

Design, methods and demographics from phase I of Alberta's Tomorrow Project cohort: a prospective cohort profile

Paula J. Robson PhD RNutr (UK), Nathan M. Solbak MSc, Tiffany R. Haig BA, Heather K. Whelan MSc, Jennifer E. Vena PhD, Alianu K. Akawung MSc, William K. Rosner MSc, Darren R. Brenner PhD, Linda S. Cook PhD, Ilona Csizmadi PhD, Karen A. Kopciuk PhD, S. Elizabeth McGregor PhD, Christine M. Friedenreich PhD

Abstract

Background: Prospective cohorts have the potential to support multifactorial, health-related research, particularly if they are drawn from the general population, incorporate active and passive follow-up and permission is obtained to allow access by researchers to data repositories. This paper describes Phase I of the Alberta's Tomorrow Project cohort, a broad-based research platform designed to support investigations into factors that influence cancer and chronic disease risk.

Methods: Adults aged 35–69 years living in Alberta, Canada, with no previous cancer diagnosis other than nonmelanoma skin cancer were recruited to the project by telephone-based random digit dialling. Participants were enrolled if they returned a Health and Lifestyle Questionnaire. Past year diet and physical activity questionnaires were mailed 3 months after enrolment. Consent was sought for active follow-up and linkage with administrative databases. Depending on enrolment date, participants were invited to complete up to 2 follow-up questionnaires (2004 and 2008).

Results: Between 2001 and 2009, 31 072 (39% men) participants (mean age 50.2 [\pm 9.2] yr) were enrolled and 99% consented to linkage with administrative databases. Participants reported a wide range of educational attainment and household income. Compared with provincial surveillance data from the Canadian Community Health Survey, Alberta's Tomorrow Project participants had higher body mass index, lower prevalence of smoking and similar distribution of chronic health conditions. Follow-up questionnaires were completed by 83% and 72% of participants in 2004 and 2008, respectively. Robust quality control measures resulted in low frequencies of missing data.

Interpretation: Alberta's Tomorrow Project provides a robust platform, based on a prospective cohort design, to support research into risk factors for cancer and chronic disease.

About 38 million deaths per year are attributable to chronic disease,¹ a number that is projected to increase to 52 million by 2030.² In Canada, cancer and chronic disease contributed to 88% of deaths between 2000 and 2012,¹ which resulted in increased recognition that prevention should be a research priority.^{3–5} Specifically, more research is needed to understand how potentially modifiable risk factors^{6–9} interact within complex, multilevel systems that include social, cultural, psychological, environmental and geographic variables.^{10,11}

Prospective cohorts have the potential to substantially enable such research,^{12,13} particularly if participants are recruited from the general population, provide a wide range of information about lifestyle and environment, are followed

actively over time to enrich and update information and provide consent for administrative database linkage. Although they are challenging to establish, longitudinal cohort-based research platforms facilitate a wide range of studies that provide insight into how lifestyle, behaviour, comorbidities and environmental exposures interact to impact long-term health.¹⁴

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Paula J. Robson, Paula.Robson@ahs.ca

CMAJ Open 2016. DOI:10.9778/cmajo.20160005

We previously reported a series of studies undertaken to explore the feasibility of establishing a cohort in Alberta, Canada;¹⁵ enrolment and follow-up of the cohort continued until 2009. In this paper, we provide a profile of Alberta's Tomorrow Project that aims to describe the following components of the project: the recruitment, enrolment, data collection and quality control methods used to create it between 2001 and 2009 (Phase I); participant characteristics at enrolment and comparison with national surveys; characteristics of participants who completed follow-up questionnaires compared with those who did not; and changes over time in participant characteristics from enrolment to follow-up.

Methods

Recruitment and enrolment

Studies exploring the feasibility of recruitment, enrolment and data collection are described in detail elsewhere.¹⁵ Participants were recruited through 8 waves of telephone-based random digit dialling using regional health authority boundaries within the province of Alberta as the sampling frame, and a 2-stage method to identify individuals. The first stage identified a household and the second selected 1 or 2 eligible adults, choosing the adult with the most recent birthday first. In wave 1, attempts were made to recruit a second household member; this practice was discontinued in subsequent waves because the assumption of random sampling in random digit dialling was not maintained.¹⁶

Using a standard script, the random digit dialling team explained the rationale for Alberta's Tomorrow Project and checked eligibility against 4 criteria: age between 35 and 69 years; no previous personal history of cancer, other than non-melanoma skin cancer; plans to reside in Alberta for at least 1 year; and ability to complete written questionnaires in English. Interested people were sent a cover letter, consent form, study information booklet, Health and Lifestyle Questionnaire and measuring tape (for anthropometric measurements). Participants were enrolled if they returned a completed Health and Lifestyle Questionnaire and signed consent form. Participants were also asked to provide their Personal Health Number for data linkage with the Alberta Cancer Registry and the provincial health ministry to facilitate future research on health service use and outcomes, including cancer diagnoses.

Ethical approval for recruitment and data collection was obtained from the former Alberta Cancer Board Research Ethics Committee and the University of Calgary Conjoint Health Research Ethics Board. Ethical approval for the current study was obtained from the Health Research Ethics Board of Alberta — Cancer Committee.

Data collection

The Health and Lifestyle Questionnaire captured information concerning personal and family health, lifestyle, anthropometrics and sociodemographic characteristics (Table 1). Nonrespondents were sent reminder postcards 6 and 16 weeks after the initial mailing and semiannual newsletters for 1 year, after which no further contact occurred. Twelve weeks after

returning a completed Health and Lifestyle Questionnaire, participants were mailed a Canadian Diet History Questionnaire I¹⁷⁻¹⁹ and Past Year Total Physical Activity Questionnaire.²⁰ These questionnaires had either been adapted (diet) or specifically developed and tested for reliability and validity (physical activity) for use in this cohort. Nonrespondents to the diet and activity questionnaires were sent a reminder postcard 6 weeks after the initial mailing and a replacement package after 16 weeks.

Enrolment continued between 2001 and 2009; depending on enrolment date, participants were sent follow-up questionnaires in 2004 and 2008. Participants enrolled from 2001 to 2003 were mailed Survey 2004, designed to update information on factors assessed at enrolment, as well as new exposures including sleep habits, hormone therapy, cancer risk perception, health-related quality of life²¹ and weight patterns in adulthood (Table 1). Nonrespondents received 2 reminders, and a shorter version of the questionnaire (Survey 2004 Express) was sent after 24 weeks in an attempt to obtain critical information from participants. Participants enrolled from 2001 to 2007 were mailed Survey 2008, designed to collect updated information pertaining to a number of factors assessed on the Health and Lifestyle Questionnaire, as well as information on consumption of fruits and vegetables, quality of life,²² physical activity,²³ lifetime shift work and environmental exposures, including the built environment²⁴ and types of occupations (employment for ≥ 6 mo). Current residence, lifetime residential history (street address, city or town, country, postal code, dates of occupancy of lifetime residences ≥ 1 yr), and birth places of participants, their parents and grandparents were also captured (Table 1). A reminder postcard and replacement package were sent to nonrespondents at 6 and 16 weeks, respectively.

