Catalogue no. 82-003-X
ISSN 1209-1367

Health Reports

Smoking patterns based on birth-cohort-specific histories from 1965 to 2013, with projections to 2041

by Douglas G. Manuel, Andrew S. Wilton, Carol Bennett, Adrian Rohit Dass, Audrey Laporte, and Theodore R. Holford

Release date: November 18, 2020

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Smoking patterns based on birth-cohort-specific histories from 1965 to 2013, with projections to 2041

by Douglas G. Manuel, Andrew S. Wilton, Carol Bennett, Adrian Rohit Dass, Audrey Laporte, and Theodore R. Holford

DOI: https://www.doi.org/10.25318/82-003-x202001100002-eng

ABSTRACT

Background

Characterizing smoking patterns over time is essential for evaluating the impact of tobacco control interventions and predicting smoking-related mortality. Beginning with a 1920s birth cohort, smoking histories (i.e., estimates of smoking initiation and cessation, and prevalence of current and former smokers) were generated.

Data

The Ontario sample (n = 238,411) of the 2003 to 2013 cycles of the Canadian Community Health Survey, which is conducted biennially, was used to obtain cross-sectional information on current smoking behaviour.

Methods

Age at smoking initiation and age at smoking cessation were used to construct smoking histories for each respondent, up to the survey date. An age-period-cohort model was generated and used to examine survival differences by smoking status. Using the model, and adjusting for survival differences in smoking status, the prevalence of current, former and never smokers was estimated in cohorts from 1920 to 1985. Smoking initiation, cessation and intensity were then estimated for age-specific distributions of each birth cohort. These rates were projected forward through to 2041. Smoking patterns by highest level of education were generated using education-stratified models.

Results

Smoking histories show clear trends over time by sex, cohort and age. If current patterns persist, smoking prevalence is projected to decline to single digits (below 10%) by 2023 for women and 2040 for men.

Interpretation

Birth-cohort-specific smoking histories can be generated using cross-sectional health surveys. These cohort histories can describe smoking patterns over time and into the future. In turn, these histories can be used in micro-simulation models to evaluate historic or planned tobacco control interventions, and to project smoking prevalence.

Keywords

models; statistical prevalence; smoking, epidemiology and trends; smoking cessation; statistics and numerical data; education

AUTHORS

Douglas G. Manuel (dmanuel@ohri.ca), and Carol Bennett are with the Ottawa Hospital Research Institute and ICES in Ottawa, Ontario. Andrew S. Wilton is also with ICES. Douglas G. Manuel is also with Statistics Canada, and the Department of Family Medicine and the School of Epidemiology, Public Health and Preventative Medicine at the University of Ottawa. Adrian Rohit Dass, and Audrey Laporte are with Institute of Health Policy, Management and Evaluation, at the Dalla Lana School of Public Health, University of Toronto, and the Canadian Centre for Health Economics. Theodore R. Holford is with the Department of Biostatistics at the Yale School of Public Health in New Haven, Connecticut.

What is already known on this subject?

- Holford et al. (2014) developed a smoking history generator model that uses birth-cohort-specific smoking histories to model smoking patters. Using United States survey data, the age-period-cohort model predicted historic smoking prevalence that closely approximated observed smoking prevalence.
- The smoking history generator model has been used to project future smoking prevalence, incidence and cessation. Using microsimulation models, the model has also been used to assess the effectiveness of smoking prevention strategies.

What does this study add?

- Smoking patterns based on birth-cohort-specific histories can be generated using the Canadian Community Health Survey and other surveys that contain questions on timing of smoking initiation and cessation.
- This study found a close approximation between modelled and observed smoking prevalence in Canada. The replication of the same modelling approach in two countries (Canada and the United States) suggests that the approach is robust.
- To the authors' knowledge, this the first study to estimate historic smoking initiation and cessation by sociodemographic groups and to project future smoking prevalence. The projections suggest that, in Canada, a large, persistent difference in smoking prevalence will remain through to 2041 between people with a postsecondary education and those with less than a high school diploma.
- There is merit to replicating this study in other countries with similar population health surveys, given the robustness of the approach and the importance of projecting smoking prevention.

In most countries, including Canada, smoking remains the leading cause of mortality that is attributable to health behaviour.[1-3] To monitor progress in reducing smoking prevalence, studies have relied on observing smoking prevalence trends using either general population health surveys, such as the Canadian Community Health Survey (CCHS), or tobacco-focused surveys, such as the Canadian Tobacco, Alcohol and Drugs Survey.[4] Two properties of smoking history are important in assessing the effect of anti-smoking strategies: life-course smoking exposure and smoking initiation and cessation information. Life-course smoking exposure is important because smoking-attributable chronic diseases have a long latency. Initiation and cessation information are essential to assess the two main strategies for reducing smoking prevalence (i.e., strategies to reduce smoking initiation or increase smoking cessation).

This study's objective was to characterize smoking history by sex using birth cohorts beginning in 1920. Smoking histories for each birth cohort included age at smoking initiation and cessation, which was used to construct smoking prevalence for each calendar year from 1971 to 2041. Given known differences in smoking rates by socioeconomic status, and evidence that inequalities are widening over time in Canada and in many other countries,[5-7] a secondary objective was to characterize smoking history by socioeconomic status.

Methods

Data

The Ontario samples from the 2003 to 2004, 2005 to 2006, 2007 to 2008, 2009 to 2010, 2011 to 2012 and 2013 to 2014 CCHS cycles were used to retrospectively reconstruct smoking histories. The CCHS uses a multistage stratified cluster design that represents approximately 98% of the Canadian population aged 12 and older (average response rate of 80.5%).[8] Individuals in each household were automatically selected to participate in the survey using various selection probabilities based on age and household composition. Each respondent was assigned a weight that signified the number of individuals that the respondent represented in the target population. Individuals living on reserves, institutionalized residents, full-time members of the Canadian Forces and residents of certain remote areas were excluded from the survey.

Ascertainment of smoking exposure and sociodemographic status

Smoking exposure was measured by self-reported smoking status at the time of the survey (current, former or never smoker), age of initiation and age of cessation (see Appendix 1). A lifetime smoking history was created for each respondent from birth to survey date. Smoking history begins at birth, when an individual is a "never smoker." Next, there is a possible transition to "current smoker," followed by a further possible transition to "former smoker" (see Appendix 2 for details on the smoking history model). The probability of smoking initiation was assumed to be zero before age 8, and the probability of smoking cessation was assumed to be zero before age 15.[9]

Highest level of education was ascertained from the following two questions: "Did you graduate from high school (secondary school)?" and "What is the highest degree, certificate or diploma you have obtained?"

Immigration status was ascertained from the following two questions: "In what country were you born?" and "In what year did you first come to Canada to live?"

For most variables, less than 2% of data were missing (see Table 1). Missing data were imputed using a multiple imputation model based on sex, birth year, immigration status and ethnicity.[10]

Details on variable transformation and recoding are available in the *cchsflow* library.[11]

Analysis

Smoking history model

This study used the Holford et al. (2014) age-period-cohort approach, which was used to create the National Cancer Institute's smoking history generator model for the population of the United States.[9] The smoking history generator model contains two parts: a smoking initiation model (for people who had not already started smoking) and a smoking cessation model (for those who were currently smoking).

This study differs in two ways from the study conducted by Holford et al.[9] First, immigrants in this study were excluded from the smoking models for the period prior to their immigration to Canada. Second, survival differences by smoking status were estimated using a multivariable predictive model that considered detailed smoking history and other health behaviours.[1] Comparatively, the Holford et al. (2014) approach used a constant mortality risk ratio based on smoking status (see the following section on mortality adjustment for further details).

