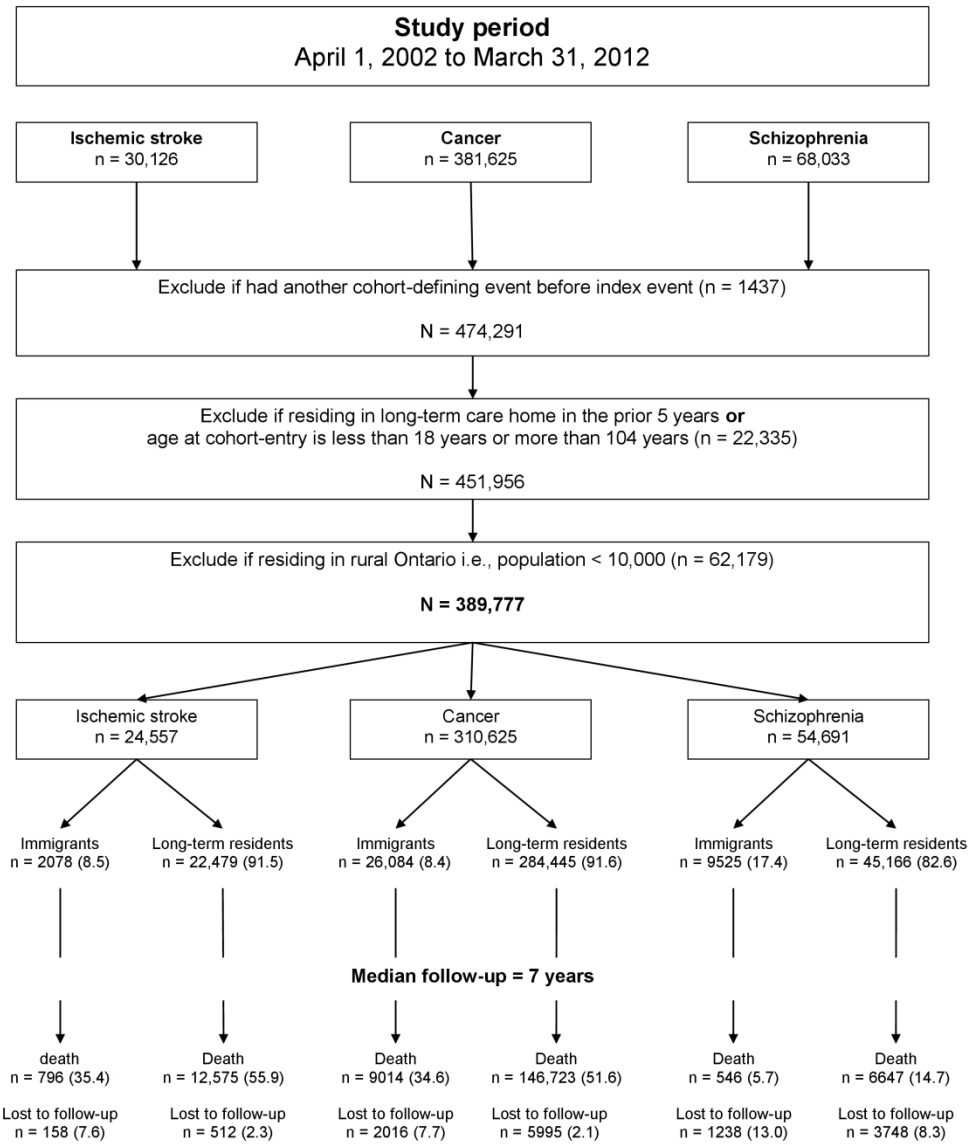

Study period
April 1, 2002 to March 31, 2012
Ischemic stroke
n = 30,126
Cancer
n = 381,625
Schizophrenia
n = 68,033
Exclude if had another cohort-defining event before index event (n = 1437)
N = 474,291
Exclude if residing in long-term care home in the prior 5 years or
age at cohort-entry is less than 18 years or more than 104 years (n = 22,335)
N = 451,956
Exclude if residing in rural Ontario i.e., population < 10,000 (n = 62,179)
N = 389,777
Ischemic stroke
n = 24,557
Cancer
n = 310,625
Schizophrenia
n = 54,691
Immigrants
n = 2078 (8.5)
Long-term residents
n = 22,479 (91.5)
Immigrants
n = 26,084 (8.4)
Long-term residents
n = 284,445 (91.6)
Immigrants
n = 9525 (17.4)
Long-term residents
n = 45,166 (82.6)
Median follow-up = 7 years
death
n = 796 (35.4)
Death
n = 12,575 (55.9)
Death
n = 9014 (34.6)
Death
n = 146,723 (51.6)
Death
n = 546 (5.7)
Death
n = 6647 (14.7)
Lost to follow-up
n = 158 (7.6)
Lost to follow-up
n = 512 (2.3)
Lost to follow-up
n = 2016 (7.7)
Lost to follow-up
n = 5995 (2.1)
Lost to follow-up
n = 1238 (13.0)
Lost to follow-up
n = 3748 (8.3)

Figure 1

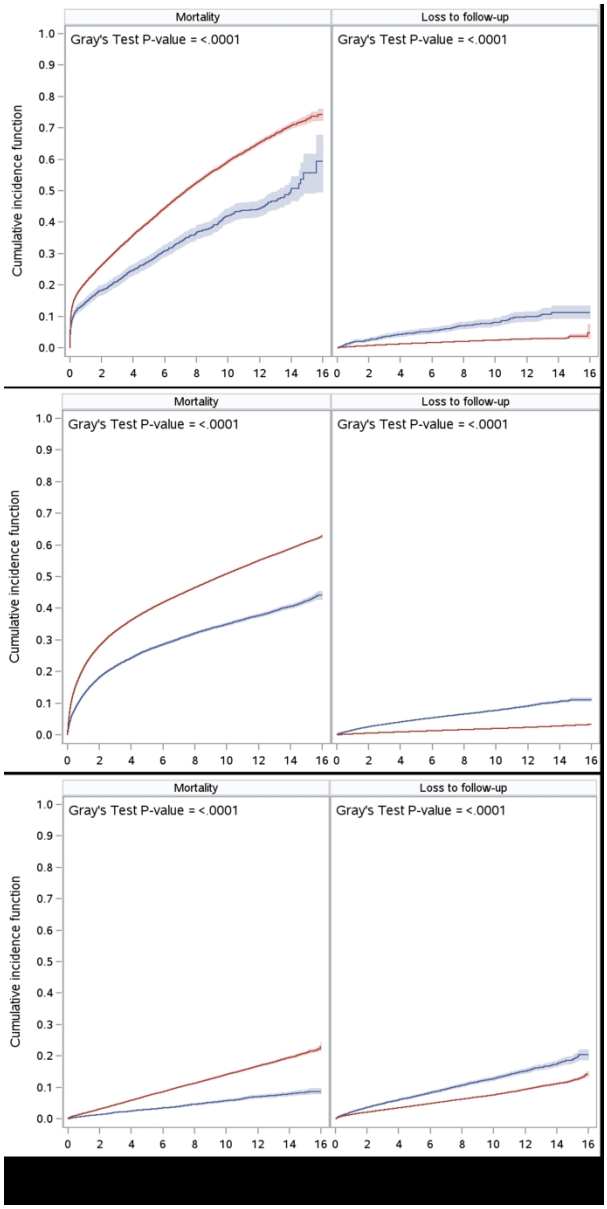

Mortality
Loss to follow-up
Gray's Test P-value = <.0001
Gray's Test P-value = <.0001
Cumulative incidence function
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
2
4
6
8
10
12
14
16

Figure 2

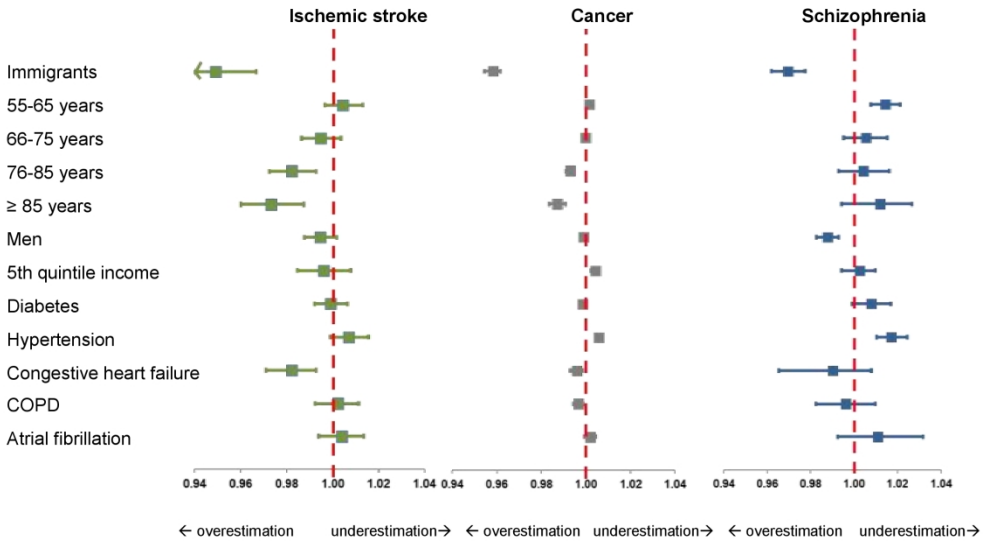

Ischemic stroke
Cancer
Schizophrenia
Immigrants
55-65 years
66-75 years
76-85 years
≥ 85 years
Men
5th quintile income
Diabetes
Hypertension
Congestive heart failure
COPD
Atrial fibrillation
0.94 0.96 0.98 1.00 1.02 1.04
← overestimation underestimation→

Figure 3

Appendix

e-table 1. Definitions of variables included in the study.

e-table 2. Administrative databases used to determine date of last health system contact and statistics on contact with health care system in Ontario.

e-table 3. Characteristics of immigrants included in the study.

e-table 4. Results of sensitivity analyses using a lag-time of 6 months when determining the date of last health system contact.

e-figure 1. Hypothetical cases to illustrate loss to follow-up using administrative database. Subject A was not lost to follow-up, Subject B would be considered lost to follow-up, and Subject C had the event of interest (death) and so is not considered lost to follow-up.

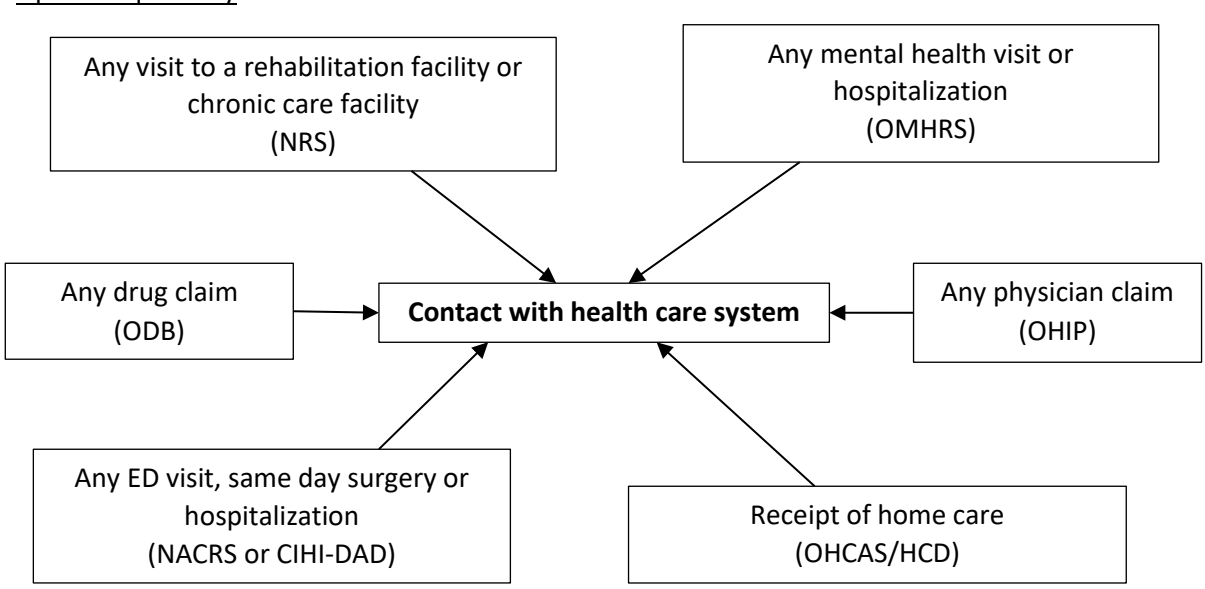
e-figure 2. Sensitivity analyses adding 180 days to last date of follow-up. Only for Subject B does addition of 180 days to follow-up change their censoring time; whereas, censoring times remain same for Subject A and C.

e-table 1. Definitions of variables included in the study.