Strategies to minimize loss to follow-up included a change of address form on the cohort website, collection of email addresses and alternative phone numbers, mailing of reminder postcards, contact information for secondary contacts (i.e., friends, family members) and rigorous follow-up procedures for returned mail and not-in-service telephone numbers. Additionally, biannual newsletters were mailed to all participants. Management of participant contact was facilitated by a custom designed software application created in C#.net 2.0 ("Cohort") that contained all participant contact information, records of questionnaire completion, records of study correspondence, and date of enrolment.

Data input and cleaning

Data entry, cleaning and verification was a multistep process. Questionnaires were first reviewed by 2 staff members to identify missing, ambiguous or contradictory information. After clarification with participants by telephone, Health and Lifestyle and Canadian Diet History Questionnaires were scanned using TELEform optical scanning software (Autonomy Company; Vista, California, USA; Versions 8.1–10.2). Blaise software (Statistics Netherlands, version 4.8; Heerlen, Netherlands) was used initially to enter the Past Year Total Physical Activity Questionnaire, but the process was transitioned to TELEform in 2007.

Table 1 (part 1 of 2) : Questionnaire domains captured in Alberta's Tomorrow Project Phase I enrolment and follow-up questionnaires

Measurements	Enrolment (2001–2009)	Follow-up (2004 and 2008)	
	Questionnaires* (n = 31 072)	Survey 2004† (n = 9 660)	Survey 2008 (n = 20 707)
Demographic information			
Current employment status	Yes	Yes	Yes
Occupational history			Yes
Shift work			Yes
Marital status	Yes	Yes	Yes
Education level	Yes		Yes
Annual household income	Yes		Yes
Ethnic background	Yes	Yes	Yes
Cancer and chronic disease			
Personal health history‡	Yes	Yes	Yes
Personal cancer diagnosis	Yes	Yes	Yes
Family structure (number of siblings, age of parents, cause of parental death)	Yes		
Family history of cancer	Yes		Yes
Family history of health conditions§	Yes		
Anthropometrics¶			
Height	Yes	Yes	Yes
Weight	Yes	Yes	Yes
Waist circumference	Yes	Yes	Yes
Hip circumference	Yes	Yes	Yes
Lifetime weight patterns		Yes	
Diet			
Food frequency questionnaire (including use of supplements)	Yes		
Past 7 day intake of fruit and vegetables			Yes
Physical activity**			
Employment/volunteer activities	Yes		Yes
Household activities	Yes		Yes
Recreation/leisure activities	Yes		Yes
Sedentary behaviours			Yes
Smoking and tobacco			
Current and past use of tobacco	Yes	Yes	Yes
Second-hand smoke exposure	Yes		
Quitting behaviours			Yes
Alcohol			
Alcohol use	Yes	Yes	
Beverage type and amount	Yes	Yes	
Sleep			
Sleep pattern		Yes	Yes

Table 1 (part 2 of 2) : Questionnaire domains captured in Alberta's Tomorrow Project Phase I enrolment and follow-up questionnaires

Measurements	Enrolment (2001–2009)	Follow-up (2004 and 2008)	
	Questionnaires* (n = 31 072)	Survey 2004† (n = 9 660)	Survey 2008 (n = 20 707)
Screening and risk behaviours			
Colon cancer screening (fecal occult blood test, colonoscopy, sigmoidoscopy)	Yes	Yes	Yes
Sun exposure — sunburn history	Yes	Yes	
Sun exposure — sunscreen use, tanning, risk of sunburn		Yes	
Primary care service use			Yes
Men's reproductive health			
Prostate specific antigen screening	Yes	Yes	Yes
Enlarged prostate	Yes		
Vasectomy	Yes		
Women's reproductive health			
Papanicolaou test screening	Yes	Yes	Yes
Mammogram screening	Yes	Yes	Yes
Breast exam	Yes		
Menstruation (age at onset)	Yes		
History of pregnancy and breastfeeding	Yes		
Oral contraceptive use	Yes		
Menopause (age, use of hormone replacement and alternative therapies)	Yes	Yes	
Oophrectomy or hysterectomy	Yes	Yes	Yes
Perceived health and quality of life			
General health rating	Yes	Yes	Yes
Stress and emotional state (anxiety, depression)	Yes	Yes	
Social support	Yes		
Spirituality	Yes		
Quality of life (mobility, self-care, pain)		Yes	Yes
Perception of risk for cancer and diabetes		Yes	
Built environment			
Built environment			Yes
Residential history			Yes
Postal code	Yes	Yes	Yes

*Enrolment measurements included the Health and Lifestyle Questionnaire (n = 31 072), the Canadian Diet History Questionnaire I (n = 26 843), and the Past Year Total Physical Activity Questionnaire (n = 26 769); 26 761 participants completed all 3 questionnaires (Health and Lifestyle Questionnaire, Canadian Diet History Questionnaire I and Past Year Total Physical Activity Questionnaire).

†An abbreviated version of Survey 2004 (Survey 2004 Express) was mailed to participants who did not return Survey 2004 to try to capture partial data on these participants. For the Express version, fewer questions were asked under each topic, but the questions that were asked were the same between questionnaires; 9197 participants completed Survey 2004, and 463 participants completed Survey 2004 Express.

‡Personal health history included self-reported physician diagnoses of chronic health conditions, including angina, chronic bronchitis, Crohn disease, cirrhosis of the liver, diabetes, emphysema, heart attack, hepatitis, high blood pressure, high cholesterol, polyps in colon and rectum, stroke and ulcerative colitis (all questionnaires), with the addition of arthritis, depression, high blood glucose, osteoporosis and thyroid problems in Survey 2004, and the addition of asthma, acid reflux, arthritis, heart problems, irritable bowel syndrome, osteoporosis and thyroid problems in Survey 2008.

§Family history of health conditions included diabetes, heart attack and stroke.

¶Anthropometric measurements are self-reported.

**Physical activity was measured over the previous year (Past Year Total Physical Activity Questionnaire) at enrolment and over the previous week (long-form International Physical Activity Questionnaire) at Survey 2008.

Several features were built into the TELEform programming to reduce data entry errors, including intelligent mark and character recognition and custom scripting to support review of complex multiquestion skip patterns. Issues were resolved according to rules established a priori or by telephone follow-up with the participant. A digital image of each page of the questionnaire was saved and, after cleaning, all data were transferred to an MS SQL server database.