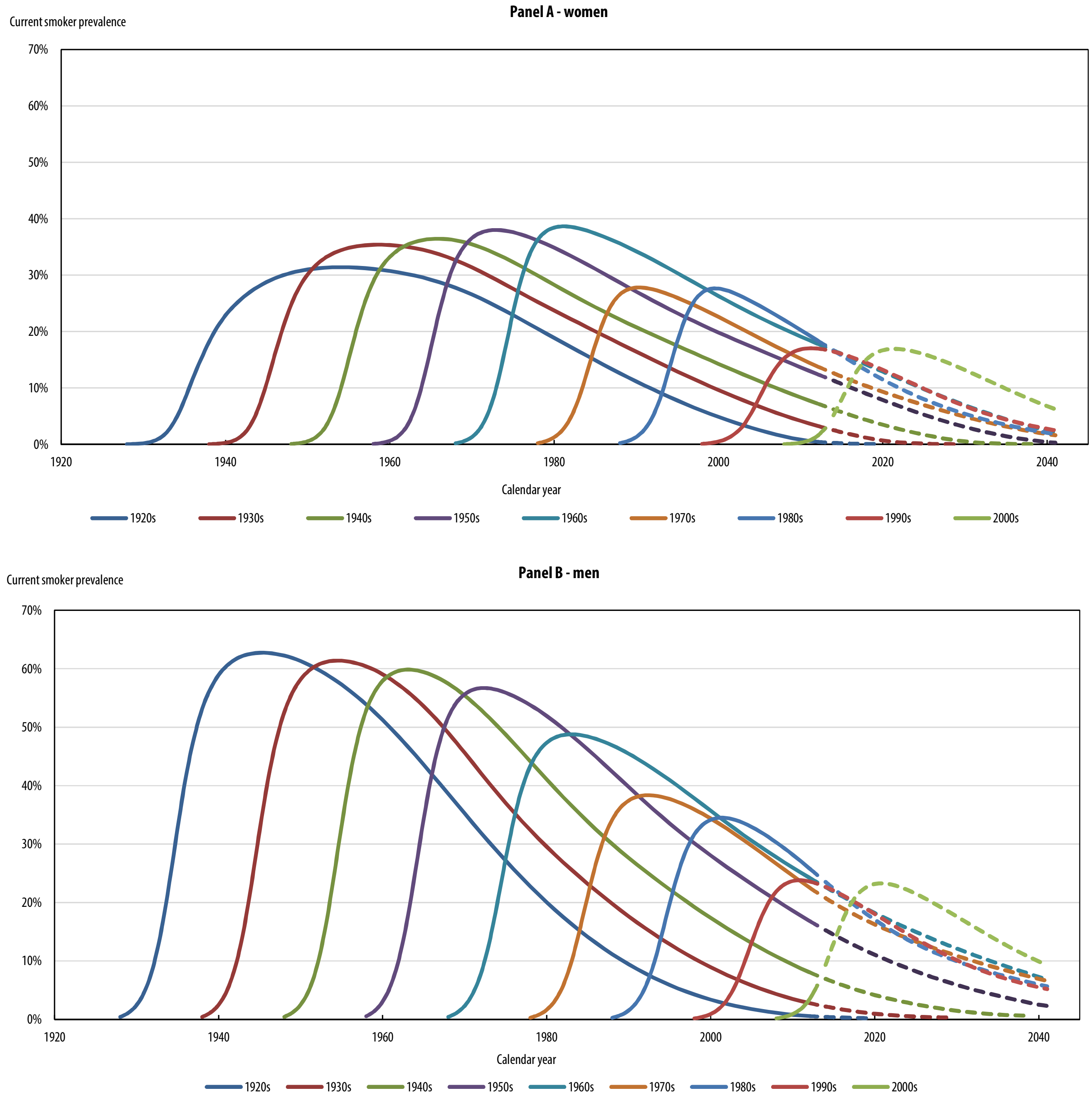
Table 1
Baseline description of study cohorts

	Male cohort		Female cohort		Overall
	Survey sample (N = 107,656)	Represented population‡	Survey sample (N = 130,755)	Represented population‡	Represented population‡
	percent				
Age group (years)					
10 to 19	13.7	12.6	10.9	11.6	12.1
20 to 29	11.8	16.5	11.4	15.7	16.1
30 to 39	13.9	15.9	13.9	16.2	16.0
40 to 49	14.8	18.9	12.9	18.1	18.5
50 to 59	15.8	16.2	16.1	16.0	16.1
60 to 69	15.1	10.9	15.7	11.3	11.1
70 to 79	10.3	6.3	11.8	7.2	6.8
80 to 89	4.3	2.4	6.5	3.5	3.0
90 to 99	0.4	0.2	0.8	0.4	0.3
Smoking status					
Current	22.0	21.9	17.6	15.9	18.9
Former	30.6	25.9	23.8	19.6	22.7
Never	47.4	52.1	58.6	64.5	58.4
Immigrant status					
Immigrant	29.8	36.6	29.8	37.0	36.8
Non-immigrant	68.6	60.8	68.5	60.5	60.7
Missing	1.6	2.5	1.6	2.5	2.5
Ethnicity					
White	76.6	67.1	77.4	67.4	67.3
Non-white	22.1	30.8	21.2	30.5	30.7
Missing	1.4	2.1	1.4	2.0	2.0
Education					
Less than high school	25.2	20.9	23.4	20.1	20.5
High school graduate	23.2	24.1	24.6	24.9	24.5
Postsecondary graduate	49.8	52.8	50.2	52.8	52.8
Missing	1.8	2.2	1.7	2.2	2.2
Survey cycle					
2003 to 2004	17.0	15.7	16.7	15.6	15.7
2005 to 2006	16.6	16.1	16.2	16.0	16.1
2007 to 2008	17.3	16.5	17.2	16.5	16.5
2009 to 2010	16.6	16.9	16.7	17.0	17.0
2011 to 2012	16.2	17.2	16.6	17.2	17.2
2013 to 2014	16.3	17.6	16.6	17.7	17.6

‡ Population estimated using the Canadian Community Health Survey sampling weights.
Source: Canadian Community Health Survey.

An age-period-cohort model was constructed for each part of the model (i.e., the smoking initiation model and the smoking cessation model). The three components of the model were related by $c = t - a$, where c represents the cohort or year of birth, a represents age, and t represents the period or calendar year.[9] The smoking initiation model was the conditional probability that an individual in cohort c started smoking at age a, given that they were a non-smoker at $a - 1$. The cumulative proportion of "ever smokers" (current or former smokers) at age a was estimated using life table methods, adjusted for mortality differences by smoking status and other predictors (see the following section on mortality adjustment).

