

**Table S1.** Variable derivation.

<b>Variable</b>	<b>Source</b>
<b>Individual-level smoking characteristics (self-report)</b>	
Frequency of smoking	STOP intake and CCHS surveys.
Number of cigarettes smoked per day	STOP intake and CCHS surveys.
Age first tried smoking	STOP intake and CCHS surveys.
Duration of smoking	Age at index – age first tried smoking.
<b>Individual-level sociodemographic characteristics (data linkage)</b>	
Age	Registered Persons Database, contains information on all persons registered with the Ontario Health Insurance Plan (OHIP).
Sex	Registered Persons Database, contains information on all persons registered with the Ontario Health Insurance Plan (OHIP).
Immigrant status	Immigration, Refugee and Citizenship Canada Permanent Resident database. Immigrants, refugee and other immigrants collapsed. If not in database, classified as not immigrant.
Prevalent comorbidities	Validated algorithms (see below) were used to determine prevalence of the following comorbidities: asthma, congestive heart failure, chronic obstructive pulmonary disease, diabetes mellitus, hypertension and myocardial infarction. Prior cancer diagnosis was ascertained using the Ontario Cancer Registry.
Comorbidity burden in 2 years up to index	Aggregated Diagnosis Groups (ADGs) score was based on previous health care use captured in the Canadian Institute for Health Information Discharge Abstract Database, Ontario Health Insurance Plan database, and the Canadian Institute for Health Information Same Day Surgery Database.
<b>Ecological-level sociodemographic characteristics (data linkage)</b>	
Neighbourhood educational quintile	Postal code from Registered Persons Database linked to 2011 Canadian Census data on neighbourhood educational attainment.
Neighbourhood employment quintile	Postal code from Registered Persons Database linked to 2011 Canadian Census data on neighbourhood employment status.
Neighbourhood income quintile and rurality	Composite variable composed of neighbourhood income quintile and rurality of the community. To determine neighbourhood income quintile, residential postal code from the Registered Persons Database was linked with 2006 Statistics Canada census data. Rurality was assessed using the Rurality Index of Ontario (RIO 2008) [1] and linked to residential postal code. Rural was defined by a RIO score $\geq 40$ and urban was defined by a RIO score $< 40$ .

Healthcare service utilization (data linkage)		
Outpatient physician visits	Ontario Health Insurance Plan database.	
Hospitalizations	Canadian Institute for Health Information Discharge Abstract Database and Ontario Mental Health Reporting System.	
ED visits	National Ambulatory Care Reporting System.	
Prevalent comorbidity algorithms		
Comorbidity	Algorithm	Validation
Asthma [2]	≥1 hospitalization <i>or</i> ≥2 physician billings in a 2-year period	80.6% sensitivity 81.4% specificity
Congestive heart failure [3]	≥1 hospitalization <i>or</i> 1 physician billing / ED visit, <i>followed by</i> ≥1 hospitalization / ED visit / physician billing within one year	84.8% sensitivity 97.0% specificity
Chronic obstructive pulmonary disease [4]	≥1 hospitalization <i>or</i> ≥1 physician billing	85.0% sensitivity 78.4% specificity
Diabetes mellitus [5]	≥2 physician billings <i>or</i> ≥1 drug claim in a 1-year period <i>or</i> ≥1 hospitalization	90.0% sensitivity 97.7% specificity
Hypertension [6]	≥1 hospitalization <i>or</i> ≥2 physician billings in a 2-year period.	72% sensitivity 95% specificity
Myocardial infarction [7]	≥1 hospitalization	88.8% sensitivity 92.8% specificity

Abbreviations: CCHS = Canadian Community Health Survey; ED = emergency department; STOP = Smoking Treatment for Ontario Patients.

## References

1. Kralj B. Measuring Rurality-RIO2008\_BASIC: methodology and results. Toronto: Ontario Medical Association 2009.
2. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying patients with physician-diagnosed asthma in health administrative databases. *Canadian Respiratory Journal*. 2009;16(6):183-188.
3. Schultz S, Rothwell D, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic Diseases and Injuries in Canada*. 2013;33(3).
4. 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.
5. 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(3):512-6.
6. Tu K, Campbell NR, Chen Z-L, Cauch-Dudek KJ, McAlister FA. Accuracy of administrative databases in identifying patients with hypertension. *Open Medicine*. 2007;1(1):e18.
7. Austin PC, Daly PA, Tu JV. A multicenter study of the coding accuracy of hospital discharge administrative data for patients admitted to cardiac care units in Ontario. *American Heart Journal*. 2002;144(2):290-296.