Variable	Definition
Incident ischemic stroke	Hospitalized or non-hospitalized (seen in the emergency department but not admitted) adult patients with confirmed acute ischemic stroke between April 1, 2002 and March 31, 2013 seen at all 150 acute care institutions in the province. Participating hospitals included comprehensive stroke centres and non-stroke centres. Information gathered using chart abstractors with neurological expertise, with the final diagnosis and other data elements obtained through review of clinical and neuroimaging data.
Incident primary cancer	A diagnosis of cancer either in hospitalized or non-hospitalized adult patients obtained from 4 different sources: hospital or ER visit using appropriate ICD codes, pathology reports with a diagnosis of cancer, regional cancer centres where patients with cancer are seen, and death certificates.
Incident schizophrenia	A primary diagnosis of schizophrenia or schizoaffective disorder from a general hospital bed (prior to 2002, ICD9 - 295; as of 2002 ICD10 - F20 or F25) **OR** primary diagnosis of schizophrenia from a psychiatric hospital bed (DSM-IV – 295.x) **OR** three outpatient visits with a diagnosis of schizophrenia (295 or F20/F25) from outpatient physician billings within a 3-year period. 93.1% Sensitivity - 58.7% Specificity

Hypertension	≥ 1 Hospitalization [add diagnostic codes to be consistent with DM section? Same for other variables.] **OR** ≥ 2 physician claims in a two-year period **OR** 1 physician claim followed by another physician claim or hospitalization within two years. 72% Sensitivity - 95% Specificity - 87% PPV - 88% NPV
Diabetes	≥ 3 physician claims for diagnostic code (250) in a one-year period 79.9% Sensitivity - 99.1% Specificity - 91.4% PPV
CHF (congestive heart failure)	≥ 1 Hospitalization **OR** 1 physician claim in ER or clinic, followed by ≥ 1 Hospitalization, ER visit, or physician claim within one year. 84.8% Sensitivity - 97.0% Specificity - 55.6% PPV
Atrial fibrillation	1 hospitalization or 1 emergency room visit, ICD-10 (2002 onwards) – I48; ICD-9 (pre-2002) – 427.31 or 427.32 **OR** Technical billing code for cardioversion billing code Z437
COPD (chronic obstructive pulmonary disease)	≥1 Hospitalization for COPD OR ≥ 3 physician claims in a two-year period 57.5% Sensitivity - 95.4% Specificity

e-table 2. Administrative databases used to determine date of last health system contact and statistics on contact with health care system in Ontario.



Abbreviations:

NRS – National Rehabilitation Reporting System; ODB – Ontario Drug Benefit; NACRS – National Ambulatory Care Reporting System; CIHI-DAD – Canadian Institute for Health Information-Discharge Abstract Database; OHCAS – Ontario Home Care Administration System; HCD – Home Care Database; OHIP – Ontario Health Insurance Plan Claims Database; OMHRS – Ontario Mental Health Reporting System. ED – Emergency Department

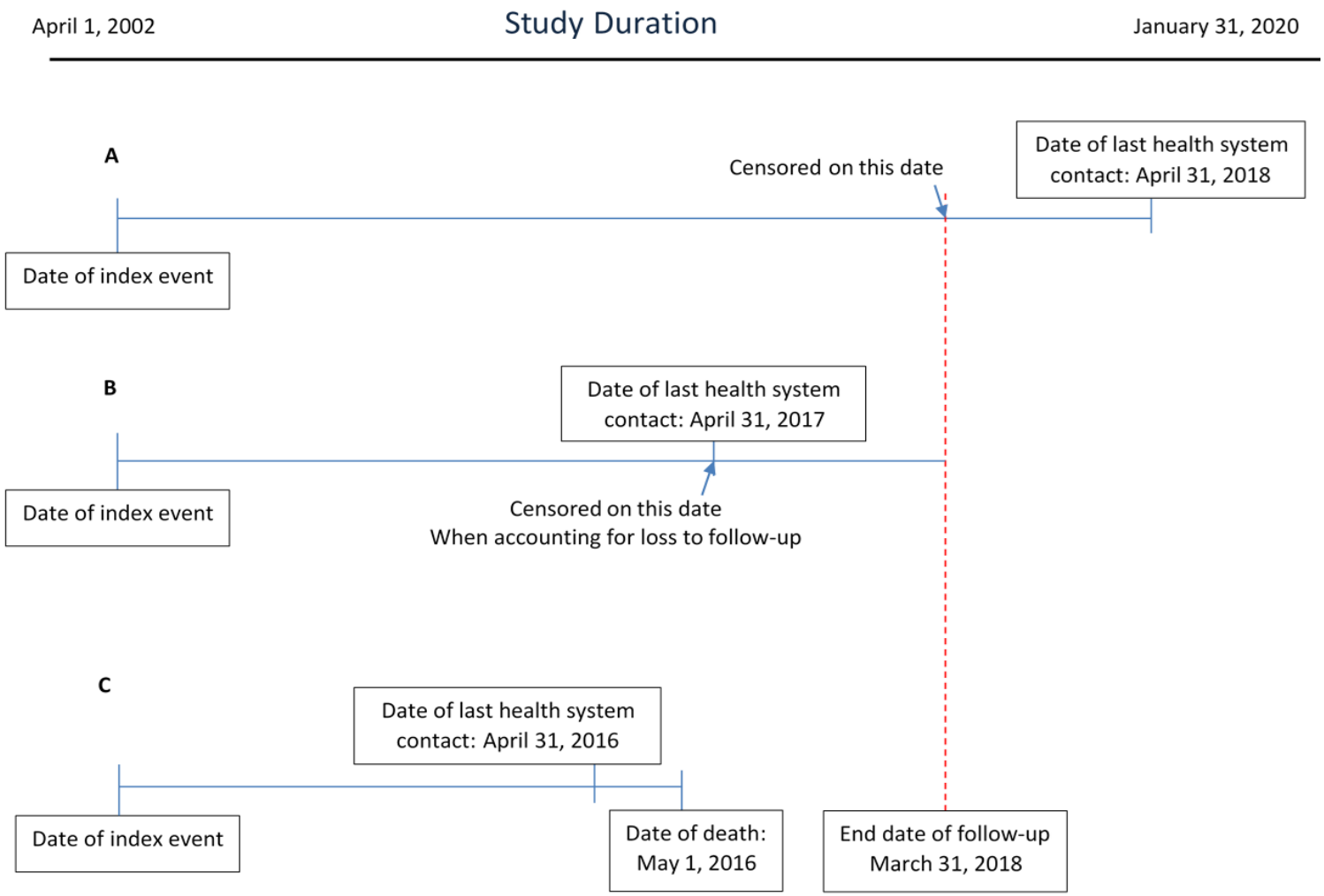
e-table 3. Characteristics of immigrants included in the study.

Immigration-specific characteristics	**Ischemic stroke n = 2078**	**Cancer n = 26,084**	**Schizophrenia n = 9525**
World region of origin			
Africa	81 (3.9)	1128 (4.3)	980 (10.3)
Caribbean	193 (9.3)	1902 (7.3)	967 (10.2)
East Asia	403 (19.4)	5331 (20.4)	1319 (13.8)
Latin America	171 (8.2)	1736 (6.7)	738 (7.7)
Middle East	194 (9.3)	2701 (10.4)	947 (9.9)
South Asia	392 (18.9)	4094 (15.7)	2271 (23.8)
Western	526 (25.3)	7525 (28.8)	1895 (19.9)
Missing	392 (18.9)	1667 (6.4)	408 (4.3)
Time since arrival			
≤ 10 years	677 (32.6)	10360 (39.7)	4763 (50.0)
> 10 years	1401 (67.4)	15724 (60.3)	4762 (50.0)
Immigration class			
Economic	468 (22.5)	9262 (35.5)	3213 (33.7)
Family or other	1273 (61.3)	13233 (50.7)	3891 (37.8)
Refugee	337 (16.2)	3589 (13.8)	2421 (25.4)

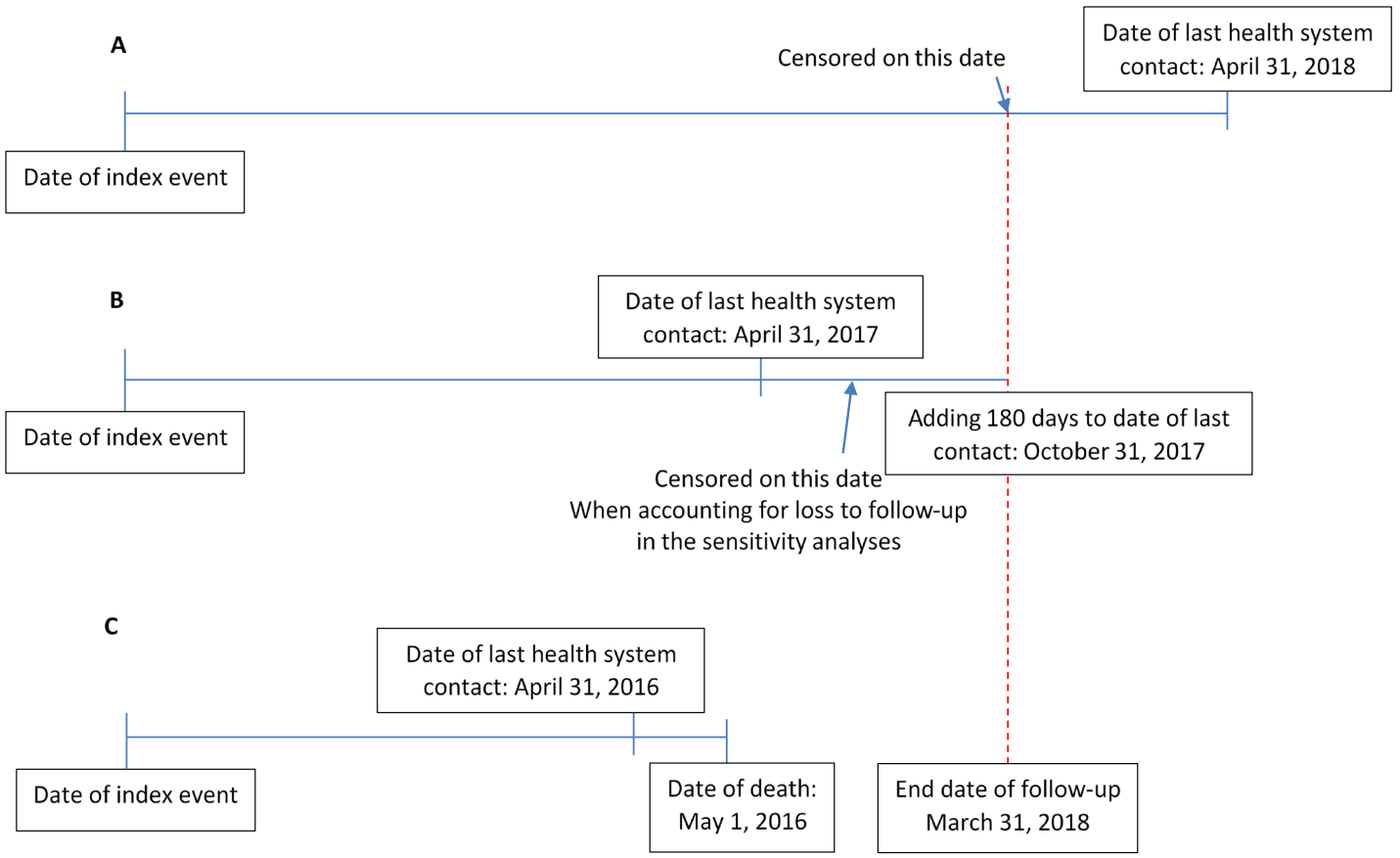
e-table 4. Results of sensitivity analyses using a lag-time of 6 months when determining the date of last health system contact.

