APPENDIX

**Supplementary Methodology**

**The CCHS-Nutrition 2004:** In 2004/2005, Statistics Canada conducted the cross-sectional, multistage, and complexCCHS- Nutrition 2004 collecting data from Canadians >0 years of age, excluding institutionalized populations, members of the Canadian Forces, those living in the territories and on reserves. In total, 35,107 respondents were sampled and appropriately weighted to represent >98% of the community-dwelling Canadian population (response rate = 76.5%). All respondents completed an initial 24-hour dietary recall with the assistance of a trained interviewer using a computerized modification of the United States Department of Agriculture’s Automated Multiple Pass Method. Seven to ten days later, approximately 30% of the respondents completed a second recall by phone. Trained dietitians coded all foods and disaggregated recipes and dishes into their main components for nutrient analysis. The nutrient and energy composition of consumed foods were obtained from the Canadian Nutrient File (supplement 2001b)[1]. Data on socio-demographic and lifestyle characteristics, and health conditions were collected via interviewer-administered questionnaires. To account for self-reported weights and heights, a correction factor was used to account for bias (e.g., under-reporting weight, over-reporting height) and produce more accurate estimates [2]. Additional details on CCHS-Nutrition 2004 survey design and data collection have been published previously[3].

**The Canadian Vital Statistics – Death Database (CVSD) (January, 2004 to December 31, 2011):** A detailed description of the linked cohort data and linkage methodology has been published[4]. All eligible CCHS-Nutrition 2004 respondents were first linked to the Historic Summary Tax File for personal identifiers using a probabilistic linkage approach. The resulting dataset was subsequently linked to the CVSD up to December 31, 2011 to ascertain mortality. Respondents were de-identified after the linkage was completed. Those still alive at the end of follow-up were censored.

**The Discharge Abstract Database (DAD) (January, 2004 to December 31, 2011):** The DAD includes discharges from all Canadian acute care hospitals from 1999/2000 to 2012 (for CCHS-Nutrition 2004), except for the province of Quebec (all years) and Manitoba (for fiscal year 1999/2000-03/04)[5]. The database includes over three million annual hospital discharges, originating from clinical chart abstractions conducted by uniformly trained personnel using standardized data definitions in every Canadian hospital[6, 7]. Individuals were identified through their unique health insurance number that provides access to the publicly funded universal health care system in their province or territory. Details on linkage methodology with CCHS-Nutrition 2004 have been previously published[8]. Participants were followed from interview date until the earliest of: loss to follow-up (i.e., loss of health care eligibility), end of study (December 31, 2011), incident CVD event, or death due to events other than CVD (i.e., competing risk).

**Institute of Medicine factorial equations for estimated energy requirements (EER) [9]:**

|  |  |
| --- | --- |
| **BMI between 18.5 kg/m2 and 25 kg/m2** | |
| **Ages 9-18** |  |
| Male | EER = 113.5 - 61.9\*age (years) + PAL \* (26.7 \* weight (kg) + 903 \* height (m)),  where PAL = 1 if sedentary, 1.13 if low active, 1.26 if active, and 1.42 if very active. |
| Female | EER = 160.3 - 30.8\*age (years) + PAL \* (10 \* weight (kg) + 934 \* height (m)),  where PAL = 1 if sedentary, 1.16 if low active, 1.31 if active, and 1.56 if very active. |
| **Ages 19 or older** |  |
| Male | EER = 661.8 - 9.53\*age (years) + PAL\*(15.91\* weight (kg) + 539.6\* height (m)),  where PAL = 1 if sedentary, 1.11 if low active, 1.25 if active, and 1.48 if very active. |
| Female | EER = 354.1 - 6.91\*age (years) + PAL\*(9.36\* weight (kg) + 726\* height (m)),  where PAL = 1 if sedentary, 1.12 if low active, 1.27 if active, and 1.45 if very active. |
| **BMI more than 25 kg/m2** | |
| **Ages 9-18** |  |
| Male | EER = -114.1-50.9\*age (years) + PAL \* (19.5\*weight (kg) + 1161.4\*height (m)),  where PAL = 1 if sedentary, 1.12 if low active, 1.24 if active, and 1.45 if very active. |
| Female | EER = 389.2 - 41.2\*age (years) + PAL \* (15 \* weight (kg) + 701.6 \* height (m)),  where PAL = 1 if sedentary, 1.18 if low active, 1.35 if active, and 1.60 if very active. |
| **Ages 19 or older** |  |
| Male | EER = 1085.6 - 10.08\*age (years) + PAL\*(13.7\* weight (kg) + 416\* height (m)),  where PAL = 1 if sedentary, 1.12 if low active, 1.29 if active and 1.59 if very active. |
| Female | EER = 447.6 - 7.95\*age (years) + PAL\*(11.4\* weight (kg) + 619\* height (m)),  where PAL = 1 if sedentary, 1.16 if low active, 1.27 if active and 1.44 if very active. |

Abbreviations: PAL = physical activity level.

