SUPPLEMENTAL MATERIAL
Data S1.

Description of Data Sources

1. Discharge abstract database (DAD): used to ascertain exclusion criteria (previous cardiovascular disease). The DAD contains records on all hospital admissions at acute, rehab, chronic, and day surgery hospitals in the province, including demographic, diagnostic, and information on procedures performed.

2. Electronic Medical Record Administrative Data Linked Database (EMRALD): used to determine outcomes (systolic blood pressures, HDL cholesterol, total cholesterol, smoking status), covariates (use of statin medication, use of anti-hypertensive medication), and auxiliary variables in multiple imputation (LDL cholesterol). Includes data from primary practice EMRs including laboratory test data, patient profiles, medications, consultation letters, and other clinical information.

3. Dynacare Medical Laboratories Database (GDML): used to ascertain individual lab values used to construct outcome variable (ACC/AHA Pooled Cohort Equation Risk Scores), secondary outcomes (HDL, total cholesterol), and auxiliary variables in multiple imputation (LDL cholesterol). Includes laboratory testing information for serum or urine samples conducted by Gamma-Dynacare Medical Laboratories, one of the largest commercial labs in the province.

4. Immigration, Refugees and Citizenship Canada Permanent Resident Database (IRCC): used for covariate ascertainment (immigration history). The extract we worked with includes all Canadian immigration applications for individuals landing first in Ontario after 1985. Data provided by Immigration, Refugees, and Citizenship Canada.
5. Ontario Canadian Census Profiles (CENSUS): used for covariate ascertainment (socioeconomic status proxy—dissemination area median household income). Contains census data aggregated to various geographic areas across Ontario.

6. Ontario Chronic Obstructive Pulmonary Disease Database (COPD): used to assess individual comorbidity (COPD). Database identifies individuals in Ontario with prevalent and incident cases of COPD among those 35 years and older based on diagnosis of COPD in physician billing claims or hospital discharge abstracts. Validation identified the algorithm had sensitivity and specificity of 85.0% and 78.5%, respectively.\(^1\)

7. Ontario Diabetes Database (ODD): used to ascertain diabetes outcomes for patients. Database identifies individuals with diabetes mellitus using algorithm based on diabetes diagnoses from physician billings claims and hospital admissions.\(^2\)

8. Ontario Drug Benefit Database (ODB): used to flag individuals residing in long-term care facilities for exclusion. Includes claims for medications insured by provincial insurance for eligible individuals.

9. Ontario Health Insurance Plan (OHIP) billings database: used to ascertain individual comorbidities using the Johns Hopkins Adjusted Clinical Groups and for exclusion criteria. Includes information on physician billings for services provided in Ontario that are insured by provincial health insurance.

10. Registered Persons Database (RPDB): used to define the study sample and for covariate ascertainment (sex, age, area-level socioeconomic status via linkage to Canadian Census data) and auxiliary variables used in multiple imputation (region of residence). Contains records for all individuals residing in Ontario who have been issued a provincial health insurance card.
11. Surname-based Ethnicity Group (ETHNIC): used for covariate ascertainment (ethnicity). The ETHNIC database includes classification of each individual into either South Asian, Chinese, or general population on the basis of surnames (surnames not included in database). Validation studies against self-identified ethnicity indicate specificities of 99.7% and 99.7% and sensitivities of 50.4% and 80.2% for South Asians and Chinese Canadians, respectively.³

12. Walkability database: used to assign exposure values (neighborhood walkability). Includes information on geographic locations and associated walkability values. Details on the index components are provided below.
Variable data collection and definitions

Walkability: The present tool was developed using objective demographic and geographic features specific to the Ontario environment with a previously described protocol. In brief, a literature review was first conducted to assess features of the built environment previously found to be related to transit activity, weight-measured (e.g. body mass index), or perceived walkability. Those factors for which data were available, scalable, and cost-permissible at the dissemination area (DA) level were then entered into a factor analysis to create a summary term explaining the common variance between built environment variables. The final index included four variables: (i) population density (population/km²), (ii) residential density (number of dwellings/km²), (iii) street connectivity (“count of 3-way or greater intersections within a 800 meter network buffer of the tract centroid”), and (iv) number of destinations within the neighbourhood (“count of locations of a given resource type within the 800 meter network buffer of each tract centroid”). Scores were Normalized with a mean of 0 and a variance of 1. Each of the items were correlated with the index with strengths ranging from 0.70 to 0.94 ($p < 0.001$), with a Chronbach’s alpha = 0.85 indicating high internal reliability. Since index values created using factor weights were highly correlated with a simpler index composed of the sum of the 4 normalized components, scores were generated using the latter approach. For analysis, walkability scores were divided into quintiles (Q1-least walkable, Q5-most walkable) based on the whole region that

1 Resource types included grocery stores and fruit & vegetable stands, convenience/variety stores, bank branches, restaurants & cafes (including fast food), and other retail services. For areal geographic units, centroids are defined as the ‘centre of mass’ for that region, often representing the centre of the region (although this may not be the case if the region has an unusual shape, e.g. donut or crescent). Network buffers are those that calculate the distance from the area centroid along roads or paths that can be traveled by individuals instead of distance ‘as the crow flies’.
walkability exposures were available for (Toronto & the Greater Toronto-Hamilton Area, Ottawa, and London).

