

Appendix 2 (as supplied by the authors): Supplementary information on methods

Table of Contents

Methods	2
eTable 2A: List of variables used collected in the sleep laboratory.....	3
eTable 2B: Description of health administrative datasets used.	4
eTable 2C: Categorization of cancers by cancer subtypes.....	5
References.....	6

Methods

Imputation of missing values

To address missing values, we used multivariate imputation by chained equations to generate five complete data sets. Eighty two variables including both the event status and the survival time were chosen for the imputation model. The following built-in imputation models were used in our analyses: for continuous variables, predictive mean matching; for binary variables, logistic regression; for unordered categorical variables, polytomous logistic regression; and for ordered categorical variables, proportional odds. The separate estimates and standard errors from imputed data sets were pooled according to Rubin's rules (1).

Model diagnostics and performance

The proportional hazards assumption for each variable was tested (2,3).

We used the bootstrap for internal validation (optimism for C-index and R²) and over-fitting-corrected calibration (predicted versus observed five-year survival) (2). Discriminative ability was assessed using Harrell's C-index and predictive ability using the model likelihood ratio χ^2 statistic (2). Likelihood ratio test was used to assess the effect of predictors on a model fit including non-linear component of continuous predictors and interactions.

More details about methods used are reported elsewhere (4).

eTable 2A: List of variables used collected in the sleep laboratory (each patient in the cohort underwent an overnight full standard PSG* recording which was scored manually by a sleep technologist and reviewed by a board-certified sleep physician). More details about variables collected in the sleep laboratory are reported elsewhere (4).

Name	Description	Measurement
Demographic characteristics (self-reported by patients)		
Sex		Male/Female
Age		years
Symptoms related to OSA (self-reported by patients)		
<i>Daytime sleepiness</i>		
Epworth Sleepiness Scale, total	Total score	ranged from 0 to 24
Item #1	Sitting and reading	0 = would never doze 1 = slight chance of dozing 2 = moderate chance of dozing 3 = high chance of dozing
Item #2	Watching TV	from 0 to 3
Item #3	Sitting inactive in a public place (e.g. a theatre or at a meeting)	from 0 to 3
Item #4	As a passenger in a car for an hour without a break	from 0 to 3
Item #5	Lying down in the afternoon when circumstances permit	from 0 to 3
Item #6	Sitting and talking to someone	from 0 to 3
Item #7	Sitting quietly after a lunch without alcohol	from 0 to 3
Item #8	In a car while stopped for a few minutes in traffic	from 0 to 3
Self-reported DS	“During the day, do you ever fall asleep unintentionally?”	Yes/No
History (self-reported by patients)		
Smoking status		current; ex-smoker; never
Physical examination by sleep technician according to the lab manual		
WGT	weight	kg
HGT	height	cm
PSG recording: PSG software used is different version of Sandman (current - 7.3)		
TST	Time in Bed (TIB) – Sleep Latency (SL)	total sleep time, hours
ArI, total	Total Arousals index in TST	events/hr.
AWK in TST	Total Awakenings, number in TST	events/TST
TiSaO ₂ <90%	Duration of SaO ₂ < 90% in TST	minutes
	% of SaO ₂ < 90%, in TST	%
Obstructive apnea/hypopnea event	Must fulfill criterion 1 or 2, plus criterion 3 of the following (5): 1. A clear decrease (>50%) from baseline in the amplitude of a valid measure of breathing during sleep. Baseline is defined as the mean amplitude of stable breathing and oxygenation in the two minutes preceding onset of the event (in individuals who have a stable breathing pattern during sleep) or the mean amplitude of the three largest breaths in the two minutes preceding onset of the event (in individuals without a stable breathing pattern). 2. A clear amplitude reduction of a validated measure of breathing during sleep that does not reach the above criterion but is associated with either an oxygen desaturation of >3% or an arousal. 3. The event lasts 10 seconds or longer	
	Obstructive Apnea-Hypopneas Index in TST	events/hr.
	Total Apnea-Hypopneas Index in TST	events/hr.

*All data from the polysomnograms performed at St. Michael's sleep laboratory are automatically exported into the database with no need for manual data entry.

eTable 2B: Description of health administrative datasets used.