		Ischemic stroke	Cancer	Schizophrenia
Immigrants	Lost to follow-up, n (%)	145 (7.0)	1895 (7.3)	1120 (11.8)
Long-term residents		472 (2.1)	5472 (1.9)	3333 (7.4)
Immigrants	Death, n (%)	796 (35.4)	9014 (34.6)	546 (5.7)
Long-term residents		12,575 (55.9)	146,723 (51.6)	6647 (14.7)
Adjusted HR of death (95% CI)^	Immigrants vs. long-term residents	0.82 (0.77-0.89)	0.77 (0.76-0.79)	0.56 (0.51-0.61)

^Accounting for loss of follow-up by censoring those lost to follow-up.



e-figure 1. Hypothetical cases to illustrate loss to follow-up using administrative database. Subject A was not lost to follow-up, Subject B would be considered lost to follow-up, and Subject C had the event of interest (death) and so is not considered lost to follow-up.



e-figure 2. Sensitivity analyses adding 180 days to last date of follow-up. Only for Subject B does addition of 180 days to follow-up change their censoring time; whereas, censoring times remain same for Subject A and C.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Location
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 4
Methods			
Study design	4	Present key elements of study design early in the paper	Page 5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 5 and e-table 2
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	n.a.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 5 and e-table 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Page 5 and e-table 1
Bias	9	Describe any efforts to address potential sources of bias	Page 6
Study size	10	Explain how the study size was arrived at	Figure 1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Table 1
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Page 6-7
		(*b*) Describe any methods used to examine subgroups and interactions	Page 6-7
		(*c*) Explain how missing data were addressed	None present
		(*d*) If applicable, explain how loss to follow-up was addressed	e-table 2 and page 7
		(*e*) Describe any sensitivity analyses	Page 8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Figure 1
		(b) Give reasons for non-participation at each stage	Figure 1

		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	None with missing data
		(c) Summarise follow-up time (eg, average and total amount)	Page 8
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 9 and Table 2
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Page 9, and table 2
		(*b*) Report category boundaries when continuous variables were categorized	Table 2
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n.a.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Figure 3
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Page 12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Page 10
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

The importance of accounting for loss to follow-up when comparing mortality between immigrants and long-term residents: a population-based retrospective cohort

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-046377.R2
Article Type:	Original research
Date Submitted by the Author:	27-Sep-2021
Complete List of Authors:	Vyas, Manav; University of Toronto, Division of Neurology Fang, Jiming; ICES Austin, Peter; ICES Laupacis, Andreas; University of Toronto, Department of Medicine Cheung, Matthew; University of Toronto, Division of Haematology & Oncology, Department of Medicine Silver, Frank ; University of Toronto, Division of Neurology, Department of Medicine Kapral, MK; University of Toronto, Division of General Internal Medicine, Department of Medicine
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Health services research
Keywords:	EPIDEMIOLOGY, SOCIAL MEDICINE, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

The importance of accounting for loss to follow-up when comparing mortality between immigrants and long-term residents: a population-based retrospective cohort

Manav V. Vyas[1,2,3], Jiming Fang[3], Peter C. Austin[2,3], Andreas Laupacis[2,3,4] Matthew C. Cheung[2,3,5], Frank L. Silver[1,3] and Moira K. Kapral[2,3,4]

1. Division of Neurology, Department of Medicine, University of Toronto, Canada
2. Institute of Health Policy, Management and Evaluation, University of Toronto, Canada
3. ICES, Toronto, Canada
4. Division of General Internal Medicine, Department of Medicine, University of Toronto, Canada
5. Division of Haematology & Oncology, Department of Medicine, University of Toronto, Canada

Abstract count: 247 words

Running Title: Immigration, Emigration, and Mortality

Word count: 3,414 words (not including abstract, references, tables, and figures)

Character count of the title: 160 (with spaces)

Number of references: 39

Keywords: immigration, emigration, mortality, lost to follow-up, emigration.

Figures: 3

Tables: 2

Corresponding author:

Dr Manav V. Vyas, MBBS MSc PhD FRCPC

Staff Neurologist, St. Michael's Hospital-Unity Health Toronto

Assistant Professor, Department of Medicine, Temerty Faculty of Medicine, University of Toronto

Office: Li Ka Shing Knowledge Institute, Rm 628, 250 Yonge Street, Toronto, ON M5B 2L7

Email: manav.vyas@mail.utoronto.ca | Phone: 416-864-5056 | Fax: 416-864-5712

ABSTRACT

Objectives: To evaluate the association between immigration status and all-cause mortality in different disease cohorts, and the impact of loss to follow-up on the observed associations.

Design: Population-based retrospective cohort study using linked administrative health data in Ontario, Canada.

Setting: We followed adults with a first-ever diagnosis of ischemic stroke, cancer, or schizophrenia between 2002 and 2013 from index event to death, loss to follow-up, or end of follow-up in 2018.

Primary and secondary outcome measures: Our outcomes of interest were all-cause mortality and loss to follow-up. For each disease cohort, we calculated adjusted hazard ratios (HR) of death in immigrants compared to long-term residents, adjusting for demographic characteristics and comorbidities, with and without censoring for those who were lost to follow-up. We calculated the ratio of two the HRs and the respective confidence limits (CL) using bootstrapping methods.

Results: Immigrants were more likely to be lost to follow-up than long-term residents in all disease cohorts. Not accounting for this loss to follow-up overestimated the magnitude of the association between immigration status and mortality in those with ischemic stroke (HR of death before vs. after accounting for censoring: 0.78 vs. 0.83, ratio = 0.95; 95% CL, 0.93-0.97), cancer (0.74 vs. 0.78, ratio = 0.96; 0.95-0.96), and schizophrenia (0.54 vs. 0.56, ratio = 0.97; 0.96-0.98).

Conclusions: Immigrants to Canada have a survival advantage that varies by the disease studied. The magnitude of this advantage is modestly overestimated by not accounting for the higher loss to follow-up in immigrants.

ARTICLE SUMMARY

Strengths and limitations of this study

- This is the first study in Canada to compare the rates of loss to follow-up in a population-based sample of immigrants and long-term residents with stroke, cancer, or schizophrenia.
- Using appropriate statistical analyses, we compared the hazard of mortality, adjusted for confounders, between immigrants and long-term residents with and without accounting for loss to follow-up.
- Loss to follow-up was determined using administrative data definitions which may not be complete.
- Long-term residents consisted of Canadian born and those who migrated before 1985, limiting generalizability of the findings to all immigrants.

INTRODUCTION

Studies from high-income countries, including Canada, have shown that immigrants have a lower mortality compared to host populations [1–3]. This immigrant health advantage has been observed not only for all-cause mortality, but also for cardiovascular [4] and cancer mortality [5], and for the incidence of non-communicable chronic conditions such as cardiovascular disease [6], cancer [7] and schizophrenia [8].

This phenomenon has been termed the *healthy immigrant effect*, and is partly attributed to a selection bias whereby only those who are healthy and have human capital are able to migrate [9,10]. However, another potential explanation for the observed immigrant health advantage is a phenomenon termed the *salmon effect,* whereby immigrants return to their home countries when they are gravely ill [11]. Thus, they are lost to follow-up and not accounted for in studies that rely on mortality statistics that do not record emigration [12]. Such lack of complete follow-up was identified in a large-scale meta-analysis on immigrant mortality in which none of the included 96 studies accounted for loss to follow-up and only 29 (28%) studies identified loss to follow-up as an issue [13]. Previous work in both observational studies and randomized controlled trials has shown that unbalanced loss to follow-up in two comparison groups can lead to biased estimates of association [14,15].

The aims of this study were to use linked population-based data from Ontario, Canada to evaluate the association between immigration status and all-cause mortality after a new diagnosis of ischemic stroke, cancer, or schizophrenia, to quantify loss to follow-up in immigrants compared to long-term residents, and to determine how accounting for loss to follow-up influences the association between immigration status and all-cause mortality in each disease cohort.

METHODS

Patient and public involvement statement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research

Setting and population

Ontario is Canada's most populous province, with an estimated population of 14 million [16]. In 2016, approximately 3.8 million immigrants were living in Ontario, accounting for about 30% of the total population [17]. All residents of Ontario (except for undocumented migrants and those visiting) are covered by the provincial health plan that includes physician services, hospital and emergency care, and investigations ordered by physicians. The linked administrative databases in Ontario capture all health care system contacts of insured Ontario residents.

We created three separate incident disease cohorts of ischemic stroke, primary cancer and schizophrenia using validated case definitions for incidence during an ascertainment period of April 1, 2002 to March 31, 2012 (e-table 1). We focused on these three diseases because they are chronic conditions associated with high morbidity that require regular health care system contact (which is captured in administrative databases) following the diagnosis, and because management of these conditions may benefit from family and social supports, which can be a factor in emigration. We identified patients with ischemic stroke using the Ontario Stroke Registry which is a province-wide registry that includes data on a random sample of consecutive patients seen at over 150 hospitals in Ontario [18]. Data collection for the registry was performed by chart abstractors with neurological expertise, with the final diagnosis and other data elements obtained through review of clinical and neuroimaging data. We identified patients with a diagnosis of a primary malignant cancer from the Ontario Cancer Registry, a population-based registry, which is created by combining information from discharge and day surgery summaries, pathology reports with any mention of cancer, or records of patients referred to specialized

institutions treating cancer patients in Ontario [19]. It captures approximately 95% of all cancer diagnoses in the province [19]. We identified patients with schizophrenia based on a validated algorithm whereby a diagnosis of schizophrenia was made if the patient had one or more hospital admissions and/or three or more outpatient visits with a diagnosis of schizophrenia or schizoaffective disorder [20].

Within each disease cohort, we excluded prevalent cases if they had a diagnosis of the specific disease prior to April 1, 2002. If patients had multiple cohort-defining events during the ascertainment period, only information at the time of the first cohort-defining event was recorded. We excluded patients who were younger than 18 years or older than 104 years at the time of the index event, those who resided in long-term care homes at the time of the index event, and those who resided in rural areas (population < 10,000) because most immigrants (> 95%) reside in large urban areas.