Johns Hopkins Collapsed Aggregated Diagnostic Groups (CADGs): The Johns Hopkins Adjusted Clinical Groups (ACG) system (Johns Hopkins ACG® System Ver. 7.0) assigns individuals to a series of non-exclusive categories (ADGs, n = 32) based on hospital admissions and billing data. These categories represent clinical and resource-intensity groups based on: duration of condition, severity of condition, diagnostic certainty, etiology, and need for specialty care. Designed to predict individuals’ health resource needs, they have also been demonstrated to be predictive of 1-year mortality in the general population of Ontario and have been used to control for comorbidity in epidemiologic analyses. To enhance the parsimony of the model, a set of 12 collapsed ADGs were generated on the basis of individuals’ ADGs using a JHADG algorithm. A count of comorbidities (1 to 10) was created to describe the burden for each individual. CADG11 is unique in that it is not reflective of morbidity itself, rather the use of preventive or administrative health services. As such it is not counted as a ‘comorbidity’ for descriptive purposes, and was not included in the count. We additionally have excluded CADG12 (pregnancy). For the secondary outcome smoking, since a variety of comorbidities could plausibly be consequences of smoking rather than confounders, we used two CADGs reflecting preventive care use and psychosocial comorbidities.
| Variable                     | Level of measurement & units/levels | Data source       |
|------------------------------|-------------------------------------|-------------------|
| Age                          | Continuous (years)                  | RPDB              |
| Sex                          | Dichotomous (male/female)           | RPDB              |
| Ethnicity                    | Nominal (Chinese, South Asian, General Population) | ETHNIC            |
| Immigration history          | Nominal (Immigration within 5 years of index, immigration 5-10 years prior to index date, other ('long-term resident')) | IRCC              |
| Neighborhood income quintile | Nominal (quintiles)                 | CENSUS (2006)     |
| COPD                         | Dichotomous (Case, non-case)        | COPD              |
| Comorbidities (JHACG)        | Continuous (number, 0 to 10)        | OHIP/DAD using John Hopkins ACG Algorithms |
| Smoking status               | Dichotomous (family physician assessed smoker, not assessed as smoker) | EMRALD            |
| **Use of anti-hypertensive medication** | Dichotomous (prescription, no prescription) | EMRALD |
|---------------------------------------|---------------------------------------------|--------|
| **Use of statin medications**         | Dichotomous (prescription, no prescription) | EMRALD |
**Multiple imputation of missing data**

Several variables had missing values in the present dataset: ACC/AHA Pooled Cohort Equation predicted risks (43.5%), systolic blood pressure (17.3%), total cholesterol (20.9%), HDL cholesterol (21.4%), current smoking status (17.7%), treated with statins within 1 year of HDL or TC measurement (21.4% and 20.9%, respectively), and neighborhood income quintile (0.4%). Missing estimated cardiovascular risk outcomes were primarily due to individuals missing one or more of the constituent components of the score (SBP, HDL, total cholesterol, smoking status) rather than individuals missing all components (1.9%). These variables are primarily outcome variables, although neighborhood income quintile and current smoking status are used as covariates in some models. Simulation studies have suggested that complete case analysis of data where only the outcome variable has missing values does not bias parameter estimates.\(^\text{10}\) However, to address potential loss of statistical power due to missing data and to explore the effect of missing smoking data on secondary outcome models, we performed multiple imputation analyses using each of the above covariates and outcomes in the imputation model. A multiple imputation using chained equations (MICE) approach was used.\(^\text{11}\) This technique has the advantage of being able to appropriately model non-continuous variables using logistic or discriminant analysis, rather than making an assumption of multivariate normality for all missing variables as is used in traditional multiple imputation.

Systolic blood pressure, HDL cholesterol, and total cholesterol were treated as continuous variables, while smoking status, treatment with statins within 1 year prior to HDL or TC measurement, and income quintile were modeled using logistic regression.
Smoking status and both statin variables were modeled with logit functions, while income quintile was modeled with a generalized logit function. Additional variables included in the imputation included diastolic blood pressure (continuous: mmHg; 17.3% missing), low density lipoprotein cholesterol (continuous: mmol/L; 22.3% missing), resource utilization band (RUB) score (0% missing), region of residence (categorical: Toronto, Ottawa, Hamilton, London, Other Greater Toronto Area; 0% missing), and treatment with statins within one year prior to LDL measurement (22.3% missing). Each individual CADG was treated separately rather than as a count, modeled as 12 individual binary variables (0% missing). RUB scores are generated with the JHACG software, and reflect differing levels of predicted resource utilization based on an individual’s demographic and clinical characteristics.