The second quality control stage involved a graduated system of double data entry. Initially, data from each questionnaire were entered twice (100% double data entry), and discrepancies were checked against the hard copy of the questionnaire. Following a satisfactory 100% double data entry, the rate dropped to 20% double data entry and then to 10%, which was the minimum performed on all questionnaires. The final data entry quality check ensured that all questionnaires in the participant tracking database were associated with a data record, a TIFF image of the questionnaire and an entry in the inventory database.

Finally, each data set was checked using SAS programming (SAS Institute Inc., Cary, North Carolina). If information was contradictory or missing and the participant could not be contacted, the affected variables were coded as missing.

Statistical analysis

Means and standard deviations (SDs) were used to describe continuous variables, and frequencies and percentages were used for categorical variables. Responses from participants within the same household (random digit dialling wave 1) were assessed for agreement using intraclass correlation coefficients. Postal codes were mapped with the Statistics Canada postal code conversion file (March 2009 postal codes) to derive urban or rural status and determine geographic location at enrolment. Characteristics of Alberta's Tomorrow Project cohort and respondents from Alberta to the Canadian Community Health Survey Cycle 3.1²⁵ were compared using χ^2 tests. This cycle of Canadian Community Health Survey was selected because it was undertaken around the midpoint of Alberta's Tomorrow Project recruitment in 2005. Completion rates were calculated for all questionnaires, and the Pearson χ^2 test was used to assess proportional differences in completion rates across sociodemographic domains. All statistical tests were performed using SAS version 9.2 (SAS Institute) on a Linux interface.

Results

The random digit dialling process resulted in 63 486 people (42.1% men, 57.9% women) interested in receiving an enrolment package. Of these, 48.8% enrolled, and the Phase I cohort consists of 31 072 participants (age 50.2 ± 9.2 yr) (Table 2). Figure 1 shows the flow of participant enrolment and questionnaire completion. Participants were enrolled across Alberta (Figure 2), mostly in urban areas (Table 3), and were mostly of European descent (92.4%).

Almost all participants consented to linking their data with administrative databases ($n = 30\,658$) and provided a valid per-

sonal health number ($n = 30\,431$; Figure 1). Intraclass correlations between characteristics reported by participants recruited from the same household ($n = 382$) were evaluated. Age, annual household income, body mass index (BMI) and Asian ethnicity had intraclass correlations of at least 0.8. Educational and occupational status, all other ethnicities, smoking status and self-reported diagnosis of chronic conditions had intraclass correlation of less than 0.8. Future analytical decisions regarding inclusion or exclusion of "second in household" participants should be made on a case-by-case basis.

Alberta's Tomorrow Project participants were distributed evenly across age categories (Table 2). There was wide variation in total annual household income, and about half of the cohort reported completing some post-secondary education, whereas 10% had not completed high school. Similar proportions of men and women were obese. However, a greater proportion of men than women were overweight (48.4% v. 33.2%), and a greater proportion of women than men had a BMI in the normal range. Most of the participants reported being current nonsmokers, although 15.5% ($n = 4816$) reported being daily smokers. High cholesterol and high blood pressure were the most commonly reported chronic health conditions (Table 2).

Compared with the Canadian Community Health Survey Cycle 3.1 Alberta weighted data, a higher proportion of Alberta's Tomorrow Project participants had participated in some postsecondary education. However, Alberta's Tomorrow Project included a lower proportion of participants with a healthy BMI. Prevalence of daily smoking was also lower than in the Canadian Community Health Survey Alberta for men, although the distribution of smoking status was similar for women in both populations (Table 2).

Completion rates for the Canadian Diet History Questionnaire I and Past Year Total Physical Activity Questionnaire varied by sex, age, working status and smoking status (Table 3). All 3 questionnaires (Health and Lifestyle Questionnaire, Canadian Diet History Questionnaire I and Past Year Total Physical Activity Questionnaire) were completed by greater proportions of women, adults aged 55–69 years (v. adults aged 35–44 yr, retirees and former/never smokers (Table 3). The overall response rate for Survey 2004 or Survey 2004 Express was 83.1% (Figure 1), with the express version completed by less than 5% of respondents. Characteristics of respondents and nonrespondents to Survey 2004 showed the same patterns as observed for diet and physical activity questionnaire completion, whereby nonrespondents were younger, daily smokers and more likely than respondents to work full time. In contrast to Survey 2004, there were few differences in sociodemographic characteristics reported by respondents and nonrespondents to Survey 2008 (Table 3).

Changes in sociodemographic and health-related variables reported from enrolment to Survey 2008 were examined (Table 4). The frequencies of self-reported physician diagnosis of high blood pressure, high cholesterol and heart attack increased from enrolment to Survey 2008. The proportion of participants with a family history of cancer increased from 53.9% at enrolment to 60.4% at Survey 2008.

Table 2: Characteristics of Alberta's Tomorrow Project (ATP) participants at enrolment compared with respondents in Alberta to the 2005 Canadian Community Health Survey Cycle 3.1

Self-reported domains	Men		Women	
	ATP, % n = 12 116	CCHS Alberta*, % n = 3376	ATP, % n = 18 956	CCHS Alberta*, % n = 3915
Sex	39.0	49.8	61.0	50.2
Age, yr†				
35–44	32.0	32.8	32.8	30.9
45–54	35.5	30.8	35.1	29.8
55–64	24.2	18.6	23.7	18.2
65–69	8.3	17.9	8.4	21.1
Missing	0.0	0.0	0.0	0.0
Education†				
High school not completed	11.0	15.1	9.0	15.7
High school completed	14.9	14.5	20.8	18.1
Some postsecondary‡	18.7	5.7	22.3	6.4
Postsecondary completed§	55.5	62.1	47.9	57.3
Missing	0.0	2.7	0.0	2.5
Marital status†				
Married/living with a partner	82.5	81.1	75.6	70.9
Single (never married)	6.6	9.2	5.5	6.4
Divorced/separated/widowed	10.9	9.7	18.9	22.6
Missing	0.0	0.0	0.0	0.1
Annual household income, \$ †				
< 30 000	9.3	11.2	15.8	16.1
30 000–59 999	24.7	14.3	28.4	14.9
60 000–89 999	27.6	22.3	23.7	20.1
≥ 90 000	36.9	38.2	29.2	30.5
Missing	1.6	14.0	3.0	18.4
Body mass index¶†				
< 18.5	0.2	0.4	1.1	2.8
18.5–24.9	23.0	34.0	39.2	46.8
25.0–29.9	48.4	45.3	33.2	29.9
≥ 30.0	28.4	19.5	26.2	15.9
Missing	0.0	0.8	0.5	4.6
Smoking status†				
Daily smoker	16.0	21.3	15.2	17.1
Occasional smoker	3.5	3.9	2.9	3.5
Not at all**	80.5	74.4	81.9	79.3
Missing	0.1	0.3	0.0	0.1
Chronic health condition††				
High blood pressure‡‡	24.8	20.3	21.7	20.0
Emphysema	0.9	1.1	0.6	0.8
Chronic bronchitis‡‡	2.8	2.2	3.9	2.8
Diabetes	5.8	6.4	4.3	5.5
Ulcerative colitis§§	0.9	0.7	1.2	0.7
Crohn disease	0.5	0.2	0.7	0.3

Note: ATP = Alberta's Tomorrow Project, CCHS = Canadian Community Health Survey.

