# BMJ Open

# Explaining regional variations in colon cancer survival in Ontario, Canada: a population-based retrospective cohort study

Colleen Webber [1,2] Michael Brundage,[3,4,5] Timothy P Hanna,[3,4,5] Christopher M Booth,[4,5] Erin Kennedy,[6,7] Weidong Kong,[5] Yingwei Peng,[3,5,8] Marlo Whitehead,[9] Patti A Groome[3,5,9]

Check for updates

For numbered affiliations see end of article.

**Correspondence to**
Professor Patti A Groome;
groomep@queensu.ca

## ABSTRACT

**Objectives** Regional variation in cancer survival is an important health system performance measurement. We evaluated if regional variation in colon cancer survival may be driven by differences in the patient population, their health and healthcare utilisation, and/or cancer care delivery.

**Design** Population-based retrospective cohort study using routinely collected linked health administrative data.

**Setting** Ontario, Canada.

**Participants** Patients with colon cancer diagnosed between 1 January 2009 and 31 December 2012.

**Outcome** Cancer-specific survival was compared across the province's 14 health regions. Using accelerated failure time models, we assessed whether regional survival variations were mediated through differences in case mix, including age, sex, comorbidities, stage at diagnosis and colon subsite, potential marginalisation and/or prediagnosis healthcare.

**Results** The study population included 16 895 patients with colon cancer. There was statistically significant regional variation in cancer-specific survival. Three regions had cancer-specific survival that was between 30% (95% CI 1.03 to 1.65) and 39% (95% CI 1.13 to 1.71) longer and one region had cancer-specific survival that was 26% shorter (95% CI 0.58 to 0.93) than the reference region. For three of these regions, case mix explained between 26% and 56% of the survival variation. Further adjustment for rurality explained 22% of the remaining survival variation in one region. Adjustment for continuity of primary care and the diagnostic interval length explained 10% and 11% of the remaining survival variation in two other regions. Socioeconomic marginalisation, recent immigration and colonoscopy history did not explain colon cancer survival variation.

**Conclusions** Case mix accounted for much of the regional variation in colon cancer survival, indicating that efforts to monitor the quality of cancer care through survival metrics should consider case mix when reporting regional survival differences. Future work should repeat this approach in other settings and other cancer sites considering a broad range of potential mediators.

## STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This was a population-based study that included all individuals diagnosed with colon cancer in Ontario, Canada over a 4-year period, thus supporting the generalisability of the findings to other jurisdictions with similar healthcare systems.

⇒ While we examined a range of potential mediators, including measures of patient health and cancer severity, marginalisation and healthcare use, there are likely additional factors that explain the significant regional survival variation that we observed.

⇒ The number of colon cancer cases in some regions was small, which may have limited our ability to identify a statistically significant difference in survival in those regions.

## INTRODUCTION

The study of patient outcomes is critical for assessing healthcare quality, including the effectiveness of cancer-related care.[1–3] Survival is a fundamental cancer outcome, reflecting effective healthcare organisation and high-quality care processes. Variations in survival across regions is therefore of interest to patients, healthcare providers, administrators and policymakers.[4–8] Previous work in Ontario found that in 2012–2016, 5-year relative survival rates for all cancers varied across health regions, with a difference of 12% between regions with the highest and lowest relative survival rates.[9 10] Similarly, we previously documented that Ontario regional cancer-specific survival in patients diagnosed in 2007–2013 ranged from 62% to 72%.[11] The magnitude of these regional differences in cancer survival raises concerns about potential corresponding variation in the quality of cancer-related health services.

Identifying factors responsible for regional survival variations is a first step to mitigating those differences. We postulated that

BMJ

regional variations in cancer survival may be driven by the care quality, and differences in the patient population, their health and healthcare utilisation. The aim of this study was, therefore, to describe regional variations in colon cancer survival in Ontario and explore mediating factors that may explain regional survival variations. We selected colon cancer because it is a relatively common cancer affecting men and women and our preliminary analyses indicated that its age-adjusted and sex-adjusted 5-year cancer-specific survival differed from 59% to 70% across regions.

## MATERIALS AND METHODS

We conducted a population-based retrospective cohort study in Ontario, Canada (population 14.6 million). Ontario's publicly funded, single-payer healthcare system allowed us to conduct this study using population-based, linked health data available through ICES (previously known as the Institute for Clinical Evaluative Sciences), an independent, non-profit research institute. This study is reported in accordance with the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement.[12]

### Data sources

The Ontario Cancer Registry was used to identify the study population and determine the date and stage of cancer at diagnosis. The Office of the Registrar General Deaths Database (ORGD) was used to determine date and cause of death. The Registered Persons Database (RPDB) was used to capture patient sociodemographic information. Healthcare utilisation databases, including the Canadian Institute for Health Information Discharge Abstract Database, Same Day Surgery Database, National Ambulatory Care Reporting System and Ontario Health Insurance Plan Claims Database were used to assign continuity of primary care, comorbidity, previous colonoscopies and the diagnostic interval. These datasets were linked using unique encoded identifiers and analysed at ICES.

### Study population

The study population and its attendant research dataset were originally developed for a previous study that examined relationships between diagnostic resource availability and the colorectal cancer (CRC) diagnostic interval.[13 14] That study included all patients with CRC (International Classification of Diseases, ninth revision (ICD-9) 153.0–153.9, 154.0–154.1, excluding 153.5 (malignant neoplasm of appendix)) diagnosed in Ontario between 1 January 2009 and 31 December 2012. Individuals were excluded from that study if they met any of the following criteria: invalid identifier for data linkage, cancer diagnosed on death certificate only, age >105 or <18 years at diagnosis, non-adenocarcinoma histology, CRC was not their first primary cancer, patient was not covered under the provincial health insurance plan in the 42 months before their cancer diagnosis date, not an Ontario resident in

the 18 months before the CRC diagnosis date, missing diagnostic interval start date or unable to assign to a colonoscopy network (for the original purpose of measuring diagnostic resource availability). For the current study, we report on survival differences for the subset of patients from that original study who were diagnosed with colon cancer (ICD-9 153.x). We thus excluded anyone who met any of the following criteria: rectal cancer subsite, invalid health region or not alive at diagnosis.

### Outcome

The outcome was death due to cancer, captured from the ORGD using ICD-9 disease codes of 140–208 or ICD for Oncology, third revision disease codes of C00–C97. Death dates and underlying cause of death were available for all patients until 31 December 2016.