To better approximate normal distributions, HDL, LDL, and total cholesterol were log transformed. Additionally, to better normalize data and preserve imputations in the allowable [0,1] range, 10-year cardiovascular disease risk was logit transformed prior to imputation. All data were converted back to their natural scale prior to analysis. Distributions of variables before and after imputation were checked to ensure imputed values were reasonable. For continuous variables, trace plots were also checked to ensure convergence before imputed values were drawn. Five imputations were generated and models were fit as usual for each of the datasets. Results for each dataset were pooled, with parameter estimates and their standard errors calculated using Rubin’s rules (SAS proc mianalyse).

Two sensitivity analyses were performed to assess the robustness of our approach. First, we performed the imputation without DBP, LDL, and treatment on
statins within 1 year prior to LDL measurement, auxiliary variables that also had some missing data. We did not find our estimates changed materially, so results including these variables are reported. To further check our results to assess whether imputation dramatically changed our estimates, we compared the analyses using multiple imputation with complete case analyses. We found that the results were similar across models, with some contrasts changing status as statistically significant, but no overall change in conclusions (Table S3).
Table S2. Association between neighbourhood walkability and 10-year cardiovascular disease risk stratified by current smoking status.

| Outcome         | Smoking Status | Model 1 OR (95% CI) | Model 2 OR (95% CI) | Model 3 OR (95% CI) |
|-----------------|----------------|---------------------|---------------------|---------------------|
| ACC/AHA ≥ 7.5%  | Non-Smoker     |                     |                     |                     |
| Walkability     |                |                     |                     |                     |
| Q1 (Low)        | 1.15 (1.03, 1.28) | 1.18 (1.06, 1.31) | 1.18 (1.06, 1.32) |                     |
| Q2              | 1.37 (1.24, 1.52) | 1.42 (1.28, 1.57) | 1.42 (1.29, 1.58) |                     |
| Q3              | 1.34 (1.21, 1.48) | 1.37 (1.24, 1.51) | 1.39 (1.26, 1.54) |                     |
| Q4              | 1.24 (1.14, 1.36) | 1.25 (1.15, 1.37) | 1.28 (1.18, 1.40) |                     |
| Q5 (High)       | Ref            | Ref                 | Ref                 | Ref                 |
| Smoker          |                |                     |                     |                     |
| Walkability     |                |                     |                     |                     |
| Q1 (Low)        | 1.08 (0.83, 1.41) | 1.06 (0.81, 1.39) | 1.09 (0.84, 1.41) |                     |
| Q2              | 1.15 (0.94, 1.41) | 1.15 (0.93, 1.42) | 1.17 (0.94, 1.46) |                     |
| Q3              | 1.37 (1.16, 1.61) | 1.39 (1.19, 1.64) | 1.40 (1.19, 1.65) |                     |
| Q4              | 1.17 (1.02, 1.34) | 1.17 (1.02, 1.34) | 1.19 (1.03, 1.37) |                     |
| Q5 (High)       | Ref            | Ref                 | Ref                 | Ref                 |
| ACC/AHA ≥ 10.0% | Non-Smoker     |                     |                     |                     |
| Walkability     |                |                     |                     |                     |
| Q1 (Low)        | 1.19 (1.06, 1.34) | 1.22 (1.09, 1.37) | 1.24 (1.10, 1.40) |                     |
| Q2              | 1.40 (1.24, 1.57) | 1.44 (1.28, 1.62) | 1.45 (1.28, 1.63) |                     |
| Q3              | 1.42 (1.27, 1.59) | 1.44 (1.29, 1.61) | 1.47 (1.32, 1.65) |                     |
| Q4              | 1.30 (1.17, 1.43) | 1.30 (1.18, 1.45) | 1.34 (1.21, 1.49) |                     |
| Q5 (High)       | Ref            | Ref                 | Ref                 | Ref                 |
| Smoker          |                |                     |                     |                     |
| Walkability     |                |                     |                     |                     |
| Q1 (Low)        | 1.14 (0.92, 1.42) | 1.12 (0.90, 1.40) | 1.15 (0.92, 1.44) |                     |
| Q2              | 1.16 (0.96, 1.41) | 1.17 (0.96, 1.42) | 1.19 (0.98, 1.45) |                     |
| Q3              | 1.27 (1.08, 1.50) | 1.30 (1.11, 1.53) | 1.31 (1.11, 1.55) |                     |
| Q4              | 1.09 (0.94, 1.27) | 1.09 (0.94, 1.27) | 1.11 (0.95, 1.30) |                     |
| Q5 (High)       | Ref            | Ref                 | Ref                 | Ref                 |