*CCHS 3.1 Alberta weighted data, restricted to CCHS 3.1 Albertan respondents aged 35–69 years as per Alberta's Tomorrow Project (ATP) inclusion criteria.

†p < 0.001 ATP enrolment v. CCHS 3.1 Alberta weighted data (χ^2 test).

‡Some postsecondary includes combined responses to: some technical school/college training completed, some part of university degree completed.

§Postsecondary completed includes combined responses to: completed technical school/college training, completed university degree, some part of post-graduate university degree completed, completed university postgraduate degree.

¶Body mass index derived from participant self-reported height and weight.

**"Not at all" smoking status includes never smokers and former smokers.

††Chronic health condition includes a self-reported physician diagnosis of one or more of the following: high blood pressure, emphysema, chronic bronchitis, diabetes, ulcerative colitis, Crohn disease.

‡‡p < 0.001 ATP enrolment v. CCHS 3.1 Alberta weighted data men only (χ^2 test).

§§p < 0.001 ATP enrolment v. CCHS 3.1 Alberta weighted data women only (χ^2 test).

Interpretation

Alberta's Tomorrow Project Phase I cohort is well placed to support a broad range of health-related research initiatives. Comprehensive data related to sociodemographic, environmental and lifestyle domains were obtained, and rigorous quality control procedures have resulted in well-documented databases with low frequencies of missing data. The

very high proportion of participants who consented to administrative database linkage will facilitate research on contextual factors that may be important in understanding how patterns of health services use and lifestyle behaviour are associated with health outcomes.^{14,26} A strength of the cohort is the ability to link current residence, and potentially historic residences, with environmental data, using

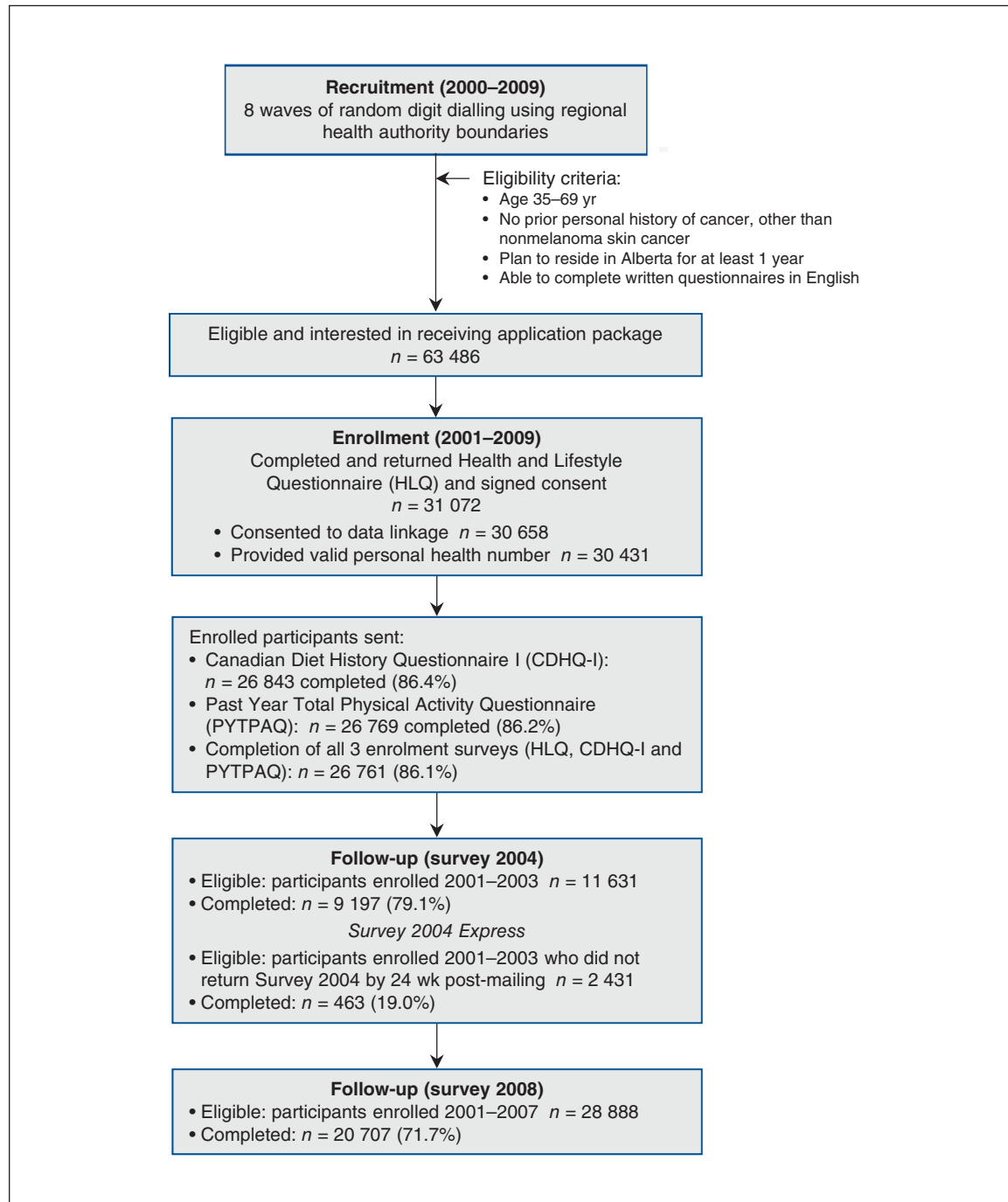


Figure 1: Enrolment and follow-up of participants in Alberta's Tomorrow Project. The follow-up questionnaire in 2008 was sent to participants enrolled between 2001 and 2007, meaning that some participants would have previously completed Survey 2004.