### Health region

At the time of this study, Ontario healthcare was organised across 14 geographically defined health regions called Local Health Integrated Networks (LHINs). Each patient was linked to their LHIN of residence at the time of cancer diagnosis based on postal code in the RPDB. We have anonymised the LHINs as the purpose of this study was to describe and explain regional survival variation, rather than to identify specific regions with better or worse survival. For clarity of reporting, LHINs will be referred to as health regions.

### Explaining regional survival differences

We first described cancer-specific survival across the health regions then identified which of our potential mediators explained region-level cancer-specific survival variation. In the first instance, regional survival differences might be explained by case mix, that is, differences in patients' baseline demographics, health and colon cancer severity. In this study, we refer to the following variables as characterising 'case mix': age at the diagnostic interval index date (<50, 50–59, 69–69, 70–79 or 80+ years, where the index date is defined as the first cancer-related healthcare encounter prior to diagnosis),[13 15] sex, comorbid disease burden, measured as the number (0, 1, 2+) of major Aggregated Diagnosis Groups (ADGs) in the 2 years prior to the diagnostic interval index date using The Johns Hopkins ACG System V.10,[16] gastrointestinal (GI) comorbidity, based on the GI Major Expanded Diagnostic Cluster from the ACG system, colon cancer subsite, defined as proximal, distal or subsite not otherwise specified and cancer stage at diagnosis measured using the American Joint Committee on Cancer tumour, node, metastases staging classification system, sixth edition. We assessed the extent to which these case mix variables explained the regional survival differences. The cancer-specific survival variation that remained was then further studied to identify targets for system improvement.

After adjusting for case mix, we explored whether membership in a marginalised group might explain the remaining survival differences. Potential marginalisation

was assessed using area-level material deprivation, measured by the Ontario Marginalisation Index,[17] with census dissemination areas categorised into deprivation quintiles 1 (lowest marginalisation) to 5 (highest marginalisation); recent immigrant, defined as those whose Ontario Health Insurance Programme (OHIP) eligibility started within the 10 years before the diagnostic interval index contact; and rurality, defined by the Rurality Index for Ontario, using categories defined by the index creators (0–9 (most urban), 10–30, 31–45, 46–55, 56–75, 75–100 (most rural)).[18]

We also explored whether health system factors might explain the regional survival differences that remained after case mix was considered. System variables included: colonoscopy history in the 5 years before the diagnostic interval index contact (0 or ≥1); continuity of care with the patient's usual provider of primary care (UPC) in the 2 years prior to the diagnostic interval index contact, defined as high (≥75% of primary care visits occurred with UPC), low (<75% of primary care visits occurred with UPC) or undefined (fewer than three primary care visits in the 2-year period), and diagnostic interval length, defined as the number of days from the first cancer-related healthcare encounter to diagnosis. The diagnostic interval was categorised as quartiles except that those with an interval longer than the 90th percentile were placed in an additional category.[13–15]

## Statistical analysis

We used descriptive statistics to characterise the study population and t-test, Kruskal-Wallis one-way analysis of variance and $\chi^2$ test to evaluate whether these characteristics differed across the health regions. A two-sided p<0.05 indicated statistical significance. Overall and cancer-specific survival across the health regions were evaluated using Kaplan-Meier survival curves and log-rank tests.

Our analytic approach treated our case mix, marginalisation and health system variables as potential mediators of the health region cancer-specific survival differences.[19] We calculated unadjusted and adjusted accelerated failure time (AFT) models to evaluate variations in cancer-specific survival across the health regions and assess for mediation. We chose the AFT model rather than the Cox proportional hazards model because AFT does not require a rare event assumption while still allowing for mediation estimation.[20 21] The AFT model considers the exponential of the mean logarithm of survival time, also known as the geometric mean survival time, instead of the regular mean survival time as the function of the case mix, marginalisation and health system variables. The interpretation of the geometric mean is similar to that of the regular mean. The AFT model calculates the ratio of geometric mean survival time (time ratio), where a ratio >1 denotes a longer time to event and a ratio <1 denotes a shorter time to event. We used health region A as the reference region because it contained the most patients and its 5-year colon cancer-specific survival was most likely that of the entire study population. Individuals with missing values for rurality or deprivation quintile were excluded from regression models due to their small number.

Focussing on the health regions with statistically significant unadjusted differences in survival, we calculated the per cent excess risk explained by influential mediators of those differences using the following formula[22]:

$$\% \text{ Excess/Reduced time explained} = \frac{\text{TRu} - \text{TRa}}{\text{TRu} - 1.0} \times 100$$

Where TRu=unadjusted time ratio and TRa=adjusted time ratio.

A priori, we defined a per cent excess or reduced time explained of 10% or more as clinically meaningful. We first adjusted for and assessed the mediating effect of case mix variables on the statistically significant unadjusted health region effects. We then ran six subsequent regression models, with each model controlling for the case mix variables and assessing one of the six potential mediators (material deprivation quintile, immigration status, rurality, colonoscopy history, continuity of care and diagnostic interval length) of the health region-survival associations. This approach to modelling allowed us to evaluate the mediating effect of each of these six variables independently on the case mix-adjusted health region time ratios, again looking for a 10% change to determine if mediation of the region survival time differences was present. We report % excess/reduced time explained when it reached our 10% threshold for health regions whose unadjusted survival was statistically significantly different from region A.

## Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

## RESULTS

The initial data capture included 30 061 patients with CRC (figure 1). Of those, 6100 were excluded to meet our original study's target population definition or because of missing information (n=23 961). The original target population exclusions were: non-adenocarcinoma histology (n=1220, 20% of excluded), a prior cancer (n=2845, 47% of excluded) or lack of continuous OHIP eligibility prior to diagnosis (n=634, 10% of excluded). Missing information excluded 746 patients (12% of excluded). For the current study, we excluded 7050 patients with rectal cancer and 16 patients who were not alive at diagnosis or who had unknown health region. Our final study population included 16 895 patients with colon cancer (table 1). The mean age was 69.7 years, 50.9% were male, 63.9% lived in the most urban areas and 3.1% were recent immigrants. There was an even distribution of patients across the material deprivation quintiles. Fifteen per cent of patients had at least one prior colonoscopy, 60.8% had a comorbid GI condition and 26.9% had two or more major comorbid conditions. Approximately half (48.8%)
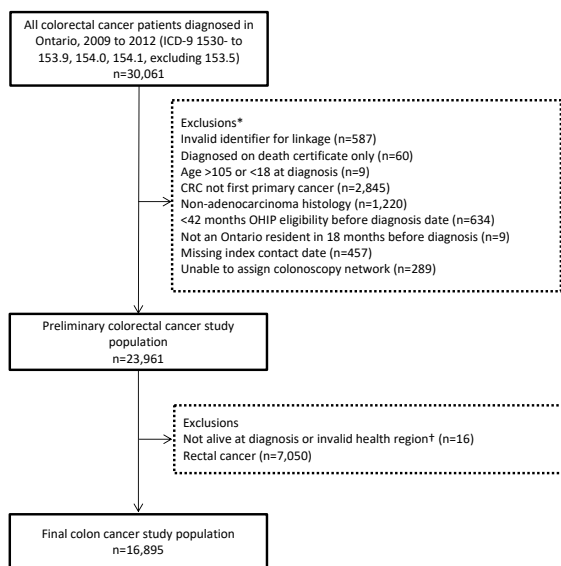
**Figure 1** Study population flow diagram. *Numbers do not add to total number of individuals excluded due to patients who met more than one exclusion criteria. †Combined for reporting due to small number (n≤5) with invalid health region. CRC, colorectal cancer; ICD, International Classification of Disease; OHIP, Ontario Health Insurance Programme.