All estimates presented are adjusted for the covariates described for the main models in the methods section. ACC/AHA: ACC/AHA pooled cohort equation risk score. CI: confidence interval. OR: odds ratio. Ref: reference category. Q: quintile.
Table S3. Model results from complete case analyses.

| Walkability | ACC/AHA ≥ 7.5% OR (95% CI) | ACC/AHA ≥ 10.0% OR (95% CI) | SBP (mmHg) β (95% CI) | Total Cholesterol (mg/dL) β (95% CI) |
|-------------|-----------------------------|-----------------------------|-----------------------|-----------------------------------|
| Q1 (Low)    | 1.05 (0.94, 1.18)           | 1.12 (0.99, 1.27)           | 2.61 (1.99, 3.23)     | 0.56 (-0.97, 2.09)                |
| Q2          | 1.26 (1.13, 1.39)           | 1.24 (1.11, 1.39)           | 2.39 (1.76, 3.01)     | -0.79 (-2.20, 0.63)               |
| Q3          | 1.34 (1.20, 1.48)           | 1.38 (1.23, 1.54)           | 2.17 (1.57, 2.76)     | 0.30 (-1.21, 1.81)                |
| Q4          | 1.20 (1.10, 1.32)           | 1.21 (1.09, 1.34)           | 1.44 (0.94, 1.95)     | 0.44 (-0.81, 1.71)                |
| Q5 (High)   | Ref                         | Ref                         | Ref                   | Ref                               |

| Walkability | HDL Cholesterol (mg/dL) β (95% CI) | Smoking Status OR (95% CI) | DM OR (95% CI) |
|-------------|-----------------------------------|---------------------------|----------------|
| Q1 (Low)    | -1.76 (-2.40, -1.13)              | 0.76 (0.67, 0.85)         | 1.27 (1.11, 1.44) |
| Q2          | -1.33 (-1.96, -0.71)              | 0.72 (0.65, 0.81)         | 1.17 (1.04, 1.33) |
| Q3          | -1.45 (-2.07, -0.82)              | 0.89 (0.80, 0.99)         | 1.14 (1.01, 1.29) |
| Q4          | -0.79 (-1.34, -0.23)              | 0.89 (0.81, 0.97)         | 1.12 (1.00, 1.25) |
| Q5 (High)   | Ref                               | Ref                       | Ref             |

All estimates presented are adjusted for the covariates described for the main models in the methods section. ACC/AHA: ACC/AHA pooled cohort equation risk score. CI: confidence interval. DM: diabetes mellitus. HDL: high density lipoprotein. OR: odds ratio. Ref: reference category. SBP: systolic blood pressure. Q: quintile.
Supplemental References:

1. Gershon A, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physician diagnosed COPD in health administrative databases. COPD. 2009;6:388–94.

2. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: Determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care. 2002;25:512–516.

3. Shah BR, Chiu M, Amin S, Ramani M, Sadry S, Tu J V. Surname lists to identify South Asian and Chinese ethnicity from secondary data in Ontario, Canada: a validation study. BMC Med Res Methodol. 2010;10:42.

4. Glazier RH, Weyman JT, Creatore MI, Gozdyra P, Moineddin R, Matheson FL, Dunn JR, Booth GL. Development and validation of an urban walkability index for Toronto, Canada. Toronto, Ontario: 2012.

5. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Reliability. In: Measurement in Medicine: A Practical Guide. New York: Cambridge University Press; 2011. p. 96–149.

6. The Johns Hopkins ACG System: Excerpt from Version 11.0 Technical Reference Guide. Baltimore: 2015.

7. Austin PC, van Walraven C, Wodchis WP, Newman A, Anderson GM. Using the Johns Hopkins Aggregated Diagnosis Groups (ADGs) to predict mortality in a general adult population cohort in Ontario, Canada. Med Care. 2011;49:932–939.

8. Dubois L, Vogt K, Vinden C, Winick-Ng J, McClure JA, Roshanov PS, Bell CM, Garg AX. Association between day of the week of elective surgery and postoperative mortality. CMAJ. 2017;189:E303–E309.

9. Nayan M, Hamilton RJ, Macdonald EM, Qing L, Mamdani MM, Earle CC, Kulkarni GS, Jarvi KA, Juurlink DN. Vasectomy and risk of prostate cancer: population based matched cohort study. BMJ. 2016;355:i5546.

10. Groenwold RHH, Donders ART, Roes KCB, Harrell FE, Moons KGM. Dealing with missing outcome data in randomized trials and observational studies. Am J Epidemiol. 2012;175:210–7.

11. White I, Royston P, Wood A. Multiple imputation using chained equations: issues for guidance and practice. Stat Med. 2011;30:377–99.