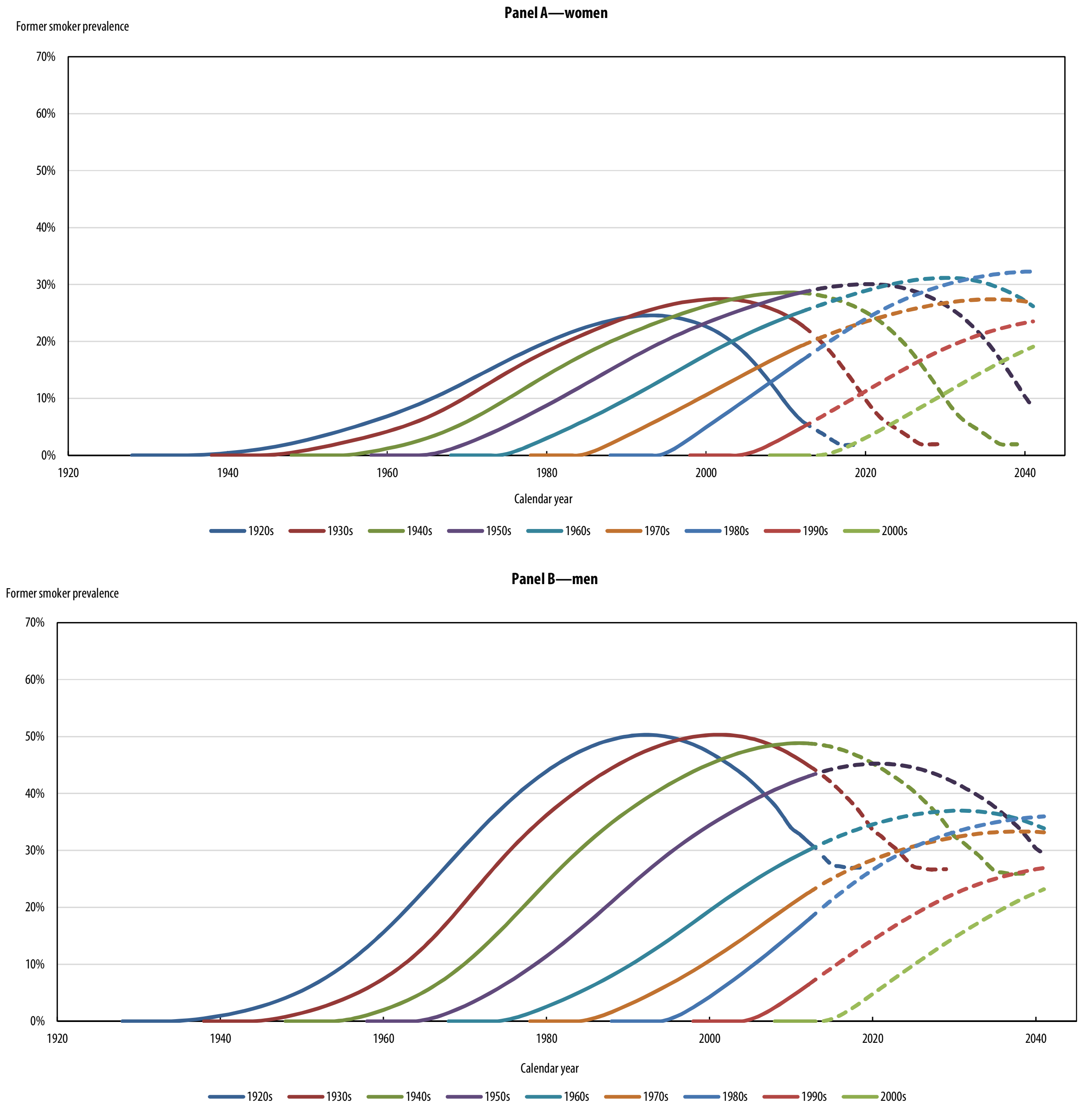
Figure 1
Historic and projected prevalence of current smokers by sex, calendar year and birth cohort, Ontario, 1928 to 2041



Note: Dashed lines indicate extrapolations.

The smoking cessation model was the conditional probability that a current smoker in cohort *c* ceased smoking at age *a*, given that they were a smoker at $a - 1$. The cessation rates provided an estimate, for each sex-age cohort, of the cumulative cessation probability for ever smokers who had not yet quit smoking. The combination of these two models provided the prevalence of current smokers and former smokers in each year for each birth cohort. The model assumes that a person smoked in each year from the age at which they started smoking until either the age they quit (former smoker) or the year of the survey (current smoker).

Figure 2
Historic and projected prevalence of former smokers by sex, calendar year and birth cohort, Ontario, 1928 to 2041



Note: Dashed lines indicate extrapolations.
Source: Author's compilation based on the 2015 Canadian Community Health Survey – Nutrition content.

At the time of this study, 2013 was the last year for which data were available for initiation and cessation estimates. Respondents born after 1985 (i.e., those younger than age 18 in 2003) were not included in the estimates. In the initiation and cessation models, the cohort effect was assumed constant from 1985 forward. The period effect of initiation was assumed constant from 1999 forward for men and from 2003 forward for women. Finally, the period effect of cessation was assumed constant from 2013 forward. The number of respondents born prior to 1920 was small. As such, the period effect was assumed constant prior to 1928, and the cohort effect was assumed constant from 1920 backward.

The model was estimated using the general linear model command (PROC GENMOD) in SAS. Using the combined CCHS surveys, the incidence of initiation (or cessation) was first estimated by age and cohort using an age-cohort linear logistic model. Constrained natural cubic splines were used to provide nonparametric additive effects, using the same knot points described by Holford et al.[9,12]

Mortality adjustment for survival bias from smoking status

Non-smokers were more likely to survive to survey date than smokers, and therefore they produced biased (downward) estimates of smoking initiation. Initiation rates were adjusted for this survival bias using a weight for never smokers and ever smokers in each historical year that reflected the proportion of ever smokers who would have died prior to the survey date. The survival bias weight was calculated using Mortality Population Risk Tool (MPoRT). The MPoRT includes age, smoking status, time since quitting, immigration status and sex, and was used to calculate each respondent's one-year probability of death for each historical year up to the survey date.[1] See Appendix 2 for details.

Population smoking prevalence

Smoking prevalence was modelled based on smoking initiation and cessation, which reflects how long a person smokes (as described in Equation 1).[9] The historic and projected population prevalence reflects an open population (i.e., a population that includes births, deaths, immigration and emigration).

Equation 1: Current smoking prevalence (Pc) for cohorts, based on initiation (ever smokers) and cessation

$$P_c\,(a,c) = P_e\,(a,c) x\, Q(a,c)$$

$P_c\,(a,c)$ = prevalence of current smokers at age a for cohort c

$P_e\,(a,c)$ = prevalence of ever smokers at age a for cohort c (P_e reflects initiation probabilities)

$Q(a,c)$ = cumulative proportion of smokers in cohort c who had not ceased smoking by age a (Q reflects cessation probabilities)

The annual age- and sex-specific prevalence rates for current and former smokers was estimated by multiplying the age- and sex-specific prevalence rates by the historic and future age- and sex-specific population projections. The population estimates were derived in DemoSim, a microsimulation demographic projection model designed to project the Canadian population according to various ethnocultural characteristics.[13] The total or overall prevalence rate was the sum of the age- and sex-specific prevalence estimates, divided by the sum of the age-specific population estimates.

Assessment of model performance

The modelled annual population prevalence (modelled using smoking initiation and cessation rates, and adjusted for the increased mortality risk of smokers) of current and former smokers was compared with historic survey results using the Survey of Smoking Habits (1971 to 1986),[11,14-22] the Canadian Health Survey (1978),[23] the General Social Survey (1985 and 1991),[16,17] the Ontario Health Survey (1990),[18] and the National Population Health Survey (1994 to 1998).[20,21,24] These surveys used different survey methods and smoking ascertainment questions. Where possible, smoking status ascertained by survey reflected the definitions used in the study's model.

Secondary analysis

To assess the impact of socioeconomic status on smoking history, the models were further stratified by highest level of education (postsecondary, high school, less than high school). To approximate highest level of education, only respondents aged 25 or older at the time of survey response were included in the education-stratified models.