Geographic Information Systems technology based on either postal code or street address to map different types of exposures against health outcomes.²⁷ Furthermore, many variables are harmonized to facilitate pooling with other large cohorts,²⁸ which will support future studies that require larger sample sizes.²⁹

Alberta's Tomorrow Project used random digit dialling in an attempt to draw a large sample of the general adult population of Alberta. This approach is somewhat different from that used by many earlier North American cohorts that focused on specific occupations (e.g., nurses,³⁰ physicians,^{31,32} teachers³³), residents of smaller geographic units³⁴ or people listed on a registry that permitted random sampling.^{35,36} During the establishment of Alberta's Tomorrow Project, there were no population-based registries that could be used as a sampling frame, and other approaches based on existing surveillance methods were unfeasible;¹⁵ thus, random digit dialling was selected as the most viable option. Given recent trends suggesting a decline in use of fixed land-lines,³⁷ it is unclear if random digit dialling will remain viable for recruitment to large cohorts. Approaches using newer technologies may help reduce costs associated with recruitment and enrolment,³⁸ but should be evaluated rigorously before implementation.

One criticism levelled at cohorts is that they often comprise highly educated, health conscious participants with rela-

tively low prevalence of chronic disease.^{39–41} Although it is true that prospective cohorts drawn from the general population rarely, if ever, comprise a truly representative sample of their source population, the use of random digit dialling in Alberta's Tomorrow Project resulted in a cohort from all over the province who reported a wide range of sociodemographic and health-related characteristics. The proportions of men and women in Alberta's Tomorrow Project were unbalanced; however this situation is common for prospective cohorts. For example, the Genetic Epidemiology Research on Adult Health and Aging cohort (NIH accession number: phs000674.v1.p1)³⁶ comprises 42% men whereas the National Institutes of Health-American Association of Retired Persons Diet and Health Study³⁵ has 60% men. Other cohorts have been limited only to one sex.^{30,32,33} Despite not being representative of their source population, these cohorts have made valuable contributions to our understanding of the antecedents of cancer and chronic disease because they have used rigorous approaches for the collection of data obtained from the same people over time, addressed internal validity, and have established the capacity to link with administrative databases to analyze information on health outcomes.

With respect to other health-related characteristics, the Alberta's Tomorrow Project cohort appears broadly similar to its source population. Proportions of participants reporting conditions such as hypertension and diabetes are comparable to those reported in national surveillance data,²⁵ as well as those reported recently by Quebec's CARTaGENE study²⁶ and the UK Biobank,⁴² both of which drew samples from provincial and national health insurance databases. However, the Alberta's Tomorrow Project cohort seems to be somewhat different from the Alberta population in terms of body size. The prevalence of obesity observed in men (28.4%) and women (26.4%) was substantially higher than was reported by Canadian Community Health Survey cycle 3.1 (19.5% and 15.9%, respectively).²⁵ However, subsequent comparisons of measured and self-reported heights and weights in a subsample of Canadian Community Health Survey participants aged 18–79 years resulted in adjusted estimates of obesity that were substantially higher in men (self-report 16.7%; measured 26.2%) and women (self-report 16.0%; measured 23.0%), suggesting bias arising from the use of self-reported information.⁴³ Although Alberta's Tomorrow Project participants also self-reported anthropometric measures, the prevalence of obesity in the cohort was closer to prevalence estimates based on measured heights and weights in Canadians aged 35 years and older.⁴⁴ We have speculated that providing detailed instructions for measurements (including diagrams) and including a tape measure may have resulted in more accurate self-reporting than simply asking for height and weight. However, more work is required to better understand how question wording and mode of administration are likely to affect the quality of self-reported anthropometric data.

Our exploration of response rates for the first follow-up questionnaire showed that repeat respondents were more likely to be women, older, retired and nonsmokers. These

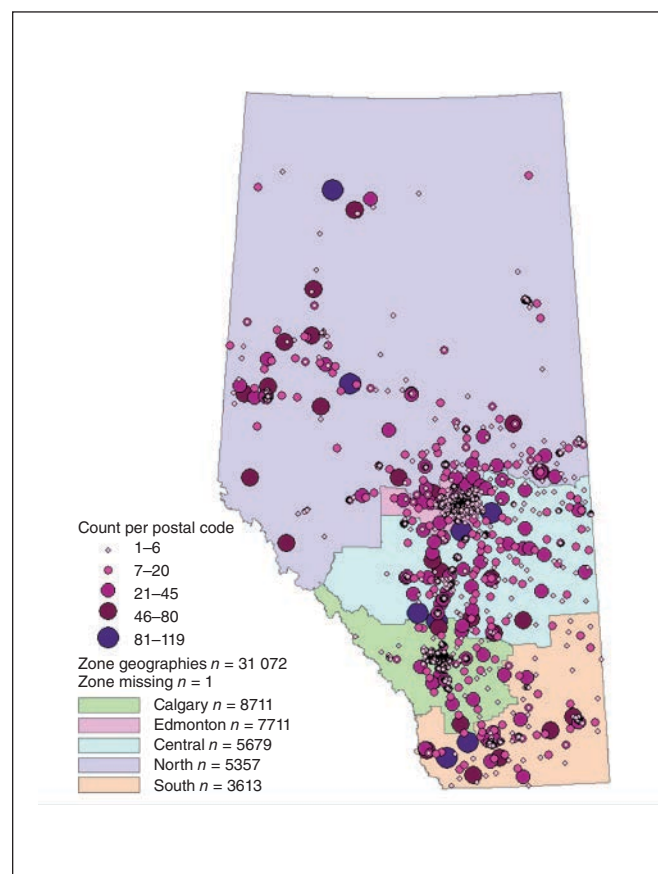


Figure 2: Geographic postal code region coverage, location of participants in Alberta's Tomorrow Project at enrolment, Alberta, Canada.

characteristics are typical of those commonly reported for participants thought to be more engaged in research.⁴⁵ Further work to elucidate patterns of response is required, taking into account health outcomes that may reduce the likelihood of questionnaire completion. Loss to follow-up is a major concern of longitudinal cohorts, and thus using retention strategies continues to be a priority for Alberta's Tomorrow Project. Although we have consent for passive follow-up through administrative database linkage, we plan to continue with active follow-up to enrich the databases and facilitate examinations of how changes over time in exposures or health status affect long-term health outcomes. To date, we have

maintained a follow-up response rate of between 72% and 83%; comparable with those reported by others such as the French E3N Cohort Study (80%)⁴⁶ and the Nord Trondelag Health Study (73%–80%).⁴¹ Nonetheless, we are increasingly aware of the need to explore and implement comprehensive strategies to engage participants effectively.

Limitations

In addition to the limitations mentioned above, recruitment for Alberta's Tomorrow Project Phase I was limited to residents with plans to reside in Alberta for at least 1 year and the ability to complete written questionnaires in English.