of patients had high continuity of care with their UPC prior to the cancer diagnostic interval start date. Almost half of patients were diagnosed with early stage (I or II) colon cancer, while 16.9% were diagnosed with stage IV cancer. Proximal colon cancer was the most common subsite (51.3%). The median diagnostic interval was 87 days. Aside from patient sex, all these characteristics varied significantly across the health regions.

The median follow-up time for survival after diagnosis was 5 years (IQR 2–6). Of the 16 895 patients, 5612 (33.2%) died of cancer-related causes, 1934 (11.4%) died of non-cancer causes and 9349 (55.3%) were alive as of 31 December 2016. The unadjusted overall and cancer-specific survival curves by health region are displayed in figures 2 and 3, respectively. In the unadjusted model (table 2), three health regions had a mean cancer-specific survival that was significantly longer than that of the reference region A, ranging from 30% longer (95% CI 1.03 to 1.65) in region M, to 34% (95% CI 1.01 to 1.78) in region D, and 39% longer (95% CI 1.13 to 1.71) in region J. Mean cancer-specific survival was 26% shorter in region C compared with A (95% CI 0.58 to 0.93).

After adjusting for case mix, the mean survival time ratio for health region C was closer to 1.0 (0.82, 95% CI 0.68 to 0.99), indicating that 46% of the unadjusted differences in survival in region C versus A were explained by case mix (table 2). The mean survival time ratio for health region D was reduced to 1.15 and no longer statistically significant (95% CI 0.92 to 1.44), indicating that 56% of the unadjusted differences in survival in region D versus A were explained by case mix. The mean survival time ratio for health region J was reduced to 1.29 (95% CI 1.09 to 1.51), indicating that 26% of the unadjusted differences in survival in region J versus A were explained by case mix. While health region D still had a significantly higher survival time ratio than health region A after case mix adjustment (1.35, 95% CI 1.12 to 1.63) and health region B had a significantly lower survival time ratio than health region A after case mix adjustment (0.82, 95% CI 0.68 to 0.99), these reflected a <10% change from the unadjusted model. For all other health regions, the case mix adjusted survival time ratios were not statistically significantly different from region A and did not change ≥10% with case mix adjustment.

Adjustment for deprivation quintile or immigration status did not account for any of the statistically significant health region survival differences that remained after adjusting for case mix (table 2). In both of those adjusted models, health regions M and J continued to have significantly longer mean cancer-specific survival than region A, while health regions C and B continued to have significantly shorter mean cancer-specific survival than health region A.

Adjustment for rurality accounted for 22% of the shorter mean survival in health region C in contrast to health region A. The region C mean survival was 14% shorter than region A (0.86, 95% CI 0.71 to 1.04) after rurality and case mix adjustment, in contrast to 18% shorter than region A (0.82, 95% CI 0.68 to 0.99) after only case mix adjustment. In all other cases, rurality accounted for <10% of the difference in survival.

Healthcare characteristics explained a small amount of the mean survival differences in health regions J and B that remained after case mix adjustment (table 3). Adjustment for continuity of care accounted for 10% of the longer survival in health region J and adjustment for the diagnostic interval length accounted for 11% of the shorter survival in health region B. Adjustment for colonoscopy history did not result in substantial changes to the model estimates. In all other cases, healthcare characteristics explained <10% of the difference in survival.

## DISCUSSION

This population-based study of colon cancer survival revealed significant variation in cancer-specific survival across the 14 health regions in Ontario. Three health regions had significantly higher mean cancer-specific survival than region A, which most closely reflected the survival experience of Ontario as a whole. The mean