Sensitivity analyses

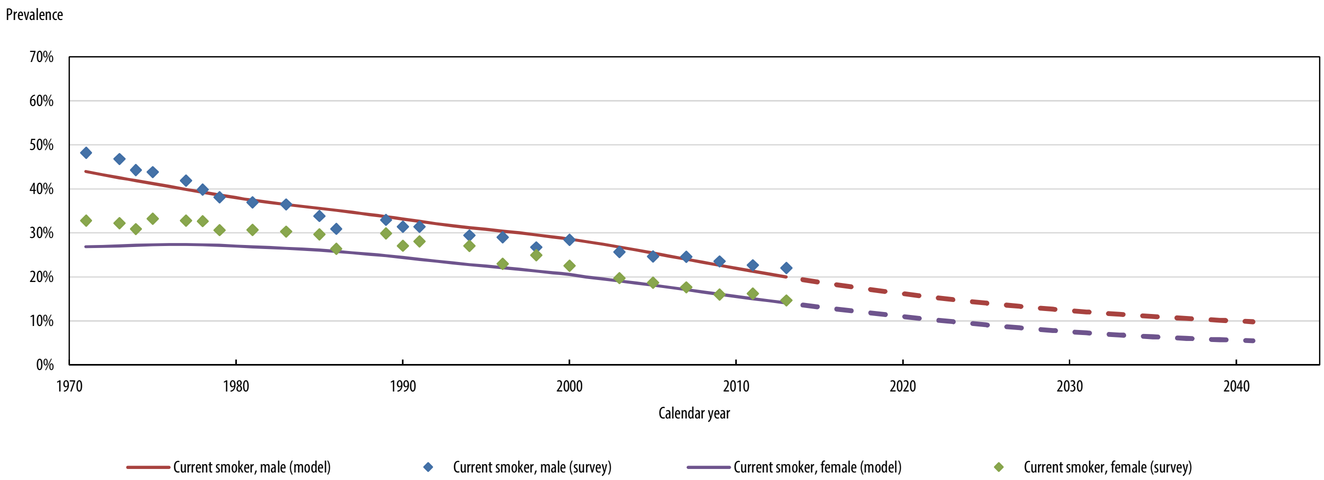
Sensitivity analyses were performed on the case ascertainment of smoking (Appendix 1) and the assumed fixed rates of initiation and cessation for recent and future years in the initiation and cessation models.

Results

Current smoker prevalence was estimated by year for birth cohorts from 1920 to 2000, as shown in Figure 1. For men, current smoker prevalence was highest for the 1920s cohort (peaking in 1945, at 63%) and then declined by almost two-thirds for the most recent cohort (projected maximum of 23% for the birth cohort from 2000 onward). The decline was steepest from 1970 to 2000, when smoking prevention strategies were progressively introduced. Current smoker prevalence in early cohorts was lower for women than for men, with current smoker prevalence increasing until the 1960s cohort (39% maximum in 1973) and then declining (17% projected maximum for the 2000s cohort).

Figure 2 shows the former smoker prevalence for birth cohorts from 1920 to 2000. Men in the cohorts from 1920 onward

Figure 3
Historic and projected prevalence of current smokers among men and women aged 20 and older, by modelled estimates and unadjusted survey results, Ontario, 1971 to 2041



Note: Dashed lines indicate extrapolations.
Source: Canadian Community Health Survey.

showed a decline in former smoking prevalence. Former smoker prevalence declined in all cohorts after age 60 (approximately), which reflects the higher mortality risk for former smokers compared with never smokers, as well as lower smoking initiation rates. For women, former smoker prevalence increased until the 1980s cohort and then declined.

The annual initiation and cessation probabilities upon which the modelled prevalence estimates were generated are shown in Appendix 3. For both men and women, the age at which they started smoking remained virtually unchanged over the past 80 years. Across all cohorts, the peak age of initiation was 15 to 17 for women and 16 for men. Over this same period, the likelihood of smoking initiation decreased quickly after age 16, with few people older than age 30 ever starting to smoke. Women born during the 1920s and 1930s were an exception: they continued to start smoking in their 30s and 40s. For both men and women, the likelihood of smoking cessation generally increased with age, and the pattern was more pronounced for women. Over the past 80 years, the increasing rate of smoking cessation contributed to the steady increase in the prevalence of former smokers. This also means that smokers in more recent birth cohorts quit smoking at earlier ages and smoked for fewer years than smokers from previous birth cohorts. Historically, few people quit smoking when they were young, and instead quit when they were in their 40s or older. In more recent years, it has become more common for smokers to quit at younger ages.

Current smoker prevalence has dropped steadily since the 1970s (Figure 3). Current smoker prevalence is projected to drop below 10% (single-digit prevalence) by 2023 for women and by 2040 for men. The projected decline in smoker prevalence is a result of recent cohorts being less likely to start smoking and, among ever smokers, being more likely to stop smoking earlier than historic cohorts. Current smoker prevalence is projected to continue to decline, even if there are no future improvements in smoking initiation and cessation rates. This inevitable decrease in current smoker prevalence is a consequence of older people (who smoke at higher rates) being slowly replaced by their children and grandchildren, who smoke at much lower rates. For men, the modelled current smoker prevalence closely approximated historic smoking surveys. For women, there was a modest underestimation of modelled smoker prevalence compared with surveys prior to 2000.

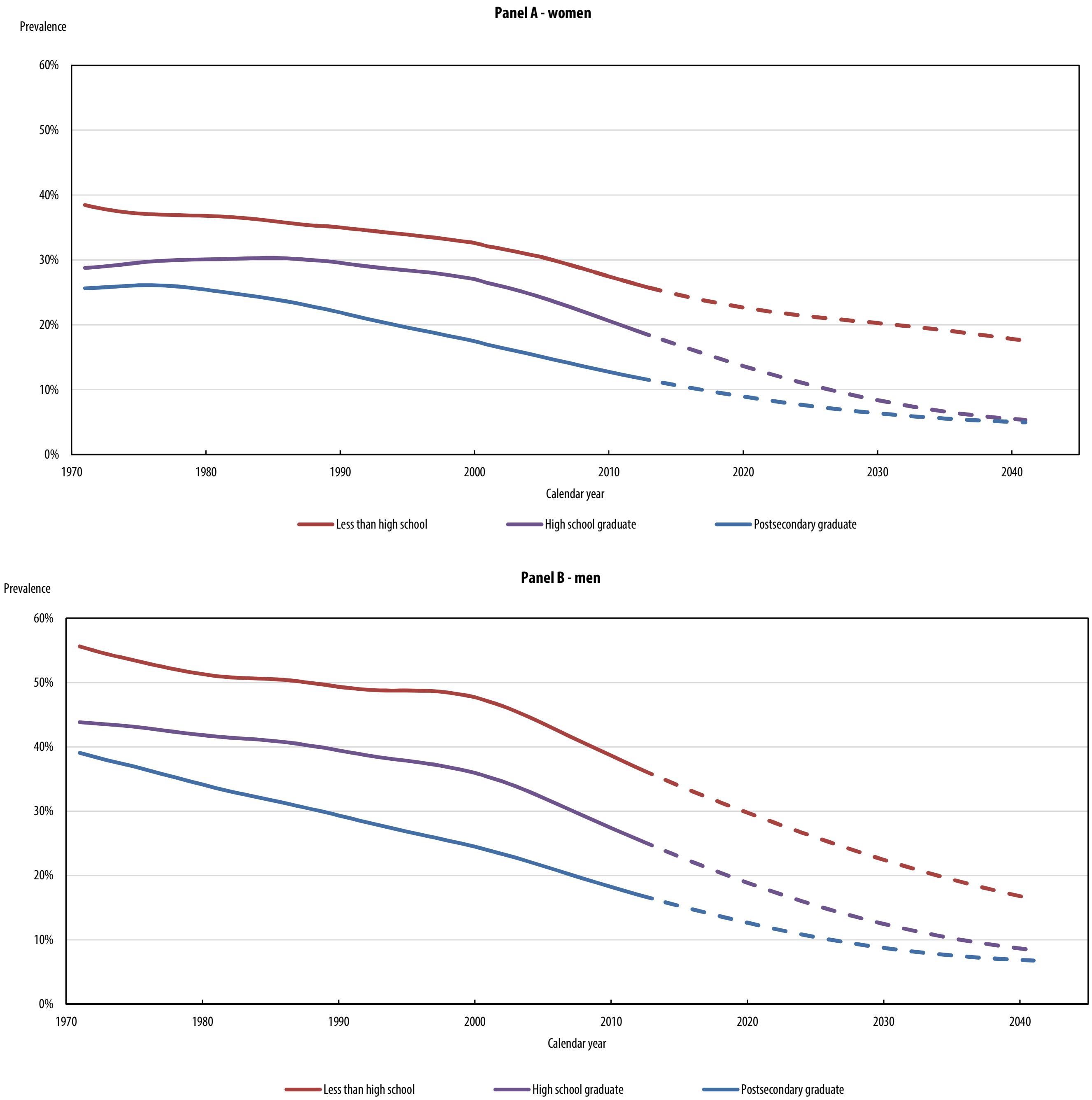
Former smoker prevalence has increased steadily since the 1970s, but plateaued in recent years (Appendix 4). The steady increase in former smoker prevalence is a consequence of the high historic smoking initiation rates combined with the long period that a person is classified as a former smoker (until they die).