**Equations used to determine physical activity level (PAL) [9]:**

ΔPAL = (MET - 1) \* Ntimes \* Minutes \* 1.34 / 1,440 (for men)

ΔPAL = (MET - 1) \* Ntimes \* Minutes \* 1.42 / 1,440 (for women)

Abbreviations: MET = metabolic equivalents; PAL = physical activity level.

**Directed Acyclic Graph of hypothesized model and relationship with covariates selected a priori:**

Diagram

Description automatically generated

Supplementary Table 1. Diagnostic codes for cardiovascular disease main events

|  |  |  |  |
| --- | --- | --- | --- |
| **Definition** | | **ICD-9\*** | **ICD-10** |
| Death (Vital Statistics) | |  |  |
|  | Ischemic Heart Disease Death | 410-414, 429.2 | I20 - I25 |
|  | Stroke Death | 430-434, 436-438 | I60 - I69 |
| Hospitalization (Main Diagnosis only) | |  |  |
| Acute Myocardial Infarction | | 410 | I21 |
|  | | 410 | I22 |
| Stroke | | 430 | I60 |
|  | | 431 | I61 |
|  | | 434 | I63 excluding I63.6 |
|  | | 436 | I64 |
|  | | 362.3 | H341 |

\*For deaths prior to 2003 and hospitalization discharges prior to 2002.

For validation studies of diagnostic codes see:

1. Tu JV, Chu A, Donovan LR, et al. The Cardiovascular Health in Ambulatory Care Research Team (CANHEART): using big data to measure and improve cardiovascular health and healthcare services. *Circ Cardiovasc Qual Outcomes.* 2015;8(2):204-212.
2. Kennedy CC, Brien SE, Tu JV. An overview of the methods and data used in the CCORT Canadian Cardiovascular Atlas project. *Can J Cardiol.* 2003;19(6):655-663.

**Supplementary Table 2.** Weighted and multivariable-adjusted hazard ratio (HR) and bootstrapped 95% confidence intervals (CIs) all-cause mortality and cardiovascular disease events (incidence and deaths) as per 1 unit increase in only the *first-day* 24-hour dietary recalls for sodium density, percentage of energy from added sugar, and percentage of energy from saturated fat at baseline to evaluate the impact of random measurement error, Canadian adults in Canadian Community Health Survey-Nutrition Linked to Canadian Vital Statistics – Death Database (CVSD) and Discharge Abstract Database (DAD), 2004/2005 to 2011\*

|  |  |  |  |
| --- | --- | --- | --- |
| **Nutrients** | | **Mid-value of 50th Percentile** | **HR (95% CI)** |
| Usual sodium density, mg/ 1000kcal, All-cause mortality HR (95% CIs) | |  |  |
|  | Model 1†,‡ | 1420.38 | 1.00 (1.00 -1.00) |
|  | Model 2§,‖ | 1429.82 | 1.00 (1.00 -1.00) |
| CVD events HR (95% CIs) † | |  |  |
|  | Model 1¶,‖ | 1408.09 | 1.00 (1.00 -1.00) |
|  | Model 2\*\*,†† | 1418.58 | 1.00 (1.00 -1.00) |
| Usual percentage of energy from added sugar, All-cause mortality HR (95% CIs) †\* | |  |  |
|  | Model 1 (λ=2.5)†,‡ | 171.70 | 1.00 (1.00 -1.00) |
|  | Model 2 (λ=2.5)§,‖ | 157.56 | 1.00 (1.00 -1.00) |
| CVD events HR (95% CIs) † | |  |  |
|  | Model 1 (λ=3)¶,‖ | 480.53 | 1.00 (1.00 -1.00) |
|  | Model 2\*\*,†† | 7.55 | 0.98 (0.98-0.98) |
| Usual percentage of energy from saturated fat, All-cause mortality HR (95% CIs) †\* | |  |  |
|  | Model 1†,‡ | 9.67 | 1.00 (0.97-1.04) |
|  | Model 2§,‖ | 9.66 | 1.00 (0.97-1.03) |
| CVD events HR (95% CIs) † | |  |  |
|  | Model 1¶,‖ | 9.54 | 0.99 (0.99-0.99) |
|  | Model 2\*\*,†† | 9.52 | 0.99 (0.99-0.99) |