Using unique identifiers, we linked these cohorts to population-based data held securely at ICES (formerly known as the Institute for Clinical Evaluative Sciences), Toronto. ICES is a prescribed entity under the Ministry of Health and Long-Term Care where Ontario's public health services data sets are stored, linked, and used for research. We obtained information on neighbourhood-level income (in quintiles) based on the postal-code files, and on previous diagnoses of hypertension [21], diabetes [22], chronic obstructive pulmonary disease (COPD) [23], congestive heart failure (CHF) [24] and atrial fibrillation [25] using validated case definitions (e-table 1).

Exposure and outcomes

Our exposure of interest was immigration status obtained from the Ministry of Immigration, Refugee and Citizenship (IRCC) Permanent Resident Database which collected information on all immigrants who arrived in Ontario after 1985. As information on immigration status was only available after 1985, we classified individuals born outside of Canada who arrived in Ontario

after 1985 as *immigrants*, and those born in Canada or those who were born outside of Canada but arrived before 1985 as *long-term residents*.

Our primary outcome was death from any cause, which was obtained from the death registry along with the date of death. We set the end date of follow-up as March 31, 2018.

We determined each person's date of last contact with the health system by using administrative databases to identify any contact with health care system such as a visit to a doctor's office, refill of prescriptions (in those over 65 years), hospitalization or emergency visits, receipt of home care, or admission to a rehabilitation facility (e-table 2) until January 31, 2020, the latest date for which information from the administrative databases was available. The health care system contact could be for any reason, and not pertaining to the index diagnosis alone. Those who were not recorded as dying prior to March 31, 2018 (end date of follow-up), and who had their last health system contact prior to this date were flagged as *lost to follow-up* at the date of last health system contact (e-Figure 1).

Statistical analyses

Analyses were conducted separately in each disease cohort. We compared baseline characteristics between immigrants and long-term residents within each disease cohort using the Chi-squared test for categorical variables and the Wilcoxon rank sum test for continuous variables.

We used the time of the index diagnosis as time zero. We estimated unadjusted cumulative incidence functions for death and loss to follow-up in immigrants and long-term residents, separately. We developed multivariable cause-specific hazards models to estimate the adjusted hazard ratio (HR) of loss to follow-up in immigrants compared to long-term residents accounting for death as a competing event, and adjusting for age, sex, neighbourhood-level income, hypertension, diabetes, COPD, CHF, and atrial fibrillation.

We then fit two multivariable Cox proportional hazards models to estimate the adjusted HR of death in immigrants compared to long-term residents, adjusting for demographic information and chronic conditions as before. In the first model, which did not account for loss to follow-up, we censored individuals only on March 31, 2018. In the second model, which accounted for loss to follow-up, we censored individuals on the first of either March 31, 2018, or the date of last health system contact.

We then calculated a ratio of the two adjusted HRs obtained from these two models and calculated 95% confidence limits around this ratio using percentile-based bootstrapping methods and 1000 bootstrap samples. If the confidence limits for the ratio included 1, it would suggest that there is no statistical difference between the adjusted HRs obtained with and without accounting for loss to follow-up. The direction and magnitude of the difference between two HRs can be inferred based on the ratio, with values under 1 suggesting overestimation of the association between immigration status and mortality when not accounting for loss to follow-up. We similarly obtained adjusted HRs of death for each covariate in the multivariable models using two separate models, with and without accounting for loss to follow-up. Using the methods described above, we also evaluated whether the association between other covariates and mortality changed after accounting for loss to follow-up. All analyses were conducted using SAS 9.4 Copyright © 2002-2012 by SAS Institute Inc., Cary, NC, USA. We did not involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

Sensitivity analyses

We re-defined the date of last health care system contact as the recorded date plus 180 days to account for patients who may not interact with health care system for up to 6 months (e-figure 2) and then re-calculated the adjusted hazard of death accounting for loss to follow-up for each disease cohort. We chose a *lag-time* of 6 months because all patients in this study had a chronic condition that would typically require follow-up within this time frame.

To evaluate how the association between immigration status and mortality would change if those lost to follow-up had died, we re-calculated the adjusted hazard of death in immigrants compared to long-term residents in two hypothetical scenarios in which patients, irrespective of their immigration status, were considered to have died at 30 days or 1 year following their last recorded health system contact.

RESULTS

The total study sample included 389,777 people (9.7% immigrants). Of these 24,557 had ischemic stroke, 310,529 had cancer and 54,691 had schizophrenia (Figure 1). A greater proportion of patients with schizophrenia were immigrants (17.4%) compared to those with ischemic stroke (8.5%) or cancer (8.4%) (Table 1). Irrespective of the underlying diagnosis, immigrants were younger at the time of the diagnosis and more likely to reside in a low-income neighbourhood compared to long-term residents (Table 1). Other characteristics of the study cohorts are shown in Table 1 and e-table 3.

During a median follow-up of 7 years, 13,667 people (3.5%) were lost to follow-up across the three disease cohorts. A greater proportion of patients with schizophrenia were lost to follow-up (9.1%) than patients with ischemic stroke (2.7%) or cancer (2.6%) (Table 2). Immigrants were more likely than long-term residents to be lost to follow up in all disease cohorts (Table 2 and Figure 2); however, the magnitude of association between immigration status and loss to follow-up was greater in patients with ischemic stroke (HR 2.87; 95% CI 2.38-3.44) and cancer (HR 3.07; 95% CI 2.91-3.23) than schizophrenia (HR 1.54; 95% CI 1.44-1.64) (Table 2).

During 2.7 million person-years of follow-up, 176,301 deaths were recorded across the three disease cohorts. The crude mortality rate was highest in patients with ischemic stroke (95.3 per 1000-person-years) followed by cancer (76.8 per 1000-person-year) and schizophrenia (13.7 per 1000-person years). In all three disease cohorts, the unadjusted hazard of mortality was

lower in immigrants compared to long-term residents (Table 2 and Figure 2). This remained true even after adjusting for baseline differences in age, comorbidity, and area-level socio-economic status, with an adjusted HR of death in immigrants compared to long-term residents of 0.78 (95% CI 0.73-0.84) in patients with ischemic stroke, a HR of 0.74 (95% CI 0.73-0.76) in patients with cancer, and a HR of 0.54 (95% CI 0.50-0.59) in patients with schizophrenia (Table 2). The magnitude of the mortality advantage in immigrants compared to long-term residents attenuated after accounting for loss to follow-up, with adjusted HR of death in immigrants compared to long-term residents of 0.83 (95% CI 0.77-0.89) for ischemic stroke, 0.78 (95% CI 0.76-0.79) for cancer, and 0.56 (95% CI 0.51-0.61) for schizophrenia (Table 2).

The ratio of the two adjusted HRs obtained using models with and without accounting for loss to follow-up was 0.95 [95% confidence limits (CL), 0.93 to 0.97] for ischemic stroke, 0.96 (95% CL, 0.95 to 0.96) for cancer and 0.97 (95% CL, 0.96 to 0.98) for schizophrenia, suggesting that not accounting for loss to follow-up overestimated the mortality advantage in immigrants in all cohorts (Figure 3). This is equivalent to a relative change in the HR of death (in immigrants vs. long-term residents) of 5% for ischemic stroke, 4% for cancer and 3% for schizophrenia. The effect of not accounting for loss to follow-up on the association between other covariates and mortality is shown in Figure 3.

Sensitivity analyses

Using a lag-time of 6 months in determining the date of last health care system contact, to account for patients who have less frequent contact with the health care system, did not alter the association between immigration status and mortality for any disease cohort (e-table 4).

In hypothetical scenarios in which, irrespective of immigration status, patients lost to follow-up were considered to be dead at 30 days and 1 year after loss to follow-up, the healthy immigrant

advantage was eliminated in patients with schizophrenia and attenuated in patients with ischemic stroke and cancer (Table 2).

DISCUSSION

In this study using linked population-based data on over 380,000 patients with a new diagnosis of ischemic stroke, cancer, or schizophrenia, we demonstrated that immigrants have a survival advantage but are also more likely to be lost to follow-up compared to long-term residents, with variations in the magnitude of both the mortality advantage and the proportion lost to follow-up across the disease groups studied. Not accounting for loss to follow-up overestimated the immigrant health advantage.

Our finding of lower mortality in immigrants compared to long-term residents with stroke, cancer and schizophrenia is consistent with previous studies, including a large-scale meta-analysis of over 15.2 million immigrants across 92 countries. [13]. Potential explanations for lower mortality in immigrants include self-selection of immigrants based on health prior to migration [26], a healthier lifestyle in immigrants [27], and return migration [28]. We found that immigrants with schizophrenia had the greatest mortality advantage compared to those with ischemic stroke or cancer. Possible explanations include the relatively younger age of immigrants and long-term residents with schizophrenia, variations in disease-specific health care provision in immigrants compared to long-term residents [29,30], or other unmeasured confounders. While certain immigrant subgroups such as refugees or asylum seekers may be at increased risk of poor mental health outcomes [31] and mortality [32], the magnitude of the mortality advantage in immigrants with schizophrenia observed in our study is consistent with previous reports of lower suicide rates in immigrants compared to long-term residents across different ethnic groups in the US [33] and in Canadian youth [34].

In our study, loss to follow-up could be explained by either emigration from the province or by failure to access the health care system while remaining in the province. Since the medical conditions included in our cohorts typically require ongoing care, it is likely that emigration rather than failure to access the health care system accounts for the majority of the loss to follow-up in our study. Although our study does not provide information on the ultimate destination of those emigrating, return to a home country and family supports at the end of life (the so-called salmon effect) has been described in immigrants with chronic conditions with physical health care needs such as ischemic stroke or cancer [35,36]. In contrast, those with schizophrenia may have less contact with the healthcare system because of their relatively young age or because of challenges in access related to mental illness, and may be less likely to return to their home country because of stigma related to mental health diagnoses in some countries of origin [37,38]. Our study did not allow us to determine whether loss to follow-up varied with disease severity, and previous studies have yielded inconsistent findings. For example, higher comorbidity in Denmark was associated with lower rates of emigration in immigrants whereas self-reported poor health in the US was associated with higher rates of emigration in Mexican immigrants [39,40].

Because immigration status was directly related to the censoring event, loss to follow-up, we found that accounting for loss to follow-up altered the magnitude of the association between immigration status and mortality. Thus, previous estimates of the mortality advantage in immigrants that have relied on death statistics alone and did not account for loss to follow-up may have overestimated the immigrant health advantage [3,12,41]. Consistent with this, a study from England and Wales found that although there was an immigrant mortality advantage, the magnitude of the association between immigration status and mortality was lower in all three hypothetical scenarios of immigrants' exits out of the country [42]. We found that accounting for loss to follow-up did not change the magnitude of the association between mortality and other

variables of interest included in the multivariable models, except for older age in the ischemic stroke and cancer cohorts. This suggests that studies using administrative health data to evaluate the association between other covariates (sex, income or comorbidities) and mortality could yield adequate results even if they fail to account for loss to follow-up.

Our study is strengthened by the use of comprehensive administrative databases that allowed us to identify loss of health system contact in three separate chronic disease cohorts. The findings are likely to be generalizable to other jurisdictions with immigrant populations and to other disease conditions not included in this study, but the magnitude of bias may vary depending on the disease condition, health care jurisdiction, and immigrant-related variables (country of origin, time since immigration or immigration class).