Table 3 (part 1 of 2): Enrolment characteristics reported by participants who completed follow-up questionnaires compared with those reported by nonrespondents to follow-up questionnaires

Response	Baseline		Survey 2004*		Survey 2008	
	Fully enrolled†	Partially enrolled‡	Returned	No response	Returned	No response
Total <i>n</i>	(<i>n</i> = 26 761), %	(<i>n</i> = 4311), %	(<i>n</i> = 9660), %	(<i>n</i> = 1971), %	(<i>n</i> = 20 707), %	(<i>n</i> = 8181), %
Sex§¶						
Men	37.7	47.3	39.8	48.7	38.9	39.6
Women	62.4	52.7	60.2	51.3	61.1	60.4
Age (yr)§¶						
35–44	31.1	41.3	35.1	44.4	32.3	32.3
45–54	35.3	34.9	34.7	36.1	35.5	35.3
55–64	24.8	18.5	22.0	15.1	24.0	24.1
65–69	8.8	5.3	8.2	4.4	8.2	8.4
Missing	0.0	0.0	0.0	0.0	0.0	0.0
Education§						
High school not completed	9.6	11.1	12.0	13.7	10.0	9.9
High school completed	18.5	18.2	20.2	21.2	18.5	18.7
Some post-secondary**	20.6	22.9	21.2	22.9	20.7	21.2
Post-secondary completed††	51.4	47.7	46.6	42.2	50.8	50.2
Missing	0.0	0.2	0.0	0.1	0.0	0.0
Marital status§ ¶						
Married/living with a partner	78.8	75.2	80.5	76.1	78.4	78.1
Single (never married)	5.8	6.8	5.2	6.9	6.0	5.4
Divorced/separated/widowed	15.5	17.7	14.3	17.0	15.6	16.4
Missing	0.0	0.2	0.0	0.1	0.0	0.0
Annual household income, \$§						
< 30 000	13.0	14.9	15.7	17.6	13.2	13.8
30 000–59 999	27.0	26.4	31.4	32.3	26.8	28.2
60 000–89 999	25.3	24.7	26.6	27.2	25.3	25.4
≥ 90 000	32.4	31.0	23.9	20.4	32.3	30.1
Missing	2.4	3.0	2.4	2.6	2.4	2.6
Geographic location‡‡						
Urban	76.5	75.2	70.7	71.6	76.2	76.0
Rural	23.5	24.8	29.3	28.4	23.8	24.0
Missing	0.0	0.0	0.0	0.0	0.0	0.0

Table 3 (part 2 of 2): Enrolment characteristics reported by participants who completed follow-up questionnaires compared with those reported by nonrespondents to follow-up questionnaires

Response	Baseline		Survey 2004*		Survey 2008	
	Fully enrolled†	Partially enrolled‡	Returned	No response	Returned	No response
Working status§¶						
Full time	55.8	64.2	54.5	63.8	57.0	57.0
Part time	16.8	13.3	17.1	12.6	16.6	15.8
Homemaker	1.9	3.0	2.0	2.9	1.9	2.2
Unemployed	8.3	6.8	9.4	7.8	8.1	8.2
Student	0.6	0.8	0.4	0.8	0.6	0.6
Retired	13.4	7.3	13.2	6.5	12.4	12.8
Other	3.2	4.5	3.3	5.5	3.4	3.4
Missing	0.0	0.1	0.1	0.1	0.0	0.1
Body mass index§¶§§						
< 18.5	0.7	0.9	0.6	0.8	0.7	0.6
18.5–24.9	33.3	30.1	33.0	29.3	32.7	33.0
25.0–29.9	39.2	37.1	40.0	38.0	39.3	38.3
≥ 30.0	26.5	30.0	26.2	30.7	26.9	27.4
Missing	0.3	1.8	0.3	1.3	0.4	0.7
Smoking status¶¶						
Current daily smoker	14.5	21.9	16.3	27.9	15.4	16.1
Current occasional smoker	3.1	3.4	3.4	4.0	3.1	3.3
Former smoker	37.6	34.3	37.6	33.2	37.1	37.8
Never smoker	44.8	40.0	42.7	35.0	44.3	42.8
Missing	0.1	0.4	0.0	0.1	0.1	0.1
Family history of cancer§¶¶¶						
Yes	53.2	48.6	51.2	46.2	53.1	51.4
No	46.8	51.5	48.8	53.8	46.9	48.6
Missing	0.0	0.0	0.0	0.0	0.0	0.0
Family history of chronic conditions***						
Diabetes	27.8	28.5	26.4	29.5	27.8	27.8
Heart attack§	22.1	19.8	23.0	20.5	21.8	22.0
Stroke	7.7	6.5	7.4	6.5	7.7	7.3

*Survey 2004 and Survey 2004 Express combined.
 †Fully enrolled = participants who completed Health and Lifestyle Questionnaire, Canadian Diet History Questionnaire I and Past Year Total Physical Activity Questionnaire.
 ‡Partially enrolled = participants who completed Health and Lifestyle Questionnaire only, and did not return the Canadian Diet History Questionnaire I or Past Year Total Physical Activity Questionnaire.
 § $p < 0.001$ fully enrolled vs. partially enrolled (χ^2 test).
 ¶ $p < 0.001$ Survey 2004 returned vs. Survey 2004 no response (χ^2 test).
 **Some post-secondary includes combined responses to some technical school/college training completed, completed technical school/college training, some part of university degree completed.
 ††Post-secondary completed includes combined responses to completed university degree, some part of post-graduate university degree completed, completed university postgraduate degree.
 †††Geographical location defined according to postal code, where “0” as the second digit indicates rural residence.
 §§Body mass index derived from participant self-reported height and weight.
 ¶¶¶Family history of cancer includes a self-reported cancer diagnosis in a first-degree relative (mother, father, brother, sister, son, daughter) of breast, ovarian, rectal, colon, prostate or other cancer.
 ***Family history of chronic conditions includes a self-reported diagnosis in a first-degree relative (mother, father, brother, sister, son, daughter) of diabetes, heart attack or stroke.

These criteria were established for practical reasons, but may have limited the representativeness of our sample. In random digit dialling wave 1, attempts were made to recruit a second household member. We found that members of the

same household shared many similar characteristics. Future analytical decisions regarding inclusion or exclusion of “second in household” participants should be made on a case-by-case basis, depending on the research.

Conclusion

Alberta's Tomorrow Project Phase I is a robust platform that will support a wide range of health-related research studies. Alberta's Tomorrow Project Phase I is currently being augmented by Phase II (the Canadian Partnership for Tomorrow Project protocol),⁴⁷ which includes the collection of additional

health and lifestyle data, measured anthropometric variables and collection of biological specimens. Both phases will result in the creation of rich repositories of data and specimens that may be accessed by researchers who have questions that will advance knowledge in the cause and early detection of cancer and chronic diseases.