\*Weighted and multivariable-adjusted HRs are calculated using regression calibration (Cox proportional hazards model) and the 95% CI calculated by bootstrapping the usual intake estimating models B=500 times at each step. λ=1 unless otherwise noted

† Unweighted n=13,473

‡**Model 1 (20≤age group):** Adjusted for baseline age (continuous) and sex in addition to education (less than secondary school graduation; secondary school graduation; some post-secondary; post-secondary graduation), smoking (daily and occasional smoker with 20≤ cigarettes per day; daily/occasional smoker with <20 cigarettes/day; former daily/occasional smoker and those who smoked ≥100 in lifetime; never smoked), misreporting (under-reporter, plausible-reporter and over-reporter), alcoholic beverage consumption (did not drink alcohol in the past 12 month; drank alcohol in the past 12 month for <once a month/once a month/2-3 times a month; drank alcohol in the past 12 month for once a week/2-3 times a week; drank alcohol in the past 12 months for 4-6 times a week/every day OR intense drinking: drank 2-3times a week/4-6 time a week/every day with the frequency of having 5 or more drinks being once/week/more than once a week), physical activity (daily energy expenditure≥3; 1.5≤daily energy expenditure<3; and 0≤daily energy expenditure<1.5), BMI categories (underweight, normal-weight, overweight, obesity), racial group (White, black, Korean/Chinese/Filipino/ Japanese South East Asian, West Asian/South Asian/Arab, Multiple ethnicity/Others), in addition to the Healthy Eating Index (HEI) 2010 score minus the nutrient (sodium, added sugar, or saturated fat depending on the model being evaluated)

§Unweighted n=14706

‖**Model 2 (20≤age group):** Model 1 covariates in addition to diabetes and cancer as covariates, only heart disease is removed in this dataset

¶Unweighted n=11546

\*\*Unweighted n=12643

††**Model 2 for CVD (20≤age group):** Model 1 covariates in addition to diabetes as covariate, only heart disease is removed in this dataset

**References**

1. Health Canada. The Canadian Nutrient File. Ottawa, ON, Canada: Nutrition Research Division; 2001.

2. Shields M, Connor Gorber S, Janssen I, et al. Bias in self-reported estimates of obesity in Canadian health surveys: an update on correction equations for adults. *Health Rep*. 2011;22(3):35-45.

3. Health Canada. Canadian Community Health Survey, Cycle 2.2, Nutrition (2004): A Guide to Accessing and Interpreting the Data. Available from: <https://www.canada.ca/en/health-canada/services/food-nutrition/food-nutrition-surveillance/health-nutrition-surveys/canadian-community-health-survey-cchs/canadian-community-health-survey-cycle-2-2-nutrition-2004-guide-accessing-interpreting-data-health-canada-2006.html> [Accessed: 01 Apr. 2019]

4. Sanmartin C, Decady Y, Trudeau R, et al. Linking the Canadian Community Health Survey and the Canadian Mortality Database: An enhanced data source for the study of mortality. *Health Rep*. 2016;27(12):10-8.

5. Canadian Institute for Health Information. Discharge Abstract Database (DAD) Metadata. Avaialble from: <https://www.cihi.ca/en/discharge-abstract-database-metadata> [Accessed: 01 Apr. 2019].

6. Garland A, Gershengorn HB, Marrie RA, et al. A Practical, Global Perspective on Using Administrative Data to Conduct Intensive Care Unit Research. *Ann Am Thorac Soc*. 2015;12(9):1373-86.

7. Canadian Institute for Health Information. CIHI Data Quality Study of the 2009-2010 Discharge Abstract Database. 2012. <https://secure.cihi.ca/free_products/Reabstraction_june19revised_09_10_en.pdf> [Accessed: 01 Apr. 2019].

8. Sanmartin C, Reicker A, Dasylva A, et al. Data resource profile: The Canadian Hospitalization and Taxation Database (C-HAT). *Int J Epidemiol*. 2018.

9. Garriguet D. Under-reporting of energy intake in the Canadian Community Health Survey. *Health Rep*. 2008;19(4):37.