Some limitations merit discussion. We were only able to define people as immigrants based on their immigration records, and because these were collected systematically only after 1985, immigrants who arrived prior to 1985 had to be classified as long-term residents. We did not have information on factors such as physical activity [43] and smoking [44], or other chronic conditions that may be associated with mortality, and we did not have information specific to each disease condition such as disease severity, disability, response to treatment, or palliative care status, all of which could influence mortality. Because we only included people with a known medical condition, we are unable to comment on patterns of loss to follow-up in healthy immigrants and long-term residents. We used area-level income as a proxy for socioeconomic status, and recognize that this may not reflect individual level income or other measures of socioeconomic status such as wealth, education, or occupation. We also assumed that loss of health system contact equated to patients leaving the health care jurisdiction rather than reflecting an excellent recovery negating the need for ongoing medical management. However, the misclassification introduced by this assumption should not vary based on immigration status. In addition, we assumed that, at least in immigrants, loss to follow-up was likely to be due to

emigration to their home countries rather than to other parts of Canada or onwards to other regions of the world. A study from the IRCC found that only 9% of immigrants who landed in Ontario between 1991 and 2006 had moved to other provinces by 2006 [45]. Lastly, movement of individuals in and out of a health care jurisdiction is a dynamic process, and those who emigrate can return. If such individuals return after the end date of follow-up, they could be falsely censored at the date of their emigration.

This study highlights the lower mortality in immigrants compared to long-term residents previously observed in other studies, but also demonstrates that inadequate handling of loss to follow-up can lead to biased estimates of the immigrant health advantage, as immigrant deaths may not be captured if immigrants return to their home region when gravely ill. Based on these findings, we recommend that future studies comparing mortality and other long-term outcomes in immigrants and non-immigrants carefully record loss to follow-up in both groups, quantify it, and account for it using appropriate methodology. When this information is not available, other measures could include use of updated postal code files during follow-up [46], measuring outcomes in the short term, or assuming specific rates of emigration based on previous reports. Future research could evaluate reasons for the variation in the magnitude of the association between immigration status and mortality based on the disease cohort, and evaluate the association between immigration-specific (immigration class, country of origin and time since immigration) and disease-specific (severity, palliative status and disease-related disability) factors and loss to follow-up.

Data sharing statement– The data used in this study is held securely in coded form at ICES. Data sharing agreements prohibit ICES from making the dataset publicly available, but access may be granted to those who meet prespecified criteria for confidential access. Please contact corresponding author for details.

Code availability – Can be made available upon request to the corresponding author.

Ethics approval – This study was approved by Research Ethics Board at Sunnybrook Health Sciences Centre, Canada (ID: 158-2017).

Disclaimer

The study was supported by ICES (formerly the Institute for Clinical Evaluative Sciences), which is supported by an annual grant from the Ontario Ministry of Health and Long-Term Care. Data were made available with help from the Immigration, Refugees and Citizenship Canada. Parts of this material are based on data and/or information compiled and provided by the Canadian Institute for Health Information. The analyses, opinions, results and conclusions reported in this article are those of the authors and are independent from the funding sources. No endorsement by ICES, the Ontario Ministry of Health and Long-Term Care, Canadian Institute for Health Information, or the Immigration, Refugees and Citizenship Canada is intended or should be inferred.

Contributorship statement

MVV, JF, and MKK were involved in the concept and design. MVV, PCA, JF were involved in data acquisition, and analysis. FLS and MKK were involved in the primary data acquisition data for Ontario Stroke Registry. MVV, JF, PA, MCC, AL, FLS and MKK were involved in developing the project and interpreting the results. MVV was responsible for drafting the manuscript which was critically revised by everyone. MKK supervised the study and is the guarantor.

Funding

This study was funded from Heart and Stroke Foundation of Canada (HSFC GIA19-26333). MKK and PCA hold Mid-Career Investigator Awards from the Heart and Stroke Foundation of Canada. MKK holds the Lillian Love Chair in Women's Health from the University Health Network/University of Toronto. MVV holds a Fellowship Award from the Canadian Institutes of Health Research.

Competing interests – none.

References

1 Vang Z, Sigouin J, Flenon A, *et al.* The Healthy Immigrant Effect in Canada: A Systematic Review. *Population Change and Lifecourse Strategic Knowledge Cluster Discussion Paper Series/ Un Réseau stratégique de connaissances Changements de population et parcours de vie Document de travail* 2015;**3**.http://ir.lib.uwo.ca/pclc/vol3/iss1/4

2 Syse A, Dzamarija MT, Kumar BN, *et al.* An observational study of immigrant mortality differences in Norway by reason for migration, length of stay and characteristics of sending countries. *BMC Public Health* 2018;**18**:508. doi:10.1186/s12889-018-5435-4

3 Omariba DWR, Ng E, Vissandjée B. Differences between immigrants at various durations of residence and host population in all-cause mortality, Canada 1991-2006. *Population Studies* 2014;**68**:339–57. doi:10.1080/00324728.2014.915050

4 McKeigue PM, Marmot MG. Mortality from coronary heart disease in Asian communities in London. *BMJ* 1988;**297**:903. doi:10.1136/bmj.297.6653.903

5 Cheung MC, Earle CC, Fischer HD, *et al.* Impact of Immigration Status on Cancer Outcomes in Ontario, Canada. *J Oncol Pract* 2017;**13**:e602–12. doi:10.1200/JOP.2016.019497

6 Tu JV, Chu A, Donovan LR, *et al.* The Cardiovascular Health in Ambulatory Care Research Team (CANHEART). *Circulation: Cardiovascular Quality and Outcomes* 2015;**8**:204–12. doi:10.1161/CIRCOUTCOMES.114.001416

7 Paszat L, Sutradhar R, Liu Y, *et al.* Risk of colorectal cancer among immigrants to Ontario, Canada. *BMC Gastroenterol* 2017;**17**:85. doi:10.1186/s12876-017-0642-5

8 Anderson KK, Cheng J, Susser E, *et al.* Incidence of psychotic disorders among first-generation immigrants and refugees in Ontario. *CMAJ* 2015;**187**:E279–86. doi:10.1503/cmaj.141420

9 Marmot MG, Adelstein AM, Bulusu L. Lessons from the study of immigrant mortality. *Lancet* 1984;**1**:1455–7.

10 McDonald JT, Kennedy S. Insights into the 'healthy immigrant effect': health status and health service use of immigrants to Canada. *Social Science & Medicine (1982)* 2004;**59**:1613–27. doi:10.1016/j.socscimed.2004.02.004

11 Markides KS, Eschbach K. Aging, migration, and mortality: current status of research on the Hispanic paradox. *J Gerontol B Psychol Sci Soc Sci* 2005;**60 Spec No 2**:68–75. doi:10.1093/geronb/60.special_issue_2.s68

12 Khan AM, Urquia M, Kornas K, *et al.* Socioeconomic gradients in all-cause, premature and avoidable mortality among immigrants and long-term residents using linked death records in Ontario, Canada. *J Epidemiol Community Health* 2017;**71**:625–32. doi:10.1136/jech-2016-208525

13 Aldridge RW, Nellums LB, Bartlett S, *et al.* Global patterns of mortality in international migrants: a systematic review and meta-analysis. *The Lancet* 2018;**392**:2553–66. doi:10.1016/S0140-6736(18)32781-8

14 Akl EA, Briel M, You JJ, *et al.* Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ* 2012;**344**. doi:10.1136/bmj.e2809

15 Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol* 1977;**106**:184–7. doi:10.1093/oxfordjournals.aje.a112451

16 Ontario Demographic Quarterly: Highlights of first quarter 2020 | Ontario.ca. https://www.ontario.ca/page/ontario-demographic-quarterly-highlights-first-quarter-2020 (accessed 26 Apr 2021).

17 Government of Canada SC. The Daily — Immigration and ethnocultural diversity: Key results from the 2016 Census. 2017.http://www.statcan.gc.ca/daily-quotidien/171025/dq171025b-eng.htm (accessed 26 Oct 2017).

18 Kapral MK, Hall RE, Silver FL, *et al.* Registry of the Canadian Stroke Network. Report on the 2004/05 Ontario Stroke Audit. https://www.ices.on.ca/~/media/Files/Atlases-Reports/2009/RCSN-report-on-2004-05-Ontario-stroke-audit/Full-report.ashx

19 Hall S, Schulze K, Groome P, *et al.* Using cancer registry data for survival studies: the example of the Ontario Cancer Registry. *J Clin Epidemiol* 2006;**59**:67–76. doi:10.1016/j.jclinepi.2005.05.001

20 Kurdyak P, Lin E, Green D, *et al.* Validation of a Population-Based Algorithm to Detect Chronic Psychotic Illness. *Can J Psychiatry* 2015;**60**:362–8. doi:10.1177/070674371506000805

21 Tu K, Chen Z, Lipscombe LL. Prevalence and incidence of hypertension from 1995 to 2005: a population-based study. *CMAJ* 2008;**178**:1429–35. doi:10.1503/cmaj.071283

22 Hux JE, Ivis F, Flintoft V, *et al.* Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002;**25**:512–6.

23 Gershon AS, Wang C, Guan J, *et al.* Identifying individuals with physcian diagnosed COPD in health administrative databases. *COPD* 2009;**6**:388–94.

24 Schultz SE, Rothwell DM, Chen Z, *et al.* Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic Dis Inj Can* 2013;**33**:160–6.

25 Tu K, Nieuwlaat R, Cheng SY, *et al.* Identifying Patients With Atrial Fibrillation in Administrative Data. *Canadian Journal of Cardiology* 2016;**32**:1561–5. doi:10.1016/j.cjca.2016.06.006

26 Osypuk TL, Alonso A, Bates LM. Understanding the Healthy Immigrant Effect and Cardiovascular Disease: Looking to Big Data and Beyond. *Circulation* 2015;**132**:1522–4. doi:10.1161/CIRCULATIONAHA.115.018707

27 Yi SS, Thorpe LE, Zanowiak JM, *et al.* Clinical Characteristics and Lifestyle Behaviors in a Population-Based Sample of Chinese and South Asian Immigrants With Hypertension. *Am J Hypertens* 2016;**29**:941–7. doi:10.1093/ajh/hpw014

28 Abraído-Lanza AF, Dohrenwend BP, Ng-Mak DS, *et al.* The Latino mortality paradox: a test of the 'salmon bias' and healthy migrant hypotheses. *Am J Public Health* 1999;**89**:1543–8. doi:10.2105/ajph.89.10.1543

29 Yarnell CJ, Fu L, Manuel D, *et al.* Association Between Immigrant Status and End-of-Life Care in Ontario, Canada. *JAMA* 2017;**318**:1479–88. doi:10.1001/jama.2017.14418

30 Chen A, Kazanjian A. 516: Ethnicity, Time in Canada and Pap Testing among Recent Immigrants to British Columbia, Canada. *Am J Epidemiol* 2005;**161**:S129–S129. doi:10.1093/aje/161.Supplement_1.S129c

31 Berthold SM, Kong S, Mollica RF, *et al.* Comorbid mental and physical health and health access in Cambodian refugees in the US. *J Community Health* 2014;**39**:1045–52. doi:10.1007/s10900-014-9861-7

32 Hollander A-C. Social inequalities in mental health and mortality among refugees and other immigrants to Sweden – epidemiological studies of register data. *Glob Health Action* 2013;**6**. doi:10.3402/gha.v6i0.21059

33 Borges G, Orozco R, Rafful C, *et al.* Suicidality, ethnicity and immigration in the USA. *Psychol Med* 2012;**42**:1175–84. doi:10.1017/S0033291711002340