Table 4: Characteristics reported by Alberta's Tomorrow Project participants from enrolment to Survey 2008 follow-up questionnaire

Characteristic	Men (n = 7788)		Women (n = 12 919)	
	Enrolment*, %	Survey 2008, %	Enrolment*, %	Survey 2008, %
†Body mass index‡				
< 18.5	0.2	0.2	1.0	0.9
18.5–24.9	22.9	21.8	40.4	38.2
25.0–29.9	50.3	48.9	33.6	33.2
≥ 30.0	26.3	28.0	24.7	25.7
Missing	0.2	1.1	0.3	2.1
§Smoking status‡				
Current daily smoker	13.4	11.5	12.8	10.9
Current occasional smoker	3.2	2.5	2.7	2.0
Former smoker	40.4	43.1	36.7	39.4
Never smoker	42.9	42.8	47.8	47.7
Missing	0.1	0.1	0.1	0.1
¶Annual household income, \$‡				
< 30 000	8.3	5.5	14.8	10.3
30 000–59 999	23.7	17.3	28.1	21.5
60 000–89 999	28.1	20.4	24.5	19.4
≥ 90 000	38.4	50.9	29.8	38.8
Missing	1.5	5.9	2.8	10.1
§Personal history of chronic conditions‡				
High blood pressure	24.7	33.9	21.6	29.5
Emphysema	1.0	1.4	0.6	1.0
Chronic bronchitis	2.7	4.1	3.6	5.5
Diabetes	5.4	8.5	3.9	5.9
Ulcerative colitis	0.9	1.3	1.1	1.4
Crohn disease	0.5	0.6	0.7	0.8
Angina	4.0	5.7	1.8	2.8
High cholesterol	31.5	41.9	24.5	33.5
Heart attack	2.8	10.4	0.8	6.6
Stroke	0.7	1.4	0.8	1.4
Hepatitis	3.2	3.9	3.0	3.8
Cirrhosis	0.2	0.4	0.2	0.3
¶Family history of cancer‡¶				
Yes	51.4	58.3	55.4	61.7
No	48.6	41.7	44.6	38.3
Missing	0.0	0.0	0.0	0.0

*Enrolment data presented only for the participants who completed Survey 2008 (total n = 20 707).

†Body mass index was derived from participant self-reported height and weight.

‡p < 0.001 Fully enrolled v. partially enrolled (Pearson χ^2 test).

§Personal history of chronic conditions: a self-reported physician diagnosis of one or more of the following: High blood pressure, emphysema, chronic bronchitis, diabetes, ulcerative colitis, Crohn disease, angina, high cholesterol, heart attack, stroke, hepatitis, cirrhosis of the liver.

¶A self-reported cancer diagnosis in a first-degree relative (mother, father, brother, sister, son, daughter) of breast, ovarian, rectal, colon, prostate or other cancer.