Current smoker prevalence stratified by highest level of education is shown in Figure 4. From 1971 through to 2041, respondents whose highest level of education was less than a high school diploma had persistently higher rates of current smoker prevalence compared with other respondents. Those with less than a high school diploma were not projected to achieve single-digit current smoker prevalence during the study period. Conversely, women and men with a postsecondary education were projected to attain single-digit current smoker prevalence by 2017 and 2027, respectively.

Discussion

This study shows that smoking patterns based on birth-cohort-specific histories can be generated using cross-sectional surveys

Figure 4
Current smoker, stratified model



Note: Dashed lines indicate extrapolations.
Source: Author's compilation based on the 2015 Canadian Community Health Survey – Nutrition content.

that contain questions on the timing of smoking initiation and cessation. Using the birth-cohort models, current smoker prevalence was projected to decline to single digits (less than 10% by 2023 for women, and 2040 for men) in Ontario, Canada. However, this decline to single digits is not universal across socioeconomic groups. Current smoker prevalence is projected to continue to decline, even if there are no future improvements in initiation and cessation, because new cohorts are projected to have a lower lifetime smoking prevalence and will progressively replace historic cohorts that had much higher lifetime smoking prevalence.

Information on smoking initiation and cessation is crucial for evaluating anti-smoking strategies, given that these strategies often aim to either reduce initiation or increase cessation. Cohort-specific histories form the basis of robust population models whose purpose is to project smoking prevalence and evaluate tobacco-control strategies designed to change smoking patterns.[25] In Canada, this smoking model was used to estimate a $51-billion reduction in smoking-attributable health care expenditures between 2003 and 2041.[26] In the United States, tobacco control was assessed using a similar model, and was found to be associated with the avoidance of 8 million premature deaths and an estimated extended mean life span of 19 to 20 years.[25]

Comparison with other studies

This study replicated the Holford et al. (2014) modelling approach, which used United Stated health surveys from 1965 to 2009.[9] Compared with the Holford et al. study, the present study used historic health surveys that covered a shorter period (11 years, from 2003 to 2014), but had a larger sample size for each available survey. Additionally, because the historic data available covered a shorter period, this study had a shorter period of historic cohorts (starting in 1920 versus 1890). Despite these differences, this study was able to replicate accurate smoking histories. The successful creation of smoking histories in Canada and the United States using the same approach suggests that the modelling approach is robust. Further replication in other countries with different periods and intensity of policy implementation could provide further data for evaluating prevention and cessation strategies.

To the authors' knowledge, this is the first study to estimate historic smoking initiation and cessation by sociodemographic status and to project future smoking prevalence. There are significant inequities in smoking-attributable deaths and diseases in Canada, but the degree to which these inequities will persist has been unclear.[1,27] This study's projections suggest that a large, persistent difference in smoking prevalence will remain between people with a postsecondary education and those with less than a high school diploma.

In the Holford et al. (2014) study, historic smoking prevalence more closely approximated historic surveys, likely because the historic data used in the model covered a longer period (46 years). It is likely that more years of historic data improve the characterization of smoking histories and are important for evaluating historic smoking prevention and reduction strategies. However, this study's close adherence to more recent prevalence estimates using only 11 years of historic data suggests that future projections can be generated with historic data that cover shorter periods.

This study builds on previous studies that used Canada's population health model to project smoking prevalence using linear interpolation and a probability matrix.[28,29] The current study showed a closer approximation to historical smoking prevalence and a steeper decline in future smoking prevalence. Other studies have also generated historic cohorts using an observed rather than a modelled approach, with similar findings.[30] Beyond Canada and the United States, the authors are not aware of any age-period-cohort models of smoking history that can be used to project smoking prevalence. However, there are historic cohorts for a number of countries.[30]

Limitations

Although the modelled estimates closely approximated survey estimates from the 2003 to 2013 cycles of the CCHS, there was modest underreporting of current smoking for women compared with other historic survey-reported prevalence. It is noteworthy that this study modelled smoking prevalence, unlike other studies that generated historic prevalence entirely from recall.[30] Furthermore, comparisons with previous health surveys are not straightforward because of the different survey designs (population coverage, sampling approach, etc.), response rates and smoking questions.

Several assumptions affect this study's results. The smoking history model's main assumption is that CCHS respondents accurately reported their historic smoking patterns, specifically when they started and stopped smoking. Although there is strong validation for self-reported smoking status among Canadians,[31] there is no similar validation for age of smoking initiation or cessation. That stated, this study's comparison of the modelled prevalence estimates with historical surveys found a close approximation between the two approaches, which suggests that any telescoping effect was modest. There are other supporting studies and improved ascertainment methods that consider smoking duration, quantity and quality.[30,32,33]

A constant smoking initiation and cessation rate after 1985 was also assumed. This assumption did not affect model accuracy up to 2013, but future projections may be lower—if the effect of control measures or social influence increase—or higher—because of the introduction of vaping, which may translate into increased smoking initiation.

This study's analysis suggests that differences in smoking rates by socioeconomic status will persist over the next couple of decades. However, the analysis by socioeconomic status was challenging. The biggest challenge was defining life-course socioeconomic status using the cross-sectional CCHS data. Although level of education was the most feasible measure, the extent to which it corresponds to socioeconomic status over generations may have changed over time.

Finally, for this study, the smoking questions were limited to cigarette smoking and did not include pipes, cigars, e-cigarettes, vaping, etc. The impact on smoking prevalence of more recent legislation with respect to e-cigarettes and other vaping products may not be adequately captured by the current model if smoking initiation and cessation patterns are impacted by these products. That stated, the model can be used to assess potential effects relative to the status quo.

Future directions

This study's focus was Ontario, but CCHS data are available for all of Canada. Adding national data should improve the robustness of the model. Given the robustness of the approach and the importance of projecting smoking prevention, further replication in other countries with similar population health surveys that include questions on smoking initiation and cessation would be beneficial. The models used in the United States and elsewhere have also included smoking intensity which, if incorporated into the Canadian data, would enhance studies on smoking-attributable burdens in the Canadian context.[9,32]

Conclusions

Birth-cohort-specific smoking histories can be generated using cross-sectional health surveys. Cohort smoking histories can describe smoking patterns over time and into the future, which in turn can be used in microsimulation models to evaluate historic or planned tobacco control interventions, and project smoking prevalence.

Acknowledgements

The authors would like to acknowledge the contributions of Rochelle Garner and Bill Flanagan, who provided methodological advice. Data were linked using unique encoded identifiers and analyzed at ICES (formerly the Institute for Clinical Evaluative Sciences). The dataset from this study is held securely in coded form at ICES. While data sharing agreements prohibit ICES from making the dataset publicly available, access may be granted to those who meet the criteria for confidential access, available at www.ices.on.ca/DAS. The data can also be accessed at Statistics Canada's research data centres. The CCHS has a public use file (PUMF) with an open licence, which could be used to replicate this study. However, the PUMF categorizes the age variable into five-year age groups.