34 Saunders NR, Lebenbaum M, Stukel TA, *et al.* Suicide and self-harm trends in recent immigrant youth in Ontario, 1996-2012: a population-based longitudinal cohort study. *BMJ Open* 2017;**7**:e014863. doi:10.1136/bmjopen-2016-014863

35 Bonifazi C, Paparusso A. Remain or return home: The migration intentions of first-generation migrants in Italy. *Population, Space and Place* 2019;**25**:e2174. doi:10.1002/psp.2174

36 Diaz CJ, Koning SM, Martinez-Donate AP. Moving Beyond Salmon Bias: Mexican Return Migration and Health Selection. *Demography* 2016;**53**:2005–30. doi:10.1007/s13524-016-0526-2

37 Saunders NR, Gill PJ, Holder L, *et al.* Use of the emergency department as a first point of contact for mental health care by immigrant youth in Canada: a population-based study. *CMAJ* 2018;**190**:E1183–91. doi:10.1503/cmaj.180277

38 Derr AS. Mental Health Service Use Among Immigrants in the United States: A Systematic Review. *Psychiatr Serv* 2016;**67**:265–74. doi:10.1176/appi.ps.201500004

39 Norredam M, Hansen OH, Petersen JH, *et al.* Remigration of migrants with severe disease: myth or reality?--a register-based cohort study. *Eur J Public Health* 2015;**25**:84–9. doi:10.1093/eurpub/cku138

40 Arenas E, Goldman N, Pebley AR, *et al.* Return Migration to Mexico: Does Health Matter? *Demography* 2015;**52**:1853–68. doi:10.1007/s13524-015-0429-7

41 Kennedy S, Kidd MP, McDonald JT, *et al.* The Healthy Immigrant Effect: Patterns and Evidence from Four Countries. *Int Migration & Integration* 2015;**16**:317–32. doi:10.1007/s12134-014-0340-x

42 Wallace M, Kulu H. Low immigrant mortality in England and Wales: a data artefact? *Soc Sci Med* 2014;**120**:100–9. doi:10.1016/j.socscimed.2014.08.032

43 Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet* 1997;**349**:1436–42. doi:10.1016/S0140-6736(96)07495-8

44 GBD 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. *Lancet* 2017;**389**:1885–906. doi:10.1016/S0140-6736(17)30819-X

45 Interprovincial Mobility of Immigrants in Canada 2006-2011 - Canada.ca. https://www.canada.ca/en/immigration-refugees-citizenship/corporate/reports-statistics/research/interprovincial-mobility-immigrants-canada-2006-2011.html (accessed 2 Mar 2020).

46 Youens D, Preen DB, Harris MN, *et al.* The importance of historical residential address information in longitudinal studies using administrative health data. *Int J Epidemiol* 2018;**47**:69–80. doi:10.1093/ije/dyx156

Figure 1. Cohort selection and follow-up.

Footnote: values in parenthesis represent proportion.

Table 1. Baseline characteristics in immigrants and long-term residents with a first-ever diagnosis of ischemic stroke, cancer or schizophrenia between 2002 and 2012 in Ontario, Canada.

	Ischemic stroke		Cancer		Schizophrenia	
	Immigrants	Long-term residents	Immigrants	Long-term residents	Immigrants	Long-term residents
	2078 (8.5)	22,479 (91.5)	26,084 (8.4)	284,445 (91.6)	9525 (17.4)	45,166 (82.6)
Female, n (%)	982 (47.3)	10,697 (47.6)	13,602 (52.1)	130,324 (45.8)	4346 (45.6)	19,943 (44.2)
Median age in years at index event (Q1-Q3)	68 (55-78)	74 (63-82)	58 (48-70)	67 (58-76)	34 (25-45)	40 (26-53)
Neighbourhood-level income, n (%)						
Lowest quintile (1st)	668 (32.1)	5043 (22.4)	7041 (27.0)	50,044 (17.6)	3803 (39.9)	13,525 (29.9)
Highest quintile (5th)	201 (9.7)	4330 (19.3)	3326 (12.8)	62,667 (22.0)	734 (7.7)	6434 (14.2)
Hypertension, n (%)	1420 (68.3)	16,046 (71.4)	11,120 (42.6)	152,177 (53.5)	1165 (12.2)	8253 (18.3)
Diabetes, n (%)	727 (35.0)	6495 (28.9)	4850 (18.6)	53,444 (18.8)	737 (7.7)	4178 (9.3)
Congestive heart failure, n (%)	258 (12.4)	3728 (16.6)	878 (3.4)	20,721 (7.3)	59 (0.6)	807 (1.8)
COPD, n (%)	111 (5.3)	2547 (11.3)	1023 (3.9)	31,745 (11.2)	60 (0.6)	1494 (3.3)
Atrial fibrillation, n (%)	243 (11.7)	3786 (16.8)	777 (3.0)	19,278 (6.8)	34 (0.4)	525 (1.2)

COPD – chronic obstructive pulmonary disease.

Table 2. Loss to follow-up and mortality in immigrants and long-term residents in Ontario, Canada.

	Ischemic stroke		Cancer		Schizophrenia	
	Immigrants	**Long-term residents**	**Immigrants**	**Long-term residents**	**Immigrants**	**Long-term residents**
N (%)	2078 (8.5)	22,479 (91.5)	26,084 (8.4)	284,445 (91.6)	9525 (17.4)	45,166 (82.6)
Loss to follow-up, n (%)	158 (7.6)	512 (2.3)	2016 (7.7)	5995 (2.1)	1238 (13.0)	3748 (8.3)
Adjusted HR of loss to follow-up (95% CI)[a] accounting for the competing risk of death	2.87 (2.38-3.44)	1.00	3.07 (2.91-3.23)	1.00	1.54 (1.44-1.64)	1.00
Death, n (%)	796 (35.4)	12,575 (55.9)	9014 (34.6)	146,723 (51.6)	546 (5.7)	6647 (14.7)
Unadjusted HR of death (95 %CI)	0.61 (0.56-0.65)	1.00	0.60 (0.59-0.62)	1.00	0.39 (0.35-0.42)	1.00
Adjusted HR[a] (95% CI) not accounting for loss to follow-up	0.78 (0.73-0.84)	1.00	0.74 (0.73-0.76)	1.00	0.54 (0.50-0.59)	1.00
Adjusted HR[a] (95% CI) accounting for loss to follow-up[b]	0.83 (0.77-0.89)	1.00	0.78 (0.76-0.79)	1.00	0.56 (0.51-0.61)	1.00
Sensitivity analyses[c]						
Death within 30 days of loss to follow-up						
Adjusted HR (95% CI)	0.93 (0.87-1.00)	1.00	0.90 (0.88-0.91)	1.00	1.00 (0.95-1.05)	1.00
Death within 1 year of loss to follow-up						
Adjusted HR (95% CI)	0.93 (0.87-0.99)	1.00	0.89 (0.87-0.91)	1.00	1.00 (0.95-1.06)	1.00

[a]Multivariable model adjusting for the following: age, sex, neighbourhood-level income, and comorbidities (known hypertension, diabetes, congestive heart failure, chronic obstructive pulmonary disease, and atrial fibrillation); [b]censoring those who were lost to follow-up, which was determined when date of last health system contact occurred before end of follow-up among those alive; [c]assigning date of death among those lost to follow-up and re-calculating adjusted hazard of death.
Abbreviations: HR – hazard ratio, CI – confidence interval.

Figure 2. Unadjusted cumulative incidence functions in immigrants (blue) and long-term residents (red) showing probabilty of death and of loss to follow-up in patients with ischemic stroke (top), cancer (middle) and schizophrenia (bottom).

Figure 3. Ratios of adjusted hazard ratios of death obtained using two multivariable cox-regression models with and without accounting for loss to follow-up. Each box represents the point estimate of this ratio, and the error bars represent 95% confidence limits. Values less than 1 suggest overestimation of the magnitude of association when loss to follow-up is not accounted for.

Footnote: Immigrants are compared to long-term residents; age less than 55 years is the comparison group; and the 5th quintile of income represent the HR of death in the highest quintile compared to lowest quintile based on neighbourhood-level income.

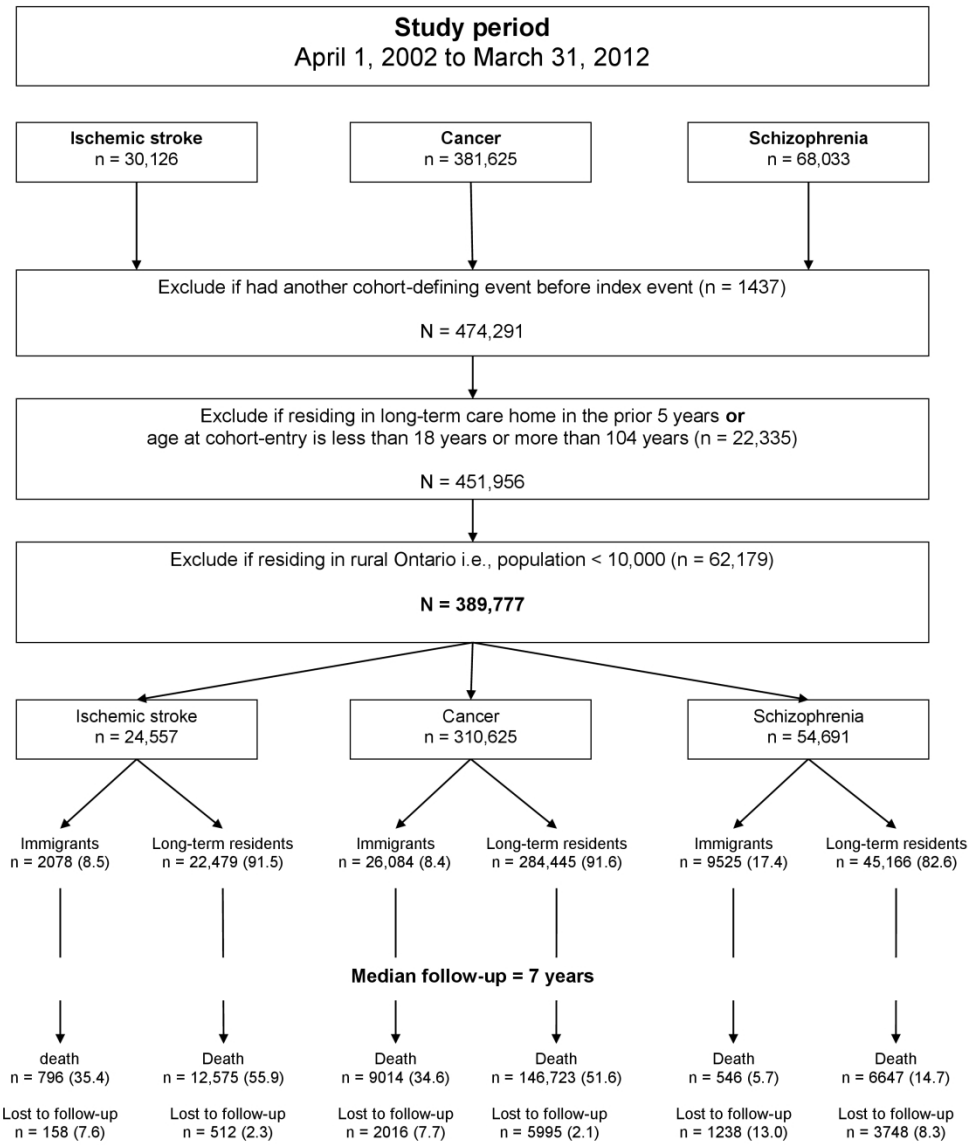

Study period
April 1, 2002 to March 31, 2012
Ischemic stroke
n = 30,126
Cancer
n = 381,625
Schizophrenia
n = 68,033
Exclude if had another cohort-defining event before index event (n = 1437)
N = 474,291
Exclude if residing in long-term care home in the prior 5 years or
age at cohort-entry is less than 18 years or more than 104 years (n = 22,335)
N = 451,956
Exclude if residing in rural Ontario i.e., population < 10,000 (n = 62,179)
N = 389,777
Ischemic stroke
n = 24,557
Cancer
n = 310,625
Schizophrenia
n = 54,691
Immigrants
n = 2078 (8.5)
Long-term residents
n = 22,479 (91.5)
Immigrants
n = 26,084 (8.4)
Long-term residents
n = 284,445 (91.6)
Immigrants
n = 9525 (17.4)
Long-term residents
n = 45,166 (82.6)
Median follow-up = 7 years
death
n = 796 (35.4)
Death
n = 12,575 (55.9)
Death
n = 9014 (34.6)
Death
n = 146,723 (51.6)
Death
n = 546 (5.7)
Death
n = 6647 (14.7)
Lost to follow-up
n = 158 (7.6)
Lost to follow-up
n = 512 (2.3)
Lost to follow-up
n = 2016 (7.7)
Lost to follow-up
n = 5995 (2.1)
Lost to follow-up
n = 1238 (13.0)
Lost to follow-up
n = 3748 (8.3)