**Table 1** Patient characteristics overall and by health region

| | Overall N=16895 | Health region | | | | | | | | | | | | | | | P value |
| | | A N=2142 | B N=992 | C N=927 | D N=714 | E N=366 | F N=899 | G N=1922 | H N=971 | I N=1416 | J N=1868 | K N=728 | L N=1148 | M N=1146 | N N=1656 | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Age (years)** | | | | | | | | | | | | | | | | <0.001 |
| <50 | 6.8 | 5.7 | 4.3 | 5.5 | 10.5 | 6.3 | 5.6 | 7.2 | 7.5 | 5.3 | 8.7 | 5.9 | 8.4 | 9.9 | 5.5 | |
| 50–59 | 14.5 | 13.7 | 15.4 | 12.6 | 17.6 | 15.6 | 12.6 | 14.3 | 14.4 | 13.4 | 14.4 | 13.5 | 15.3 | 17.7 | 14.7 | |
| 60–69 | 24.8 | 24.7 | 26.4 | 25.4 | 27.6 | 21.9 | 25.8 | 23.5 | 23.5 | 26.6 | 22.5 | 25.1 | 23.2 | 24.2 | 27.5 | |
| 70–79 | 28.4 | 30.1 | 31.8 | 28.5 | 26.8 | 27.9 | 30.9 | 27.5 | 30 | 28.8 | 27.2 | 30.2 | 26.6 | 24.7 | 28.2 | |
| 80+ | 25.4 | 25.8 | 22.1 | 28 | 17.5 | 28.4 | 25.1 | 27.5 | 24.6 | 25.9 | 27.2 | 25.3 | 26.6 | 23.6 | 24.1 | |
| Male | 59.9 | 50.9 | 53.1 | 52.1 | 50 | 49.5 | 51.8 | 49.4 | 47.1 | 51.3 | 51 | 48.9 | 51.3 | 54 | 51 | 0.21 |
| **Number of major ADGs** | | | | | | | | | | | | | | | | <0.001 |
| 0 | 39.7 | 37.5 | 38.7 | 36.2 | 45.4 | 37.7 | 35.5 | 37.6 | 45 | 38.9 | 40.8 | 40.1 | 40.4 | 42.7 | 41 | |
| 1 | 33.4 | 33.6 | 32.7 | 33.5 | 31.8 | 32.5 | 35.2 | 34.7 | 32.6 | 36.4 | 32 | 32.7 | 32.1 | 33.1 | 31.7 | |
| ≥2 | 26.9 | 28.9 | 27.2 | 30.2 | 22.8 | 29.8 | 29.4 | 27.8 | 22.3 | 24.7 | 27.2 | 27.2 | 27.4 | 24.3 | 27.3 | |
| Has GI comorbidity | 60.8 | 61.6 | 56.9 | 61.8 | 59.8 | 65.6 | 55.1 | 63.5 | 57.7 | 57 | 63 | 57.7 | 66.4 | 61.1 | 60.3 | <0.001 |
| **Subsite** | | | | | | | | | | | | | | | | <0.001 |
| Distal | 33.4 | 33.4 | 30.6 | 33 | 28.7 | 35 | 33.4 | 33.7 | 36.3 | 32.6 | 33.7 | 28.4 | 34.7 | 32.4 | 36.8 | |
| Proximal | 51.3 | 53 | 54.3 | 54.5 | 47.9 | 58.5 | 52.1 | 50.6 | 50.9 | 54.1 | 47.8 | 56.5 | 45.1 | 46.8 | 52.9 | |
| NOS | 15.3 | 13.6 | 15 | 12.5 | 23.4 | 6.6 | 14.6 | 15.7 | 12.9 | 13.3 | 18.5 | 15.1 | 20.2 | 20.9 | 10.2 | |
| **Stage** | | | | | | | | | | | | | | | | <0.001 |
| I | 20.6 | 20.5 | 18.2 | 21.9 | 24.2 | 22.4 | 17.4 | 22.2 | 19.8 | 21.4 | 21.9 | 21.3 | 19.7 | 19.5 | 19 | |
| II | 25.7 | 25.5 | 26.5 | 24.6 | 20.6 | 24 | 27 | 24.5 | 25.8 | 28.2 | 25.6 | 24.2 | 26.5 | 24.2 | 28.4 | |
| III | 24.7 | 23.9 | 22.1 | 23.7 | 25.1 | 19.9 | 27.1 | 25.5 | 26.1 | 25.6 | 25.5 | 23.6 | 23.3 | 22 | 27.2 | |
| IV | 16.9 | 18.1 | 14.1 | 19.8 | 15.4 | 15.3 | 18.9 | 17.1 | 17.1 | 15.3 | 15.8 | 17.3 | 18.8 | 16.8 | 16.3 | |
| Unknown | 12 | 12 | 19.1 | 9.9 | 14.7 | 18.3 | 9.6 | 10.6 | 11.2 | 9.5 | 11.1 | 13.6 | 11.7 | 17.5 | 9 | |
| **Deprivation quintile** | | | | | | | | | | | | | | | | <0.001 |
| 1 (least deprived) | 18 | 19.4 | 12.6 | 18.2 | 10.2 | 12.8 | 14 | 12.1 | 30.9 | 23.2 | 12.7 | 9.2 | 15.5 | 22.5 | 28.9 | |
| 2 | 19 | 20.3 | 17.8 | 18.9 | 13.7 | 15.6 | 21.1 | 15.9 | 19.9 | 21.5 | 16.6 | 21.2 | 17.4 | 21.6 | 21.7 | |
| 3 | 20.5 | 17.4 | 19.4 | 20.2 | 19.9 | 23.5 | 21.2 | 18.7 | 15 | 22.5 | 26.3 | 29 | 16.6 | 22.7 | 19 | |
| 4 | 20.9 | 17.6 | 24.9 | 15.7 | 31.7 | 24.6 | 18.9 | 27.8 | 19.1 | 17.7 | 23.7 | 22.4 | 21.4 | 19.1 | 14.6 | |
| 5 (most deprived) | 20.9 | 24.8 | 22.6 | 26.8 | 24.5 | 15.6 | 24.2 | 24.9 | 15 | 14.6 | 20.4 | 17.6 | 28 | 14.1 | 15.3 | |
| Unknown | 0.7 | 0.5 | 2.7 | 0.2 | 0 | 7.9 | 0.4 | 0.6 | 0.1 | 0.5 | 0.2 | 0.7 | 1 | 0 | 0.4 | |
| Recent immigrant | 3.1 | 2 | 1.7 | 1.9 | 6.3 | 2.2 | 2.3 | 2.4 | 1.3 | 1.4 | 5.4 | 2.2 | 3.7 | 4.5 | 5.1 | <0.001 |
| **RIO score** | | | | | | | | | | | | | | | | <0.001 |
| 0–9 (most urban) | 63.9 | 73.9 | 24.3 | 29.1 | 93 | 43.4 | 34.5 | 60.7 | 75 | 37.9 | 87.3 | 22.4 | >99.0* | >99.0* | 63.9 | |