The full dataset creation plan and underlying analytical code are available from the authors upon request, with the understanding that the computer programs may rely on coding templates or macros that are unique to ICES and are therefore either inaccessible or may require modification. Data pre-processing and transformations are available in *cchsflow*, an R package that can be adapted to other programming languages.[11]

Conflict of interest statement: This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by ICES or the Ontario MOHLTC is intended or should be inferred. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Financial disclosure: No financial disclosures were reported by the authors of this paper.

APPENDIX

Appendix 1
Ascertaining smoking exposure from the Canadian Community Health Survey

See *cchsflow* for details on ascertaining smoking exposure for different versions of the Canadian Community Health Survey.

Canadian Community Health Survey – flow: Transforming and Harmonizing CCHS Variables (available in English only)

https://cran.r-project.org/web/packages/cchsflow/index.html or

Canadian Community Health Survey – flow (available in English only)

https://big-life-lab.github.io/cchsflow/

Smoking status questions

1a) Have you smoked >100 cigarettes in your lifetime? Yes/No/Don't know

1b) Have you ever smoked a whole cigarette? Yes/No/Don't know

2) What is your current smoking status? Daily/Occasional/Non-smoker/Don't know

Smoking initiation question

Age at which you smoked first whole cigarette? #/Don't know

Smoking cessation questions

When did you stop smoking?

How many years ago did you stop completely? #/Don't know (derived from previous question)

Non-smokers are defined as individuals who smoked fewer than 100 cigarettes in their lifetime or, if they did not know if they had smoked fewer than 100 cigarettes in their lifetime, individuals who had never smoked a whole cigarette.

Current smokers are defined as individuals who smoked at least 100 cigarettes in their lifetime or, if they did not know how many cigarettes they had smoked in their lifetime, individuals who gave an age of initiation and reported their current smoking status as "daily" or "occasional."

Former smokers are defined as current non-smokers who smoked at least 100 cigarettes in their lifetime.

Age at cessation is defined as numbers of years since the individual stopped smoking completely, subtracted from their age at the time of the survey.

Exclusions were made if start or stop (where appropriate) age could not be determined, or if responses were inconsistent. Where available, age at first whole cigarette was used as the start age.

Case ascertainment sensitivity analyses

Several different approaches were assessed to ascertain smoking status. These approaches had a small influence on the study's main findings.

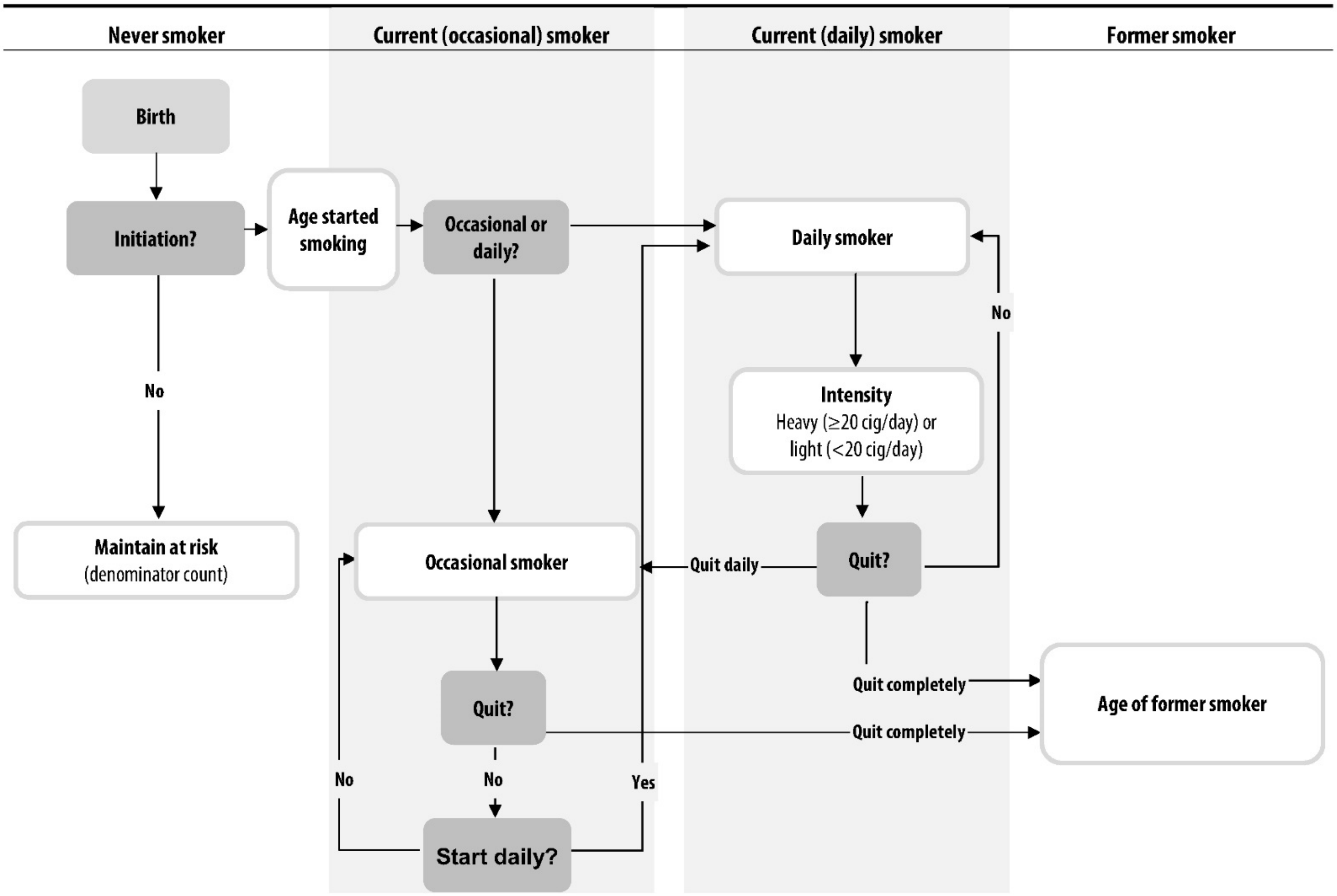
First, in more recent surveys, those who smoked fewer than 100 lifetime cigarettes were classified as occasional smokers, not current smokers. In this study's primary analysis, these occasional smokers were treated the same as current daily smokers. In the first sensitivity analysis, these individuals were excluded (i.e., treated as "never smokers").

The second sensitivity analysis classified those who smoked less than once a year as never smokers instead of current smokers (how they were classified in the primary analysis).

In the primary analysis, because smoking cessation is often unsuccessful, people were considered to be former smokers only if they had ceased smoking for at least two years before the survey interview; otherwise, their observation period was truncated to their age of reported cessation. In the third sensitivity analysis, this truncation rule was reduced to one year, or there with no truncation and anyone who ceased smoking was classified as a former smoker.

Appendix 2

Smoking history model—lifetime smoking patterns



Source: Author's compilation based on the 2015 Canadian Community Health Survey – Nutrition content.

A simple compartment model was used for smoking history. The model begins at birth, when individuals are classified as "never smokers," and then there is a possible transition to "current smoker," followed by a further possible transition to "former smoker." Yearly transition probabilities were conditional on a hypothetical scenario of no transitions. Smoking history parameters were estimated using an age-period-cohort model. Temporal indicators are related by $c = t - a$, where a represents age, t represents period or calendar year, and c represents cohort or year of birth.