Figure 1

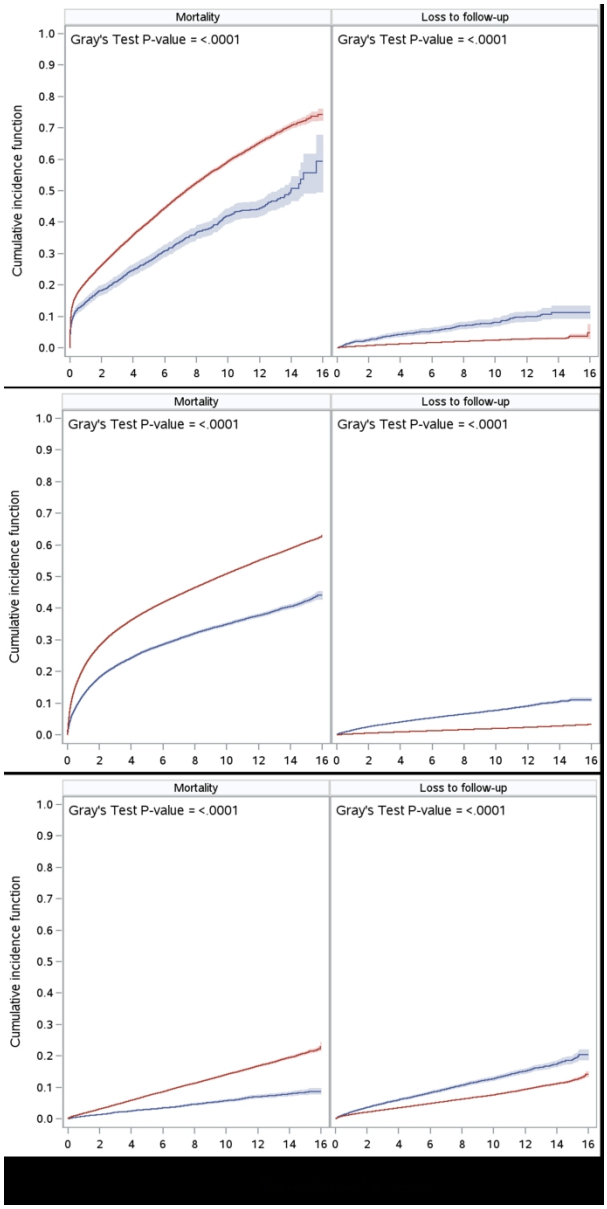

Mortality
Loss to follow-up
Gray's Test P-value = <.0001
Gray's Test P-value = <.0001
Cumulative incidence function
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
2
4
6
8
10
12
14
16

Figure 2

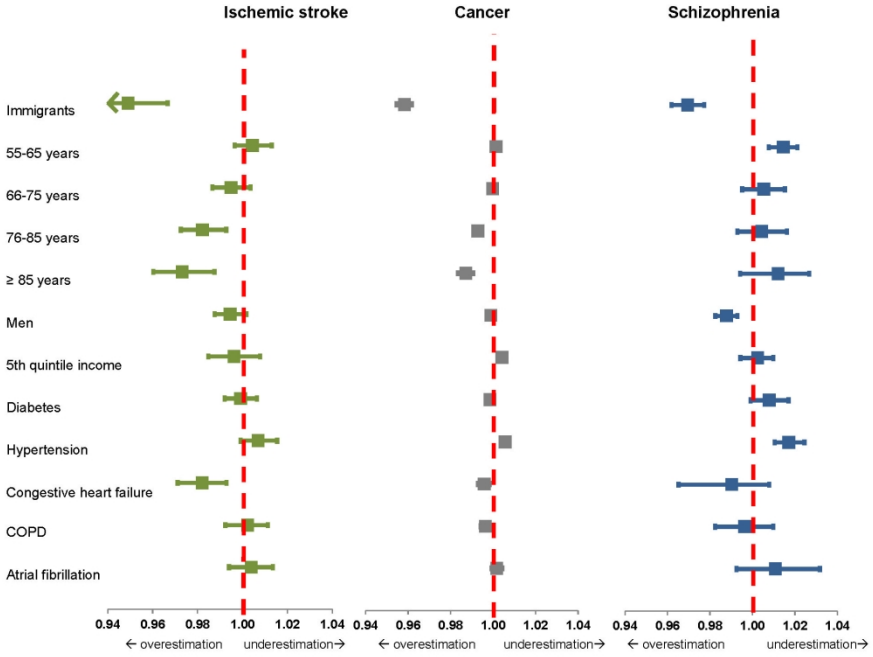


Figure 3

279x215mm (200 x 200 DPI)

Appendix

e-table 1. Definitions of variables included in the study.

e-table 2. Administrative databases used to determine date of last health system contact and statistics on contact with health care system in Ontario.

e-table 3. Characteristics of immigrants included in the study.

e-table 4. Results of sensitivity analyses using a lag-time of 6 months when determining the date of last health system contact.

e-figure 1. Hypothetical cases to illustrate loss to follow-up using administrative database. Subject A was not lost to follow-up, Subject B would be considered lost to follow-up, and Subject C had the event of interest (death) and so is not considered lost to follow-up.

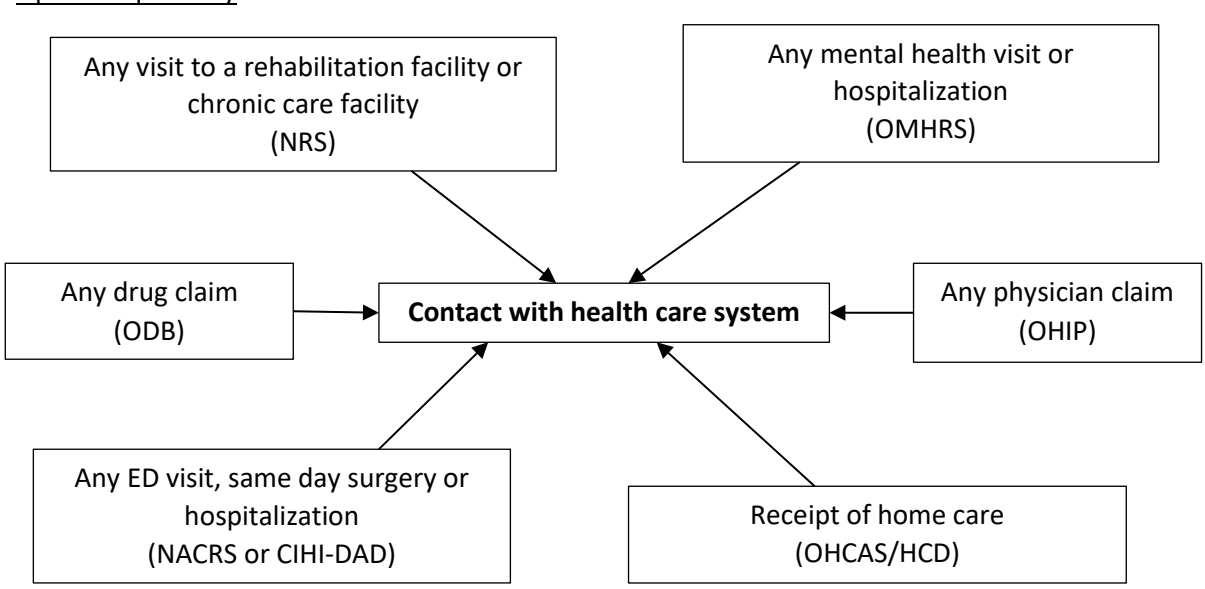
e-figure 2. Sensitivity analyses adding 180 days to last date of follow-up. Only for Subject B does addition of 180 days to follow-up change their censoring time; whereas, censoring times remain same for Subject A and C.

e-table 1. Definitions of variables included in the study.

Variable	Definition
Incident ischemic stroke	Hospitalized or non-hospitalized (seen in the emergency department but not admitted) adult patients with confirmed acute ischemic stroke between April 1, 2002 and March 31, 2013 seen at all 150 acute care institutions in the province. Participating hospitals included comprehensive stroke centres and non-stroke centres. Information gathered using chart abstractors with neurological expertise, with the final diagnosis and other data elements obtained through review of clinical and neuroimaging data.
Incident primary cancer	A diagnosis of cancer either in hospitalized or non-hospitalized adult patients obtained from 4 different sources: hospital or ER visit using appropriate ICD codes, pathology reports with a diagnosis of cancer, regional cancer centres where patients with cancer are seen, and death certificates.
Incident schizophrenia	A primary diagnosis of schizophrenia or schizoaffective disorder from a general hospital bed (prior to 2002, ICD9 - 295; as of 2002 ICD10 - F20 or F25) **OR** primary diagnosis of schizophrenia from a psychiatric hospital bed (DSM-IV – 295.x) **OR** three outpatient visits with a diagnosis of schizophrenia (295 or F20/F25) from outpatient physician billings within a 3-year period. 93.1% Sensitivity - 58.7% Specificity

Hypertension	≥ 1 Hospitalization [add diagnostic codes to be consistent with DM section? Same for other variables.] **OR** ≥ 2 physician claims in a two-year period **OR** 1 physician claim followed by another physician claim or hospitalization within two years. 72% Sensitivity - 95% Specificity - 87% PPV - 88% NPV
Diabetes	≥ 3 physician claims for diagnostic code (250) in a one-year period 79.9% Sensitivity - 99.1% Specificity - 91.4% PPV
CHF (congestive heart failure)	≥ 1 Hospitalization **OR** 1 physician claim in ER or clinic, followed by ≥ 1 Hospitalization, ER visit, or physician claim within one year. 84.8% Sensitivity - 97.0% Specificity - 55.6% PPV
Atrial fibrillation	1 hospitalization or 1 emergency room visit, ICD-10 (2002 onwards) – I48; ICD-9 (pre-2002) – 427.31 or 427.32 **OR** Technical billing code for cardioversion billing code Z437
COPD (chronic obstructive pulmonary disease)	≥1 Hospitalization for COPD OR ≥ 3 physician claims in a two-year period 57.5% Sensitivity - 95.4% Specificity

e-table 2. Administrative databases used to determine date of last health system contact and statistics on contact with health care system in Ontario.