Continued

**Table 1** Continued

| | Overall N=16895 | Health region | | | | | | | | | | | | | | P value |
| | | A N=2142 | B N=992 | C N=927 | D N=714 | E N=366 | F N=899 | G N=1922 | H N=971 | I N=1416 | J N=1868 | K N=728 | L N=1148 | M N=1146 | N N=1656 | |
| 10–30 | 18.1 | 23.9 | 30.8 | 61.4 | 2.9 | 0 | 27.4 | 24.4 | 15.1 | 22.3 | 11.8 | 20.6 | <1.0* | 0 | 6 | |
| 31–45 | 10 | 1.6 | 3.3 | 9.1 | <4.0* | 4.4 | 31.8 | 10.5 | 6.6 | 22.9 | 0.7 | 36.8 | 0 | <1.0* | 20 | |
| 46–55 | 2.8 | 0 | 5.6 | 0 | <4.0* | 0 | 2.2 | 1.8 | 3.3 | 14 | 0.1 | 6.6 | 0 | 0 | 5.2 | |
| 56–75 | 2.9 | 0 | 20.7 | 0 | 0 | 1.9 | 3.8 | 2.1 | 0 | 2.7 | 0.1 | 12.6 | <1.0* | 0 | 4.7 | |
| 76+ (most rural) | 1.1 | 0 | 6.5 | 0 | 0 | 31.7 | 0 | 0.3 | 0 | 0 | 0 | <1.0* | 0 | 0 | 0 | |
| Unknown | 1.1 | 0.5 | 8.8 | 0.4 | 0 | 18.6 | 0.3 | 0.3 | 0 | 0.3 | 0.1 | <1.0* | 0 | 0 | 0.2 | |
| Number of colonoscopies in previous 5 years | | | | | | | | | | | | | | | | 0.006 |
| 0 | 85.2 | 86.6 | 84.7 | 82.4 | 87.7 | 85.2 | 87.8 | 85.3 | 87.6 | 84.5 | 84.3 | 81.5 | 85.2 | 86.6 | 86 | |
| 1 | 11.3 | 10.2 | 11.6 | 13.9 | 9.5 | 11.2 | 9.1 | 13 | 8.9 | 12 | 12.2 | 14.4 | 10.4 | 10.6 | 10.9 | |
| >1 | 3.4 | 3.3 | 3.7 | 3.7 | 2.8 | 3.6 | 3.1 | 3.5 | 3.5 | 3.5 | 3.5 | 4.1 | 4.4 | 2.7 | 3.1 | |
| Continuity of Care Index | | | | | | | | | | | | | | | | <0.001 |
| Low | 10 | 8.8 | 8 | 9.7 | 12.5 | 11.2 | 9.6 | 9.8 | 7.1 | 7.8 | 12.5 | 7.7 | 9.6 | 14.7 | 10.7 | |
| High | 48.8 | 49.3 | 43.2 | 53.4 | 52.4 | 37.2 | 43.7 | 52.5 | 40 | 47 | 54.3 | 48.4 | 52.3 | 51.5 | 44.5 | |
| 0 visits | 14.8 | 14.9 | 20.8 | 14.9 | 12.6 | 20.8 | 16 | 12.8 | 17.6 | 16.9 | 11.3 | 14.4 | 15.2 | 10.1 | 15.5 | |
| 1–2 visits | 26.5 | 27 | 28 | 22 | 22.5 | 30.9 | 30.7 | 24.8 | 35.3 | 28.3 | 21.8 | 29.5 | 23 | 23.6 | 29.2 | |
| Diagnostic interval length percentile | | | | | | | | | | | | | | | | <0.001 |
| ≤25th | 25.2 | 23.9 | 27.9 | 24.2 | 24.4 | 27.6 | 25.3 | 25.9 | 23.4 | 24.6 | 26.2 | 28.8 | 27.5 | 21 | 25.3 | |
| >25th to ≤50th | 25 | 26 | 22.8 | 24.4 | 23.8 | 23.2 | 26.1 | 26.8 | 31.3 | 24.9 | 24.5 | 25.3 | 20.7 | 26.6 | 22.1 | |
| >50th to ≤75th | 24.9 | 25.2 | 25.2 | 24.1 | 26.8 | 25.7 | 26.9 | 22.4 | 24.5 | 25.9 | 24.1 | 24.2 | 24.6 | 25.9 | 26.3 | |
| >75th to ≤90th | 14.9 | 15.1 | 13.9 | 15.2 | 16.1 | 14.5 | 13 | 16.1 | 11.9 | 14.1 | 15.8 | 12.1 | 16.3 | 15.6 | 15.3 | |
| >90th | 9.9 | 9.9 | 10.2 | 12.2 | 9 | 9 | 8.7 | 8.9 | 8.9 | 10.5 | 9.3 | 9.6 | 10.9 | 10.8 | 10.9 | |

*Exact proportions not reported due to small cell counts.
ADGs, Aggregated Diagnostic Groups; NOS, not otherwise specified; RIO, Rurality Index for Ontario.
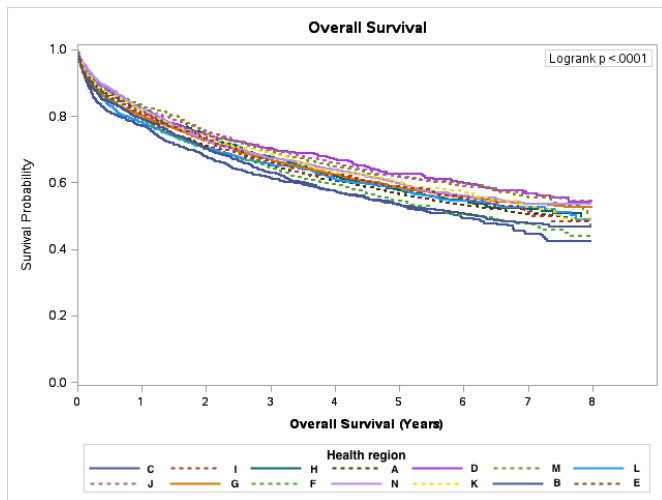
**Figure 2** Overall survival curves by health region with log-rank test p value.

survival times for these three health regions were between 30% and 39% higher than region A. One health region had mean cancer-specific survival that was 26% lower than that of region A. Case mix explained the greatest proportion of variation in survival across the health regions, while system-level factors explained the least. However, even after accounting for these mediating factors, significant survival differences across health regions remained. This study builds on our understanding of factors underpinning regional differences in cancer survival. A recent international comparison of cancer survival across seven high-income countries identified stage of disease at diagnosis, timely access to effective treatment and comorbidity as potential determinants of international survival disparities.[4] Our findings suggest that case mix, which includes stage at diagnosis as well as other indicators of patient health and disease severity, are important determinants of colon cancer survival variation. Furthermore, in some regions, rurality and continuity of primary care may also be important.
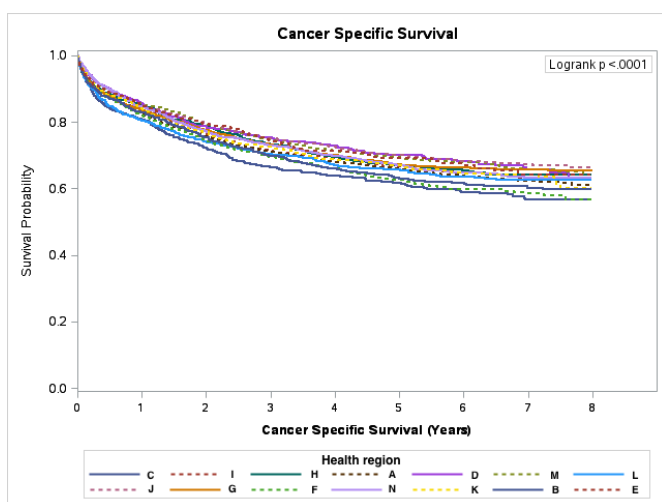