References

1. *Noncommunicable diseases country profiles 2014*. Geneva: World Health Organization; 2014.
2. *Global status report on noncommunicable diseases 2014*. Geneva: World Health Organization; 2014.
3. Bauer UE, Briss PA, Goodman RA, et al. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *Lancet* 2014;384:45-52.
4. Hanusaik N, Contandriopoulos D, Kishchuk N, et al. Chronicling changes to the chronic disease prevention landscape in Canada's public health system 2004-2010. *Public Health* 2014;128:716-24.
5. Ashbury FD, Little J, Ioannidis JPA, et al. A vision for chronic disease prevention and intervention research: report from a workshop. *Can J Public Health* 2014;105:e150-3.
6. Meschia JF, Bushnell C, Boden-Albala B, et al. Guidelines for the primary prevention of stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2014;45:3754-832.
7. Dartois L, Fagherazzi G, Boutron-Ruault M-C, et al. Association between five lifestyle habits and cancer risk: results from the E3N cohort. *Cancer Prev Res (Phila)* 2014;7:516-25.
8. *Breast cancer facts & figures 2013-14*. Atlanta: American Cancer Society; 2013.
9. *Food, nutrition, physical activity, and the prevention of cancer: a global perspective*. Washington (DC): World Cancer Research Fund, American Institute for Cancer Research; 2007.
10. Hystad P, Brauer M, Demers PA, et al. Geographic variation in radon and associated lung cancer risk in Canada. *Can J Public Health* 2014;105:e4-10.
11. Colditz GA, Wei EK, Edu C, et al. Preventability of cancer: the relative contributions of biologic and social and physical environmental determinants of cancer mortality. *Annu Rev Public Health* 2012;33:137-56.
12. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. *Nat Rev Genet* 2006;7:812-20.
13. Scholtens S, Smidt N, Swertz MA, et al. Cohort profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol* 2015;44:1172-80.
14. Doiron D, Raina P, Fortier I. Linking Canadian population health data: maximizing the potential of cohort and administrative data. *Can J Public Health* 2013;104:e258-61.
15. Bryant H, Robson PJ, Ullman R, et al. Population-based cohort development in Alberta, Canada: A feasibility study. *Chronic Dis Can* 2006;27:51-9.
16. Potthoff RF. Telephone sampling in epidemiologic research: To reap the benefits, avoid the pitfalls. *Am J Epidemiol* 1994;139:967-978.
17. Diet History Questionnaire. Washington (DC): National Cancer Institute; 2007. Available: <http://epi.grants.cancer.gov/DHQ/> (accessed 2015 Jan 15).
18. Cszimadi I, Kahle L, Ullman R, et al. Adaptation and evaluation of the National Cancer Institute's Diet History Questionnaire and nutrient database for Canadian populations. *Public Health Nutr* 2007;10:88-96.
19. Cszimadi I, Boucher B, Lo Siou G, et al. Using national dietary intake data to evaluate and adapt the US Diet History Questionnaire: the stepwise tailoring of an FFQ for Canadian populations. *Public Health Nutr* 2016;28:1-9.
20. Friedenreich CM, Courneya KS, Neilson HK, et al. Reliability and validity of the Past Year Total Physical Activity Questionnaire. *Am J Epidemiol* 2006;163:959-70.
21. Ware JE, Sherbourne CD, Davies AR. Developing and testing the MOS 20-item short-form health survey. A general population application. In: Stewart AL, Ware JE, editors. *Measuring functioning and well-being: the Medical Outcomes Study approach*. Duke University Press: Durham (NC); 1992:277-90.
22. EuroQol Group. EuroQol[®] — a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-206.
23. Craig CL, Marshall A, Sjostrom M, et al. International physical activity questionnaire: 12-Country reliability and validity. *Med Sci Sports Exerc* 2003;35:1381-95.
24. Cerin E, Saelens B, Sallis J, et al. Neighborhood Environment Walkability Scale: validity and development of a short form. *Med Sci Sports Exerc* 2006;38:1682-91.
25. Canadian Community Health Survey: Detailed information for 2005 (cycle 3.1). Ottawa: Statistics Canada; 2015. Available: www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=22642 (accessed 2015 Jan. 15).
26. Awadalla P, Boileau C, Payette Y, et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics on behalf of the CARTaGENE Project. *Int J Epidemiol* 2013;42:1285-99.
27. Edwards SE, Strauss B, Lynn Miranda M. Geocoding large population-level administrative datasets at highly resolved spatial scales. *Trans GIS* 2014;18:586-603.
28. Fortier I, Burton PR, Robson PJ, et al. Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. *Int J Epidemiol* 2010;39:1383-93.
29. Burton PR, Hansell AL, Fortier I, et al. Size matters: Just how big is BIG? Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol* 2009;38:263-73.
30. Belanger CF, Hennekens CH, Rosner B, et al. The Nurses' Health Study. *Am J Nurs* 1978;78:1039-40.
31. Hennekens CH, Eberlein K. A randomized trial of aspirin and B-carotene among US physicians. *Prev Med* 1985;14:165-8.
32. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking: a second report on the mortality of British doctors. *BMJ* 1956;2:1071-81.
33. Bernstein L, Allen M, Anton-Culver H, et al. High breast cancer incidence rates among California teachers: results from the California Teachers Study (United States). *Cancer Causes Control* 2002;13:625-35.
34. Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: The Framingham Study. *Am J Public Health Nations Health* 1951;41:279-81.
35. Schatzkin A, Subar AF, Thompson FE, et al. Design and serendipity in establishing a large cohort with wide dietary intake distributions. The National Institutes of Health-American Association of Retired Persons (NIH-AARP) diet and health study. *Am J Epidemiol* 2001;154:1119-25.
36. dbGaP Study Accession: phs000674.v1.p1. Resource for Genetic Epidemiology Research on Adult Health and Aging (GERA). Bethesda (MD): National Center for Biotechnology Information. Available: www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000674.v1.p1 (accessed 2015 Nov. 26).
37. *Residential Telephone Service Survey*. Ottawa: Statistics Canada; 2010. Available: www.statcan.gc.ca/daily-quotidien/110405/dq110405a-eng.htm (accessed 2015 Dec. 8).
38. Toledano MB, Smith RB, Brook JP, et al. How to establish and follow up a large prospective cohort study in the 21st century — lessons from UK COSMOS. *PLoS One* 2015;10:e0131521.
39. Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol* 2007;17:643-53.
40. Banks E, Herbert N, Mather T, et al. Characteristics of Australian cohort study participants who do and do not take up an additional invitation to join a long-term biobank: the 45 and Up Study. *BMC Res Notes* 2012;5:655.
41. Langhammer A, Krokstad S, Romundstad P, et al. The HUNT study: participation is associated with survival and depends on socioeconomic status, diseases and symptoms. *BMC Med Res Methodol* 2012;12:143.
42. Dare S, Mackay DF, Pell JP. Relationship between smoking and obesity: a cross-sectional study of 499,504 middle-aged adults in the UK general population. *PLoS One* 2015;10:e0123579.
43. Shields M, Connor-Gorbor 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;22:35-45.
44. *Table 105-0407 — Measured adult body mass index (BMI), by age group and sex, household population aged 18 and over excluding pregnant females*. In: Canadian Community Health Survey (CCHS 3.1), Canada, every 2 years. CANSIM. Ottawa: Statistics Canada; 2009.
45. Littman AJ, Boyko EJ, Jacobson IG, et al. Assessing nonresponse bias at follow-up in a large prospective cohort of relatively young and mobile military service members the Millennium Cohort Study. *BMC Med Res Methodol* 2010;10:99.
46. Clavel-Chapelon F. Cohort profile: the French E3N Cohort Study. *Int J Epidemiol* 2015;44:801-9.
47. Borugian MJ, Robson P, Fortier I, et al. The Canadian Partnership for Tomorrow Project: building a pan-Canadian research platform for disease prevention. *CMAJ* 2010;182:1197-201.

Affiliations: Cancer Measurement, Outcomes, Research and Evaluation (Robson), CancerControl Alberta, Alberta Health Services; Department of Agricultural, Food and Nutritional Science (Robson), Faculty of Agricultural, Life and Environmental Sciences, University of Alberta, Edmonton, Alta.; Cancer Measurement, Outcomes, Research and Evaluation (Solbak, Haig, Whelan, Vena, Akawung, Rosner), CancerControl Alberta, Alberta Health Services; Department of Cancer Epidemiology and Prevention Research (Brenner, Friedenreich, Cszimadi, Kopciuk), CancerControl Alberta, Alberta Health Services; Department of Oncology and Department of Community Health Sciences (Brenner, Friedenreich, Cszimadi), Cumming School of Medicine, University of Calgary, Health Sciences Centre, Calgary, Alta.; Department of Internal Medicine (Cook), University of New Mexico, Albuquerque, NM; Departments of Oncology and Mathematics and Statistics (Kopciuk), University of Calgary, Calgary, Alta.; Population, Public and Aboriginal Health (McGregor), Alberta Health Services, Calgary, Alta.

Data sharing: A list of previous publications, communications and other information can be found at www.myATP.ca. Standard operating procedures for data collection, processing and storage protocols are available upon request. Questionnaire data will be available to external researchers upon successful completion and approval of an access request. Access requests from national and international researchers are welcome. Information and details on how to request access to Alberta's Tomorrow Project data can be found at www.myATP.ca, or by emailing ATP.Research@ahs.ca.

Funding: Alberta's Tomorrow Project was made possible because of the commitment of its research participants, its staff and its funders: Alberta Cancer Foundation, Canadian Partnership Against Cancer, Alberta Cancer

Prevention Legacy Fund (administered by Alberta Innovates — Health Solutions) and substantial in-kind funding from Alberta Health Services. Dr. Friedenreich was supported by career awards from the Canadian Institutes of Health Research, Alberta Heritage Foundation for Medical Research, Alberta Innovates Health Solutions and the Alberta Cancer Foundation. Dr. Brenner is supported by a Canadian Cancer Society Research Institute Capacity Development in Cancer Prevention Award. The authors of this manuscript gratefully acknowledge the contributions made by previous researchers to the conceptualization, feasibility testing and overall design of Alberta's Tomorrow Project.

Disclaimer: The views expressed herein represent the views of the authors and not of Alberta's Tomorrow Project or any of its funders. This analysis is based on the Statistics Canada Canadian Community Health Survey, Cycle 3.1, Public Use Microdata File (catalogue no. 82M0013XCD) (2005). All computations, use and interpretation of these data are entirely those of the authors of this paper.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/4/3/E515/suppl/DC1