Abbreviations:

NRS – National Rehabilitation Reporting System; ODB – Ontario Drug Benefit; NACRS – National Ambulatory Care Reporting System; CIHI-DAD – Canadian Institute for Health Information-Discharge Abstract Database; OHCAS – Ontario Home Care Administration System; HCD – Home Care Database; OHIP – Ontario Health Insurance Plan Claims Database; OMHRS – Ontario Mental Health Reporting System. ED – Emergency Department

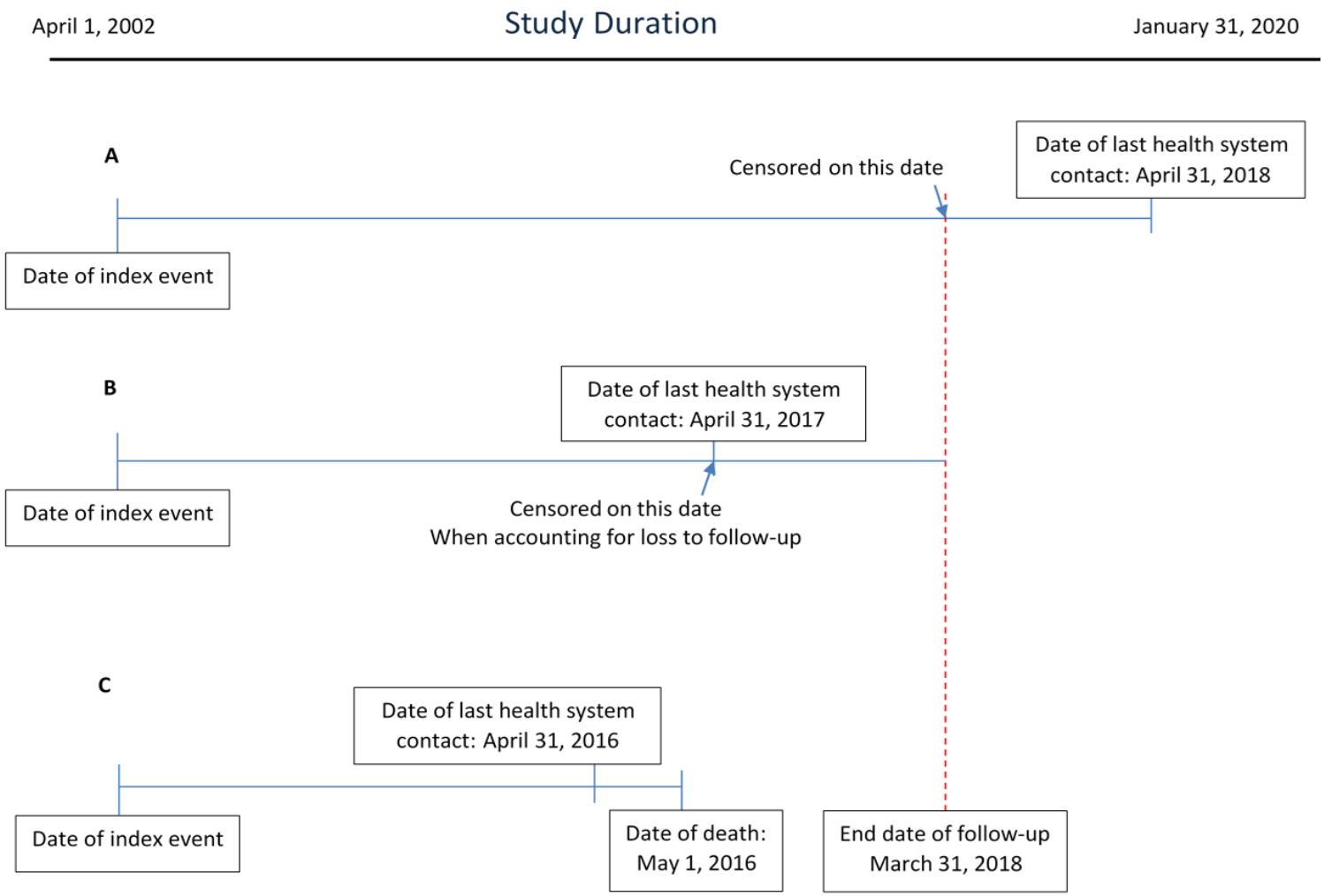
e-table 3. Characteristics of immigrants included in the study.

Immigration-specific characteristics	**Ischemic stroke n = 2078**	**Cancer n = 26,084**	**Schizophrenia n = 9525**
World region of origin			
Africa	81 (3.9)	1128 (4.3)	980 (10.3)
Caribbean	193 (9.3)	1902 (7.3)	967 (10.2)
East Asia	403 (19.4)	5331 (20.4)	1319 (13.8)
Latin America	171 (8.2)	1736 (6.7)	738 (7.7)
Middle East	194 (9.3)	2701 (10.4)	947 (9.9)
South Asia	392 (18.9)	4094 (15.7)	2271 (23.8)
Western	526 (25.3)	7525 (28.8)	1895 (19.9)
Missing	392 (18.9)	1667 (6.4)	408 (4.3)
Time since arrival			
≤ 10 years	677 (32.6)	10360 (39.7)	4763 (50.0)
> 10 years	1401 (67.4)	15724 (60.3)	4762 (50.0)
Immigration class			
Economic	468 (22.5)	9262 (35.5)	3213 (33.7)
Family or other	1273 (61.3)	13233 (50.7)	3891 (37.8)
Refugee	337 (16.2)	3589 (13.8)	2421 (25.4)

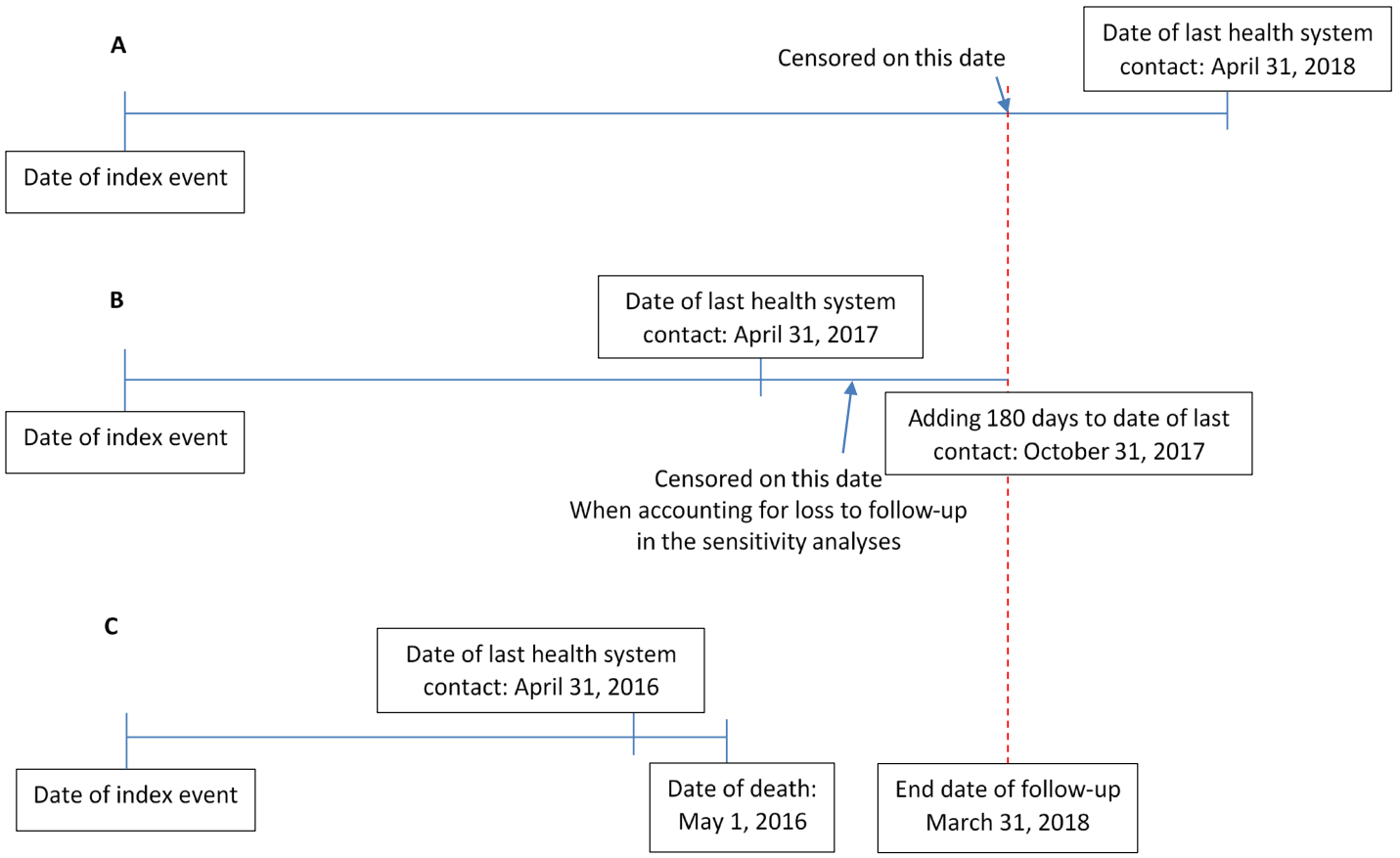
e-table 4. Results of sensitivity analyses using a lag-time of 6 months when determining the date of last health system contact.

		Ischemic stroke	Cancer	Schizophrenia
Immigrants	Lost to follow-up, n (%)	145 (7.0)	1895 (7.3)	1120 (11.8)
Long-term residents		472 (2.1)	5472 (1.9)	3333 (7.4)
Immigrants	Death, n (%)	796 (35.4)	9014 (34.6)	546 (5.7)
Long-term residents		12,575 (55.9)	146,723 (51.6)	6647 (14.7)
Adjusted HR of death (95% CI)^	Immigrants vs. long-term residents	0.82 (0.77-0.89)	0.77 (0.76-0.79)	0.56 (0.51-0.61)

^Accounting for loss of follow-up by censoring those lost to follow-up.



e-figure 1. Hypothetical cases to illustrate loss to follow-up using administrative database. Subject A was not lost to follow-up, Subject B would be considered lost to follow-up, and Subject C had the event of interest (death) and so is not considered lost to follow-up.



e-figure 2. Sensitivity analyses adding 180 days to last date of follow-up. Only for Subject B does addition of 180 days to follow-up change their censoring time; whereas, censoring times remain same for Subject A and C.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Location
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 4
Methods			
Study design	4	Present key elements of study design early in the paper	Page 5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 5 and e-table 2
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	n.a.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 5 and e-table 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Page 5 and e-table 1
Bias	9	Describe any efforts to address potential sources of bias	Page 6
Study size	10	Explain how the study size was arrived at	Figure 1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Table 1
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Page 6-7
		(*b*) Describe any methods used to examine subgroups and interactions	Page 6-7
		(*c*) Explain how missing data were addressed	None present
		(*d*) If applicable, explain how loss to follow-up was addressed	e-table 2 and page 7
		(*e*) Describe any sensitivity analyses	Page 8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Figure 1
		(b) Give reasons for non-participation at each stage	Figure 1

		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	None with missing data
		(c) Summarise follow-up time (eg, average and total amount)	Page 8
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 9 and Table 2
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Page 9, and table 2
		(*b*) Report category boundaries when continuous variables were categorized	Table 2
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n.a.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Figure 3
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Page 12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Page 10
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.