**Figure 3** Cancer-specific survival curves by health region with log-rank test p value.

Our case mix variables reflected patients' underlying health status and cancer severity. All of these variables are known to be associated with colon cancer survival. Except for sex, the distribution of all the case mix variables varied significantly across the health regions. For instance, there was a 5.0% difference in the distribution of stage I (17.4% in region F vs 22.4% in region E) and a 5.7% difference in the distribution of stage IV (14.1% of patients in region B vs 19.8% in region I). These variations in stage may reflect differences in patient characteristics and health behaviours across the health regions; for instance, uptake of colon cancer screening through the provincial, publicly funded screening programme varies across LHINs.[23] Stage variations may also be driven in part by differences in access to care and diagnostic processes that contribute to a later stage at diagnosis. The proportion of patients with unknown stage also differed across the health regions, from a low of 9.0% in region N to a high of 19.1% in region B. Previous research using Ontario cancer stage data has found that unknown cancer stage is an indicator of poor quality care, which supports its inclusion as a potential explanation of health region-level survival differences.[24]

Indicators of marginalisation, including deprivation quintile, immigration status and rurality, did not explain the significant health region differences in colon cancer survival that remained after case mix adjustment. The exception was for health region C, where the magnitude of difference in mean survival was reduced and no longer statistically significantly different after adjustment for rurality. Rurality has generally been associated with worse survival in colon and other cancers.[25] Our finding indicates that patients living in rural areas in health region C had worse survival so this region could be a target for quality improvement initiatives. Previous literature has documented differences in cancer survival in Ontario by area-level income, with patients from lower-income areas generally having poorer cancer survival.[26] While we observed variation in area-level deprivation across the health regions, this variable did not play a mediating role in the region-level differences in colon cancer survival. Similarly, while immigration status has been associated with worse survival in colorectal cancer, this characteristic did not explain the regional survival differences that were our focus.[27]

We studied three healthcare-related variables which may affect colon cancer care and thus survival: colonoscopy history, continuity of primary care prior to diagnosis and the diagnostic interval length. We included the first two variables to assess whether relevant aspects of prediagnosis healthcare use might play a role in cancer survival while the diagnostic interval has previously been associated with colorectal cancer survival.[28–30] The rates of previous colonoscopy varied across health regions but those variations did not explain any region-level survival differences. A potential explanation is that survival benefits of previous colonoscopy may occur through earlier stage at diagnosis, which was already accounted for in the

**Table 2** Unadjusted and adjusted geometric mean cause-specific survival time ratios and their 95% CI for each health region versus region A

| Health region | Unadjusted | Case mix* adjusted | % excess/reduced time explained by case mix† | Deprivation quintile and case mix adjusted | Immigration and case mix adjusted | Rurality and case mix adjusted | % excess/reduced time explained by rurality† |
|---|---|---|---|---|---|---|---|
| A | Reference | Reference | | Reference | Reference | Reference | |
| B | 0.88 (0.69 to 1.11) | **0.82 (0.68 to 0.99)‡** | | **0.82 (0.68 to 0.99)** | **0.82 (0.68 to 0.99)** | 0.83 (0.68 to 1.02) | |
| C | **0.74 (0.58 to 0.93)** | **0.82 (0.68 to 0.99)** | 46% explained | **0.82 (0.68 to 0.99)** | **0.82 (0.68 to 0.99)** | 0.86 (0.71 to 1.04) | 22% explained |
| D | **1.34 (1.01 to 1.78)** | 1.15 (0.92 to 1.44) | 56% explained | 1.15 (0.92 to 1.45) | 1.15 (0.92 to 1.44) | 1.13 (0.90 to 1.41) | |
| E | 1.25 (0.86 to 1.81) | 1.21 (0.90 to 1.62) | | 1.19 (0.88 to 1.61) | 1.21 (0.90 to 1.62) | 1.19 (0.87 to 1.62) | |
| F | 0.81 (0.64 to 1.04) | 0.96 (0.79 to 1.16) | | 0.96 (0.79 to 1.16) | 0.95 (0.79 to 1.16) | 0.97 (0.79 to 1.18) | |
| G | 1.18 (0.96 to 1.44) | 1.17 (1.00 to 1.37) | | 1.18 (1.00 to 1.38) | 1.17 (1.00 to 1.37) | 1.18 (1.00 to 1.38) | |
| H | 1.12 (0.88 to 1.44) | 1.06 (0.88 to 1.29) | | 1.03 (0.84 to 1.25) | 1.07 (0.88 to 1.30) | 1.06 (0.87 to 1.29) | |
| I | 1.07 (0.86 to 1.33) | 0.94 (0.79 to 1.12) | | 0.92 (0.78 to 1.09) | 0.94 (0.79 to 1.12) | 0.95 (0.79 to 1.13) | |
| J | **1.39 (1.13 to 1.71)** | **1.29 (1.09 to 1.51)** | 26% explained | **1.29 (1.09 to 1.52)** | **1.28 (1.09 to 1.51)** | **1.27 (1.08 to 1.50)** | |
| K | 1.07 (0.82 to 1.41) | 1.18 (0.95 to 1.46) | | 1.18 (0.95 to 1.46) | 1.18 (0.95 to 1.46) | 1.19 (0.95 to 1.49) | |
| L | 0.96 (0.76 to 1.21) | 1.00 (0.83 to 1.20) | | 1.01 (0.84 to 1.21) | 1.00 (0.84 to 1.20) | 0.98 (0.82 to 1.18) | |
| M | **1.30 (1.03 to 1.65)** | **1.35 (1.12 to 1.63)** | | **1.32 (1.09 to 1.60)** | **1.35 (1.12 to 1.63)** | **1.32 (1.09 to 1.59)** | |
| N | 1.12 (0.91 to 1.37) | 1.04 (0.88 to 1.22) | | 1.01 (0.86 to 1.20) | 1.04 (0.88 to 1.22) | 1.03 (0.87 to 1.22) | |

*Case mix variables included colon cancer subsite, age, sex, stage at diagnosis, presence of GI comorbidity and number of major ADGs.
†Computed for health regions with statistically significant different survival than region A in unadjusted model and reported when the % change from unadjusted model (for mediation by case mix) or case mix adjusted model (for mediation by deprivation, immigration and rurality) was at least 10%. The % change did not reach the 10% threshold in the models analysing mediation by deprivation index and immigration.
‡p<0.05 for bolded survival time ratios.
ADGs, Aggregated Diagnostic Groups; GI, gastrointestinal.