Mortality adjustment for survival bias from smoking status

Non-smokers are more likely to survive to survey date than smokers, which produces biased (downward) estimates of smoking initiation. Initiation rates were adjusted for this survival bias using a weight for never smokers and "ever smokers" (i.e., current or former smokers) in each historical year. This weight reflected the proportion of ever smokers that would have died prior to the survey date. The survival bias weight was calculated using the Mortality Population Risk Tool (MPoRT). The MPoRT includes age, smoking status, time since quitting, immigration status and sex, and was used to calculate each respondent's one-year probability of death for each historical year up to the survey date.[1]

First, mortality adjustment was performed for all respondents for the years when they were aged 30 or older and living in Canada. Differential mortality was not applied before age 30, under the assumption that smoking status does not have a differential effect on survival at that stage of life. The MPoRT model assumes that the proportional hazard of death for smokers is higher than that for non-smokers, and that it varies over time depending on an individual's level and intensity of smoking (heavy smoker, light smoker, former heavy smoker, former light smoker) and time since quitting.

Second, weighted age–sex mean probabilities of death were calculated yearly and compared with the published yearly age–sex probabilities of death to form a scaling factor for each year/age/sex group.[12] Third, each respondent's one-year probability of death (from MPoRT) was multiplied by the appropriate year/age/sex scaling factor to obtain a scaled probability. For years in which respondents were younger than 30 years of age, the published mortality year/age/sex probability was used as the "scaled probability." Finally, each respondent's scaled probability was used to calculate the probability of survival from any year to the survey response date, as the product of all one-year survival probabilities from that year to the survey year. The final yearly weight was calculated as the mean of the cumulative survival probabilities in each sex/smoking-status group. These weights were used in the denominator of **Equation 1** to estimate historical smoking incidence.

Equation 1: Proportion of historic smokers adjusted for survival differences between smokers and non-smokers

$$\boldsymbol{P_{tt}} = \frac{\boldsymbol{P_{tT}/S^{s}_{tT}}}{\boldsymbol{P_{tT}/S^{s}_{tT} + \left(1 - P_{tT}\right)/S^{N}_{tT}}}$$

P_{tt} = adjusted incidence proportion

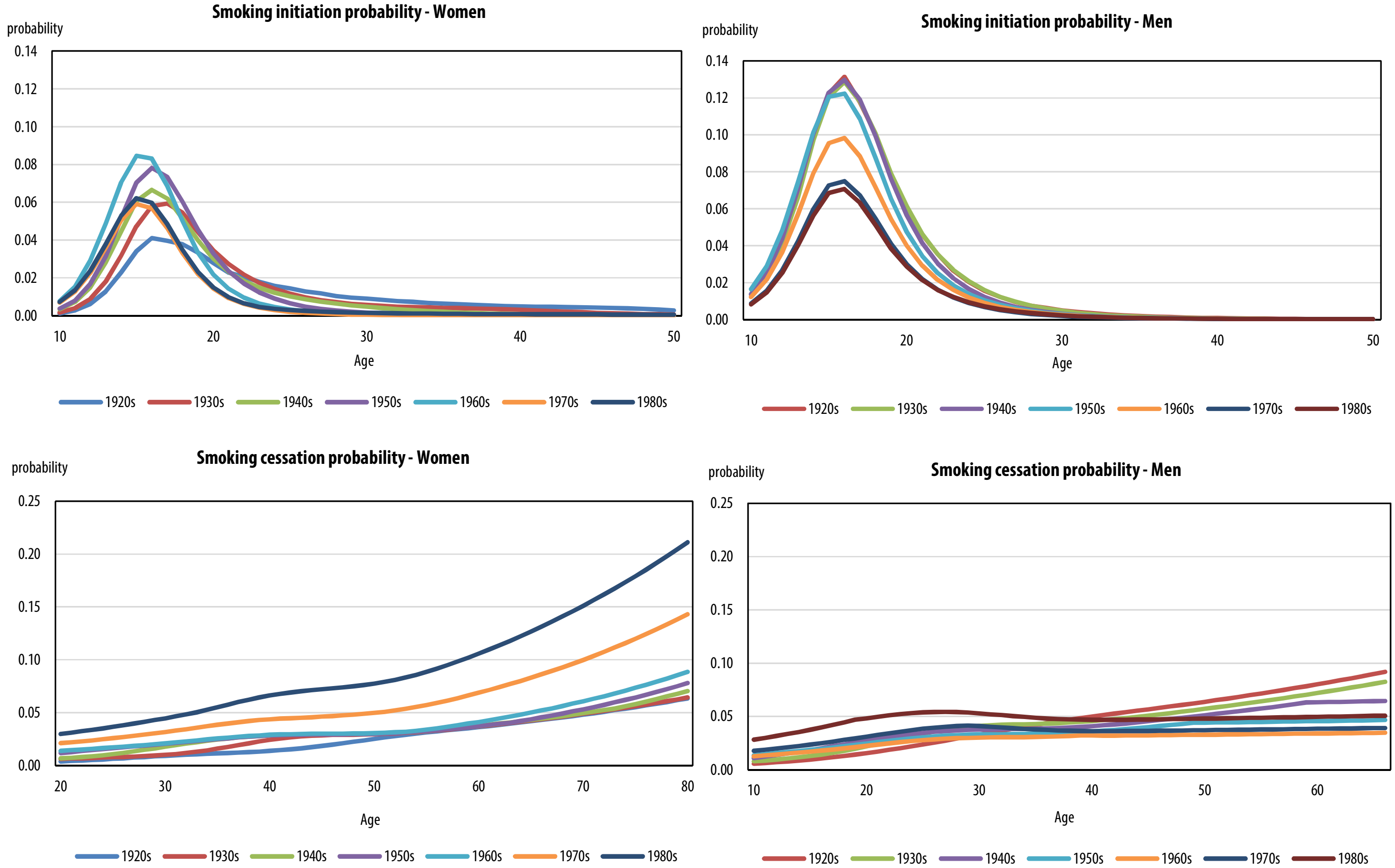
P_{tT} = crude incidence proportion

S^{S}_{tT} = survival probability of ever smokers at time T to survey

S^{N}_{tT} = survival probability of never smokers at time T to survey

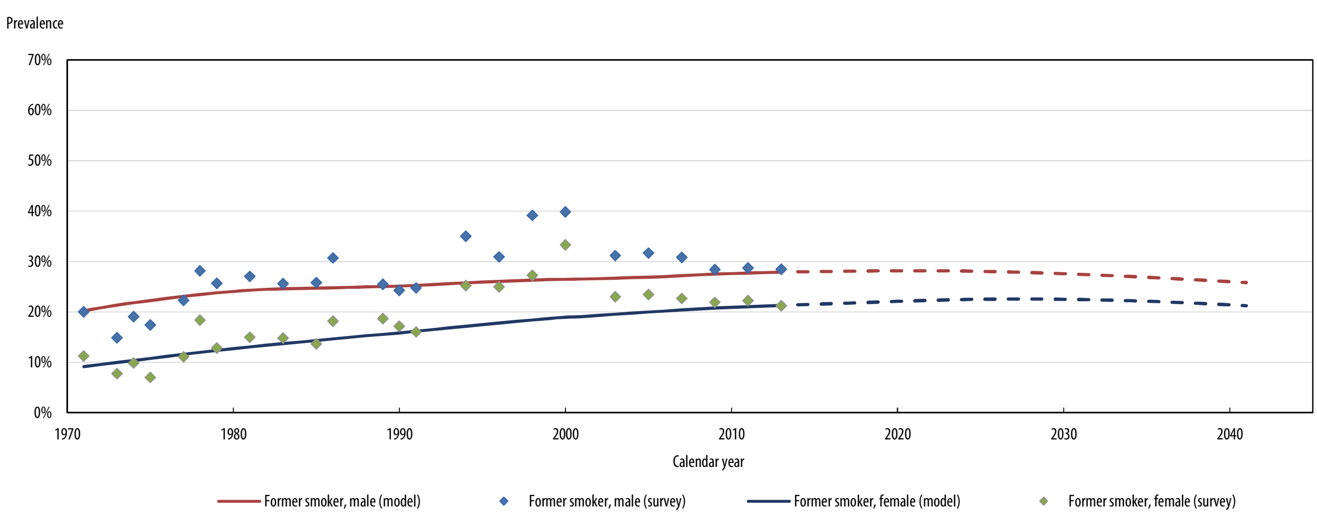
For population prevalence estimates, a second mortality adjustment was applied to account for the likelihood that never smokers have a higher one-year survival probability than ever smokers. Using the scaled one-year survival probabilities found above, the probability of death (as 1 – mean one-year survival probability) was calculated at each age for never smokers and ever smokers. The ratio of mean probabilities of death was used as the mortality adjustment on the proportion of ever smokers in the population.