Dataset	Description	Available from
RPDB	Registered Persons' Database files. Basic demographic information (e.g., sex, year of birth, date of death where applicable and postal codes) about anyone who has ever received an Ontario health card number.	1990
OHIP	Ontario Health Insurance Plan Physician Services Claims Database. Information on all physician claims in the province.	1991
CIHI-DAD /SDS	Canadian Institute for Health Information Discharge Abstract Database (information on all acute care hospitalization in Canada) and the Same Day Surgery.	1988
NACRS	Canadian Institute for Health Information National Ambulatory Care Reporting System: Emergency (ED), Dialysis and Cancer Clinics. Information on all ED visits in the province.	ED: 2002
Census	Contains aggregated data for Ontario and Canada that describe the general demographic information on 100% of the population and the remaining information for a 20% sample of the population. Ontario neighbourhoods are classified into one of the five approximately equal-sized income quintiles, ranked from poorest (Q1) to wealthiest (Q5). Each participant was assigned to the income quintile at the time of diagnostic sleep study using the patient's postal code and Statistics Canada's Postal Code Conversion File (6).	1991
ADP	Assistive Devices Program. Information on all individuals who have had their CPAP equipment paid for by the province.	2004
CHF	Ontario Congestive Heart Failure Database. It contains all Ontario individuals identified as having CHF based on OHIP/NACRS, CIHI, and RPDB data.	1991
COPD	Ontario Chronic Obstructive Pulmonary Disease Database. It contains all Ontario COPD patients based on OHIP, CIHI, and RPDB data.	1988
HYPERTENSION	Ontario Hypertension Database. It contains all Ontario hypertension patients based on OHIP, CIHI/SDS, and RPDB data.	1988
ODD	Ontario Diabetes Database. It contains all Ontario individuals identified as having diabetes based on OHIP, CIHI/SDS, and RPDB data. Once included, a person remains in the ODD until death or relocation outside of Ontario.	1991*
OMHRS	Ontario Mental Health Reporting System stand-alone admissions dataset. OMHRS is a data holding at CIHI that includes information on all adult inpatient mental health beds in Ontario. It is based on the Resident Assessment Instrument-Mental Health and includes information about mental and physical health, social support and service use.	2006
OCRD	Ontario Cancer Registry Data. It contains information on all Ontario residents who have been newly diagnosed with cancer or who have died of cancer, except non-melanoma skin cancer. Data is collected from: CIHI/DAD; Pathology Reports (paper); Pathology Data (PIMS); Registered Person Database (MOHLTC); Registrar General (Mortality Data); Chemo/Radiation Clinic visits (Integrated Cancer Programs & Princess Margaret Hospital); Data from Other Provincial Registries.	1985

*Data are available from 1988 to 1991 for hospitalized diabetes only (from CIHI).

eTable 2C: Categorization of cancers by cancer subtypes (N events = 627).

	ICD-9 Code	No. of events*
Anatomic subtype (most common)		
Prostate	185	125
Breast	174	75
Lung	162	61
Colorectal	153, 154	64
Kidney	189	34
Bladder	188	22
Etiologic type		
Potentially smoking related (7,8)†	140, 141, 143-145: lip, tongue, gum, floor of mouth, and other and unspecified parts of mouth 146-150: oropharynx, nasopharynx, hypopharynx, other and ill-defined sites within the lip, and esophagus 157: pancreas 160-163: nasal cavities, middle ear, accessory sinuses, larynx, trachea, bronchus, lung, and pleura 188- 189: bladder, kidney and other and unspecified urinary organs	156
Potentially alcohol related (9)	141: tongue 143-146: gum, floor of mouth, other and unspecified parts of mouth, and oropharynx 148-150: hypopharynx, other and ill-defined sites within the lip, and esophagus 155: liver and intrahepatic bile ducts 161: larynx	39
Potentially virus/immune related (9)	155: liver and intrahepatic bile ducts 173: skin, non-melanoma 180: cervix uteri 200-202: Non-Hodgkin and Hodgkin's lymphoma 204: leukaemia	59
Potentially hormone related (9)	174: breast 182: corpus uteri 183: ovary and other uterine adnexa 185: prostate	217
Detectable by screening		
Prostate, breast, colorectal, cervical	185, 174, 153, 154, 180	269

*Number of events of more than 100 was considered sufficient using the rule of thumb of 10 events per predictor.(10)

†Cancer of the respiratory tract (ICD-9 codes 160–163) : 41.7% of smoking-related cancers.

References

1. Rubin DB. Multiple imputation for nonresponse in surveys. New York ;: Wiley; 1987.
2. Harrell FE. Regression modeling strategies : With applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.
3. Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals *Biometrika* 1994;81:515 - 526.
4. Kendzerska T, Gershon AS, Hawker G, Leung RS, Tomlinson G. Obstructive sleep apnea and risk of cardiovascular events and all-cause mortality: A decade-long historical cohort study. *PLoS Med* 2014;11:e1001599.
5. Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. The report of an american academy of sleep medicine task force. *Sleep* 1999;22:667-689.
6. Hou F, Chen J. Neighbourhood low income, income inequality and health in toronto. *Health reports / Statistics Canada, Canadian Centre for Health Information = Rapports sur la sante / Statistique Canada, Centre canadien d'information sur la sante* 2003;14:21-34.
7. Smith GD, Shipley M, Leon DA. Height and mortality from cancer among men: Prospective observational study. *BMJ* 1998;317:1351-1352.
8. Macleod J, Smith GD, Heslop P, Metcalfe C, Carroll D, Hart C. Are the effects of psychosocial exposures attributable to confounding? Evidence from a prospective observational study on psychological stress and mortality. *Journal of epidemiology and community health* 2001;55:878-884.
9. Christensen AS, Clark A, Salo P, Nymann P, Lange P, Prescott E, Rod NH. Symptoms of sleep disordered breathing and risk of cancer: A prospective cohort study. *Sleep* 2013;36:1429-1435.
10. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. Ii. Accuracy and precision of regression estimates. *Journal of clinical epidemiology* 1995;48:1503-1510.