**Table 3** Case mix adjusted and healthcare characteristic adjusted geometric mean cause-specific survival time ratios and their 95% CIs for each health region versus region A

| | Case mix adjusted | Colonoscopy history and case mix adjusted | Continuity of care and case mix adjusted | % excess/ reduced time explained by continuity of care* | Diagnostic interval and case mix adjusted | % excess/ reduced time explained by diagnostic interval* |
|---|---|---|---|---|---|---|
| Health region | | | | | | |
| A | Reference | Reference | Reference | | Reference | |
| B | **0.82 (0.68 to 0.99)‡** | **0.81 (0.67 to 0.98)** | **0.82 (0.68 to 0.99)** | | 0.84 (0.69 to 1.01) | 11% explained |
| C | **0.82 (0.68 to 0.99)** | **0.81 (0.67 to 0.97)** | **0.82 (0.68 to 0.98)** | | **0.82 (0.68 to 0.99)** | |
| D | 1.15 (0.92 to 1.44) | 1.16 (0.93 to 1.46) | 1.13 (0.90 to 1.41) | | 1.14 (0.91 to 1.42) | |
| E | 1.21 (0.90 to 1.62) | 1.22 (0.91 to 1.64) | 1.22 (0.91 to 1.64) | | 1.21 (0.90 to 1.62) | |
| F | 0.96 (0.79 to 1.16) | 0.96 (0.79 to 1.16) | 0.96 (0.79 to 1.16) | | 0.96 (0.79 to 1.16) | |
| G | 1.17 (1.00 to 1.37) | 1.17 (1.00 to 1.37) | 1.16 (0.99 to 1.36) | | 1.17 (1.00 to 1.37) | |
| H | 1.06 (0.88 to 1.29) | 1.06 (0.87 to 1.29) | 1.07 (0.88 to 1.30) | | 1.06 (0.87 to 1.28) | |
| I | 0.94 (0.79 to 1.12) | 0.94 (0.79 to 1.11) | 0.94 (0.79 to 1.12) | | 0.94 (0.79 to 1.12) | |
| J | **1.29 (1.09 to 1.51)** | **1.27 (1.08 to 1.49)** | **1.26 (1.07 to 1.49)** | 10% explained | **1.30 (1.11 to 1.53)** | |
| K | 1.18 (0.95 to 1.46) | 1.16 (0.94 to 1.44) | 1.18 (0.95 to 1.46) | | 1.21 (0.97 to 1.49) | |
| L | 1.00 (0.83 to 1.20) | 1.00 (0.83 to 1.20) | 1.00 (0.83 to 1.20) | | 1.03 (0.86 to 1.23) | |
| M | **1.35 (1.12 to 1.63)** | **1.35 (1.12 to 1.63)** | **1.32 (1.09 to 1.60)** | | **1.35 (1.12 to 1.63)** | |
| N | 1.04 (0.88 to 1.22) | 1.04 (0.88 to 1.22) | 1.04 (0.88 to 1.22) | | 1.04 (0.88 to 1.23) | |

*Computed for health regions with statistically significantly different survival than region A in unadjusted model and reported when the % change from the case mix adjusted model was at least 10%. The % change did not reach the 10% threshold in the model analysing mediation by colonoscopy history.

†Case mix variables included colon cancer subsite, age, sex, stage, presence of GI comorbidity and number of major ADGs.

‡p<0.05 for bolded survival time ratios.

ADGs, Aggregated Diagnostic Groups; GI, gastrointestinal.

case mix adjusted model. High continuity of primary care varied across health regions, which may reflect differences in primary care access and/or structures across the regions. Adjustment for continuity of primary care explained 10% of the longer mean survival difference in health region J, implying the quality of primary care in that region contributed to better survival of its patients with colon cancer. The median diagnostic interval ranged from 77 to 96 days across health regions. These differences may be partly a result of unadjusted variations in case mix as patients with more advanced disease often receive an expedited diagnosis due to symptom severity. However, variations in the diagnostic interval length may also reflect differences in the efficiency and/or quality of healthcare delivery across the health regions. Adjustment for diagnostic interval length explained 11% of the shorter mean survival in health region B, implying that the differences in the diagnostic interval in that region were one explanation for its shorter survival. Further study is needed to understand the factors driving these regional differences in continuity of primary care and the colon cancer diagnostic interval, and the process through which they may contribute to differences in survival across regions.

Cancer survival is an important outcome to patients and providers. Evidence of substantial regional variations in cancer survival may point to issues with cancer care quality. However, as shown in this study, reporting regional cancer survival differences for the purposes of assessing healthcare quality without controlling for case mix can be misleading. Unadjusted survival differences reflect, at least in part, differences in patient's underlying health and disease severity and should thus not be interpreted as being completely due to differences in healthcare quality. The results in health region D are a case in point. Better survival in that region was partly explained by the fact that the patients in that region were younger, had less comorbid disease and more stage I cancers. This demonstrates the importance of adjusting for case mix differences when reporting cancer survival as a healthcare quality measure.

This study had several strengths and limitations. This population-level study allowed us to identify region-specific factors that may explain colon cancer survival differences in Ontario. We expect that jurisdictions with similar health systems would see similar effects of explanatory variables and we encourage ongoing surveillance using the approach we presented. We chose to anonymise the health regions as that information was not necessary to convey the main findings of our study. Furthermore, identifying the health regions may have been unnecessarily stigmatising, particularly since the survival

variations and mediation observed in this study may not persist today. Our choice of potential mediators that may explain regional survival differences was guided by our prior knowledge about factors that can affect cancer survival along with our clinical judgement. These mediators were evaluated using a mediation framework to identify those that were influential at the health region level. We evaluated each of the non-case mix variables in separate models to assess for mediation and to identify at-risk groups of patients. Some of the non-case mix variables may have been associated so that their independent mediating effect may have been smaller than we observed. However, the separate models achieve our purpose of identifying groups who may be at a higher risk of cancer death. Furthermore, we did not include a fully adjusted model as our purpose was not to assess the causal impact of these mediators. We expect that correlations between mediators would have resulted in a fully adjusted model with independent effects that would be difficult to interpret.

Mediation analysis requires control for all important confounders across all contrasts. For this study, this would include confounders of the health region-survival, health region-mediator and mediator-survival associations. We think our control for key colon cancer prognostic factors largely encompasses confounding under these three conditions since prognostic factors play such a strong role in medical decision making. That said, unmeasured confounding may be present in our results. Furthermore, this study was limited in the potential mediators studied based on their availability in the administrative data holdings. For instance, the data sources for this study do not include measures of race or ethnicity, and Indigenous status is incomplete. We were therefore not able to examine mediation by these characteristics. Further investigation of potential mediators beyond the data we had available is warranted. Finally, while the study had a large overall study population, the number of patients with colon cancer in some health regions was small which may be contributing to wider CIs for model estimates for those health regions. For instance, region E had an unadjusted survival time that was 25% longer than that of region A, but with wide CIs that overlapped 1.00.