Appendix 3
Probability of smoking initiation and cessation by sex, age, and birth cohort in Ontario



Source: Author's compilation based on the 2015 Canadian Community Health Survey - Nutrition content.

Appendix 4
Historic and projected prevalence of former smokers among men and women aged 20 and older, by modelled estimates and unadjusted survey results, Ontario, 1971 to 2041



Note: Dashed lines indicate extrapolations.
Source: Author's compilation based on the 2015 Canadian Community Health Survey – Nutrition content.

References

1. Manuel DG, Perez R, Sanmartin C, et al. Measuring Burden of Unhealthy Behaviours Using a Multivariable Predictive Approach: Life Expectancy Lost in Canada Attributable to Smoking, Alcohol, Physical Inactivity, and Diet. *PLoS Medicine* 2016; 13(8): e1002082.
2. Manuel DG, Perez R, Bennett C, et al. *Seven More Years: The impact of smoking, alcohol, diet, physical activity and stress on health and life expectancy in Ontario.* ICES/PHO Report. Toronto: Institute for Clinical Evaluative Sciences and Public Health Ontario. 2012.
3. Forouzanfar MH, Alexander L, Anderson HR, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet.* 2015.
4. Gagné T. Estimation of smoking prevalence in Canada: Implications of survey characteristics in the CCHS and CTUMS/CTADS. *Canadian Journal of Public Health* 2017; 108(3): 331-334.
5. Hitchman SC, Fong GT, Zanna MP, Thrasher JF, Chung-Hall J, Siahpush M. Socioeconomic status and smokers' number of smoking friends: findings from the International Tobacco Control (ITC) Four Country Survey. *Drug and Alcohol Dependence* 2014; 143(1): 158-166.
6. Corsi DJ, Boyle MH, Lear SA, Chow CK, Teo KK, Subramanian SV. Trends in smoking in Canada from 1950 to 2011: progression of the tobacco epidemic according to socioeconomic status and geography. *Cancer Causes & Control* 2014; 25(1): 45-57.
7. Corsi DJ, Lear SA, Chow CK, Subramanian SV, Boyle MH, Teo KK. Socioeconomic and geographic patterning of smoking behaviour in Canada: a cross-sectional multilevel analysis. *PLOS One* 2013; 8(2): e57646.
8. Beland Y. Canadian Community Health Survey - Methodological Overview. *Health Reports* 2002; 13(2): 9-14.
9. Holford TR, Levy DT, McKay LA, et al. Patterns of birth cohort-specific smoking histories, 1965-2009. *American Journal of Preventive Medicine* 2014; 46(2): e31-37.
10. *HMisc Package* [computer program]. http://biostat.mc.vanderbilt.edu/wiki/Main/Hmisc: GitHub; 2015.
11. Manuel DG, Yusuf W, Tuna M, Badder A, Yulric S, Carol B. cchsflow - An R package for the harmonization of variables across the survey cycles. Open Science Foundation. https://osf.io/hkuy3/. Published 2019. Accessed Nov. 2019.
12. Durrleman S, Simon R. Flexible regression models with cubic splines. *Statistics in Medicine* 1989; 8(5): 551-561.
13. Caron-Malenfant E, Coulombe S. *Demosim: An Overview of Methods and Data Sources.* Ottawa: Statistics Canada, 2015.
14. *Smoking habits of Canadians, 1973.* National Archives of Canada (G0000702) (Req. 194). Ottawa: Statistics Canada.
15. *Canada Health Survey, 1978.* Data Liberation Initiative. Ottawa: Statistics Canada.
16. General Social Survey, Cycle 1, 1985 [Canada]: Health and Social Support. Data Liberation Initiative, 1986. Ottawa: Statistics Canada.
17. *General Social Survey, Cycle 6, 1991 [Canada]: Health.* Data Liberation Initiative, 1992. Ottawa: Statistics Canada.
18. *Ontario Health Survey, 1990.* Ontario Ministry of Health, 1991.
19. Manuel D, Perez R, Bennett C, et al. A $4.9 Billion Decrease in Health Care Expenditure: The Ten-Year Impact of Improving Smoking, Alcohol, Diet and Physical Activity in Ontario. Toronto, Ontario: The Institute for Clinical Evaluative Sciences, April 2016.
20. *National Population Health Survey, 1998-99 [Canada]: Health.* Data Liberation Initiative, 1999. Ottawa: Statistics Canada.
21. National Population Health Survey, 1994-95 [Canada]: Health File. Data Liberation Initiative, 1995. Ottawa: Statistics Canada.
22. Costanzo S, de Gaetano G, Di Castelnuovo A, Djoussé L, Poli A, van Velden DP. Moderate alcohol consumption and lower total mortality risk: Justified doubts or established facts? *Nutrition, Metabolism and Cardiovascular Diseases* 2019; 29(10): 1003-1008.
23. Michie S, Van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science* 2011; 6(1): 42.
24. National Population Health Survey, 1996-1997 [Canada]: Household Component - Health File. Data Liberation Initiative, 2015. Ottawa: Statistics Canada.
25. Holford TR, Meza R, Warner KE, et al. Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964-2012. *Journal of the American Medical Association* 2014; 311 (2): 164-171.
26. Manuel D, Wilton AS, Rohit Dass A, Laporte A, Gandhi S, Bennett C. *Health Care Cost of Smoking in Ontario: 2003 to 2041.* Toronto, Ontario: Institute for Clinical Evaluative Sciences. 2018.
27. James PD, Wilkins R, Detsky AS, Tugwell P, Manuel DG. Avoidable mortality by neighbourhood income in Canada: 25 years after the establishment of universal health insurance. *Journal of Epidemiology and Community Health* 2007; 61(4): 287-296.
28. Hennessy DA, Flanagan WM, Tanuseputro P, et al. The Population Health Model (POHEM): an overview of rationale, methods and applications. *Population Health Metrics* 2015; 13(1).
29. Manuel DG, Tuna M, Bennett C, et al. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: the Cardiovascular Disease Population Risk Tool (CVDPoRT). *Canadian Medical Association Journal* 2018; 190(29): E871-E882.
30. Lillard DR, Christopoulou R. *Life-course Smoking Behavior: Patterns and National Context in Ten Countries.* Oxford University Press, 2015.
31. Wong SL, Shields M, Leatherdale S, Malaison E, Hammond D. Assessment of validity of self-reported smoking status. *Health Reports* 2012; 23(1): 47-53.

32. Christopoulou R, Lillard DR. A Novel Indicator of Life-Course Smoking Prevalence in the United States Combining Popularity, Duration, Quantity, and Quality of Smoking. *American Journal of Public Health* 2016:e1-e7.

33. Lillard DR, Christopoulou R, Lacruz AG. Validation of a method for reconstructing historical rates of smoking prevalence. *American Journal of Epidemiology* 2014; 180(6): 656.