This study offers a deeper understanding of regional variations in colon cancer survival in Ontario than was previously possible. Much of the regional variations in colon cancer survival were explained by differences in case mix between the health regions, although even after accounting for case mix, regional survival differences remained. In some health regions, rurality, continuity of primary care and the diagnostic interval length explained colon cancer survival variation. Our approach and findings may be used to support quality improvement initiatives that may identify and address the contributing factors driving cancer survival differences. Future work should apply this approach in other settings and cancer sites considering a broad range of potential mediators. Cancer quality surveillance efforts should consider case mix when reporting regional survival differences.

**Author affiliations**
[1]Bruyere Research Institute, Ottawa, Ontario, Canada
[2]Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
[3]Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada
[4]Department of Oncology, Queen's University, Kingston, Ontario, Canada
[5]Division of Cancer Care and Epidemiology, Cancer Research Institute, Queen's University, Kingston, Ontario, Canada
[6]University of Toronto, Toronto, Ontario, Canada
[7]Ontario Health (Cancer Care Ontario), Toronto, Ontario, Canada
[8]Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada
[9]ICES, Kingston, Ontario, Canada

**ORCID iD**
Colleen Webber http://orcid.org/0000-0001-9193-5386

## REFERENCES

1. Avedis D. *An introduction to quality assurance in health care*. New York: Oxford University Press, 2003.
2. Donabedian A. Evaluating the quality of medical care. 1966. *Milbank Q* 2005;83:691–729.
3. Porter ME, EOlmsted T. *Redefining health care: creating value-based competition on results*. Boston, Mass: Harvard Business School Press, 2006.
4. Arnold M, Rutherford MJ, Bardot A, *et al*. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995-2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol* 2019;20:1493–505.
5. Maringe C, Walters S, Rachet B, *et al*. Stage at diagnosis and colorectal cancer survival in six high-income countries: a population-based study of patients diagnosed during 2000-2007. *Acta Oncol* 2013;52:919–32.
6. Lipscomb J, Yabroff KR, Hornbrook MC, *et al*. Comparing cancer care, outcomes, and costs across health systems: Charting the course. *J Natl Cancer Inst Monogr* 2013;2013:124–30.
7. Porter ME. What is value in health care? *N Engl J Med* 2010;363:2477–81.
8. Coleman MP. Cancer survival: global surveillance will stimulate health policy and improve equity. *Lancet* 2014;383:564–73.
9. Ontario Health: Cancer Care Ontario. Ontario Cancer Profiles: Cancer Survival – All cancer >>Both sexes all ages (15 to 99) (2012-16) [Internet]. Available: https://profiles.cancercare.on.ca/Survival/atlas.html [Accessed 01 Oct 2021].
10. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *Natl Cancer Inst Monogr* 1961;6:101–21.
11. Brundage MD, Groome P, Hanna T, *et al*. Cancer survival outcomes in Ontario, Canada: significant unexplained variation. *Journal of Clinical Oncology* 2018;36:36.
12. Benchimol EI, Smeeth L, Guttmann A, *et al*. The reporting of studies conducted using observational Routinely-collected health data (record) statement. *PLoS Med* 2015;12:e1001885.
13. Webber C. *Availability and quality of colonoscopy resources and the colorectal cancer diagnostic interval [PhD*. Kingston: Queen's University, 2017.
14. Webber C, Flemming JA, Birtwhistle R, *et al*. Colonoscopy resource availability and its association with the colorectal cancer diagnostic interval: a population-based cross-sectional study. *Eur J Cancer Care* 2020;29:e13187.
15. Groome PA, Webber C, Whitehead M, *et al*. Determining the cancer diagnostic interval using administrative health care data in a breast cancer cohort. *JCO Clin Cancer Inform* 2019;3:1–10.
16. Johns Hopkins University School of Public Heatlh. *The Johns Hopkins ACG system. version 10.0 ED*. Baltimore, MD, 2011.
17. Matheson FI, Dunn JR, Smith KLW, *et al*. Development of the Canadian marginalization index: a new tool for the study of inequality. *Can J Public Health* 2012;103:S12–16.
18. Kralj B. Measuring "rurality" for purposes of health-care planning: An empirical measure for Ontario. *Ont Med Rev* 2000;67:33–52.
19. Hafeman DM, Schwartz S. Opening the black box: a motivation for the assessment of mediation. *Int J Epidemiol* 2009;38:838–45.
20. VanderWeele TJ. Mediation analysis: a practitioner's guide. *Annu Rev Public Health* 2016;37:17–32.
21. Gelfand LA, MacKinnon DP, DeRubeis RJ, *et al*. Mediation analysis with survival outcomes: accelerated failure time vs. proportional hazards models. *Front Psychol* 2016;7:423.
22. Szklo M, Nieto FJ. Identifying noncausal associations: Confounding. In: *Epidemiology: beyond the basics*. 2nd edn. Massachusetts, USA: Jones and Bartlett Publishers, 2007: 151–82.
23. Cancer Care Ontario. *Ontario cancer screening performance report 2016*. Toronto, ON: Cancer Care Ontario, 2016: 122.
24. Mahar AL, Kurdyak P, Hanna TP, *et al*. Cancer staging in individuals with a severe psychiatric illness: a cross-sectional study using population-based cancer registry data. *BMC Cancer* 2020;20:476.
25. Afshar N, English DR, Milne RL. Rural-Urban residence and cancer survival in high-income countries: a systematic review. *Cancer* 2019;125:2172–84.
26. Booth CM, Li G, Zhang-Salomons J, *et al*. The impact of socioeconomic status on stage of cancer at diagnosis and survival: a population-based study in Ontario, Canada. *Cancer* 2010;116:4160–7.
27. Coughlin SS. Social determinants of colorectal cancer risk, stage, and survival: a systematic review. *Int J Colorectal Dis* 2020;35:985–95.
28. Tørring ML, Frydenberg M, Hansen RP, *et al*. Evidence of increasing mortality with longer diagnostic intervals for five common cancers: a cohort study in primary care. *Eur J Cancer* 2013;49:2187–98.
29. Tørring ML, Frydenberg M, Hamilton W, *et al*. Diagnostic interval and mortality in colorectal cancer: U-shaped association demonstrated for three different datasets. *J Clin Epidemiol* 2012;65:669–78.
30. Tørring ML, Frydenberg M, Hansen RP, *et al*. Time to diagnosis and mortality in colorectal cancer: a cohort study in primary care. *Br J Cancer* 2011;104:934–40.