Cohort Profile: Design, Methods, and Demographics from Phase I of Alberta's Tomorrow Project Cohort

Paula J. Robson PhD RNutr(UK)*[1,2], Nathan M. Solbak MSc[3], Tiffany R. Haig BA[3], Heather K. Whelan MSc[3], Jennifer E. Vena PhD[3], Alianu K. Akawung MSc[3], William K. Rosner MSc[3], Darren R. Brenner PhD[4,5], Christine M. Friedenreich PhD[4,5]

[1]Cancer Measurement, Outcomes, Research and Evaluation, CancerControl Alberta, Alberta Health Services, Sun Life Place, 15th floor, 10123 99 Ave, Edmonton, AB, Canada, T5J 3C6

[2]Department of Agricultural, Food and Nutritional Science, Faculty of Agricultural, Life and Environmental Sciences, University of Alberta, 410 Agriculture/Forestry Centre, Edmonton, AB, Canada, T6G 2P5

[3]Cancer Measurement, Outcomes, Research and Evaluation, CancerControl Alberta, Alberta Health Services, 1820 Richmond Rd SW, Calgary, AB, Canada, T2T 5C7

[4]Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, 2210 2 St SW, Calgary, AB, Canada, T2S 3C3

[5]Department of Oncology and Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Health Sciences Centre, Foothills Campus, 3330 Hospital Dr NW, Calgary, AB, Canada, T2N 4N1

*To whom correspondence should be sent:
Paula J. Robson
Scientific Director, Alberta's Tomorrow Project
CancerControl Alberta, Alberta Health Services
Sun Life Place, 15th floor
10123 99 Ave
Edmonton, AB, Canada, T5J 3C6
Canada
Paula.Robson@ahs.ca

ABBREVIATIONS

ATP = Alberta's Tomorrow Project

BMI = Body Mass Index

CCHS = Canadian Community Health Survey

CDHQ = Canadian Diet History Questionnaire

DDE = Double Data Entry

DHQ = Diet History Questionnaire

FFQ = Food Frequency Questionnaire

HLQ = Health and Lifestyle Questionnaire

ICC = Intra-class Correlation

PHN = Personal Health Number

PYTPAQ = Past-Year Total Physical Activity Questionnaire

RDD = Random Digit Dialing

RHA = Regional Health Authority

SD = Standard Deviation

ABSTRACT

Background

Prospective cohorts have potential to support research into many factors that influence cancer and chronic disease risk, particularly if they are sampled from the general population, consent is obtained for active and passive follow-up, and permission is obtained to allow access by researchers to data repositories. This paper describes the profile of such a cohort in Alberta, Canada.

Methods

Albertans aged 35-69 years, with no previous cancer diagnosis other than non-melanoma skin, were recruited to Alberta's Tomorrow Project (ATP) by random digit dialing. Participants were enrolled if they returned a Health and Lifestyle Questionnaire and signed consent form. Past year diet and physical activity questionnaires were mailed three months following enrollment. Consent was sought for linkage with administrative databases, and active follow-up. Depending on enrollment date, participants were invited to complete up to two follow-up questionnaires (2004 and 2008).

Results

Between 2001 and 2009, 31,072 (39% men) Albertans (mean (SD) age 50.2 (9.2)) were enrolled and 99% consented to linkage with administrative databases. Participants reported a wide range of educational attainment and household income. Compared to provincial surveillance data, ATP participants had higher body mass index, lower prevalence of daily 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

ATP provides a robust platform, based on a prospective cohort design, to support research in risk factors for cancer and chronic disease.

INTRODUCTION

The World Health Organization estimates that 38 million deaths per year worldwide are attributable to chronic disease,[1] and current projections suggest this number will increase to 52 million by 2030.[2] In Canada alone, chronic disease, including cancer, contributed to 88% of all deaths between 2000 and 2012[1] leading to a growing recognition that research focused on chronic disease prevention should become a public health priority.[3–5] While several modifiable risk factors have been identified for specific cancers and other chronic diseases,[6–9] more research is needed to understand how these risk factors interact within complex, multi-level systems which also include social, cultural, psychological, environmental, and geographic variables.[10,11]

Prospective cohorts have the potential to be significant enablers of such research,[12,13] particularly if participants: (i) are recruited from the general population, (ii) are asked to provide a wide range of information pertaining to behaviour and environment, (iii) are followed actively over time to update and enrich information, (iv) provide consent for linkage with administrative databases, and (v) provide consent for the resulting data to be made available for analysis by the health research community. While these longitudinal cohort-based research platforms can be challenging to establish, they can facilitate a wide range of studies that may provide more insight into how lifestyle, behaviour, co-morbidities, and environmental exposures interact to impact long-term health.[14]

We previously reported on a series of studies that were undertaken to explore the feasibility of establishing such a cohort in Alberta, Canada.[15] Following the success of that earlier work, enrollment and follow-up to the Alberta cohort continued until 2009. In this paper, we now provide a profile of Alberta's Tomorrow Project (ATP). The objectives are to:

(i) Describe the recruitment, enrollment, data collection and quality control methods used to create the ATP between 2001 and 2009 (Phase I);

(ii) Describe characteristics reported by ATP participants at enrollment and compare with characteristics of Albertans reported in national surveys;

(iii) Explore characteristics of ATP participants who completed follow-up questionnaires compared to those who did not;

(iv) Describe changes over time in characteristics reported by ATP participants at enrollment and at follow-up.

METHODS

Recruitment, Enrollment and Consent

Studies exploring the feasibility of strategies for recruitment, enrollment and data collection in ATP are described in detail elsewhere.[15] The ATP cohort was recruited using eight waves of telephone-based random digit dialing (RDD). Regional health authority (RHA) boundaries were used as the sampling frame, and a two-stage method was used to identify eligible individuals. The first stage identified a household, while the second stage selected one or two eligible adults within each household, choosing the adult with the most recent birthday first, if eligible. In the first recruitment wave, a second household member of the opposite gender was recruited if possible. This practice was discontinued in subsequent recruitment waves because of concerns that issues could arise in future data analyses from high correlation between individuals living within the same household.

Using a standard script, the RDD team explained the rationale for ATP, and checked eligibility against four criteria: (i) aged 35-69 years; (ii) no prior personal history of cancer, other than non-melanoma skin cancer; (iii) plans to reside in Alberta for at least one year; and (iv) able to complete written questionnaires in English. Those who indicated that they were eligible and interested in receiving an ATP enrollment package were sent a cover letter, consent form, study information booklet, Health and Lifestyle Questionnaire (HLQ), and measuring tape to assist with anthropometric measurements. Participants were considered enrolled if they returned a completed HLQ and signed consent form.

In addition to completing self-administered questionnaires at enrollment and follow-up, participants were asked to provide consent and their Personal Health Number (PHN) for data linkage with the Alberta Cancer Registry and the provincial health ministry in order to facilitate future research on health services utilization and health outcomes, including cancer diagnoses. Participants were also asked if they would be willing to receive invitations to provide a biospecimen in the future.

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

Information about lifestyle-related risk factors and exposures was collected from participants at enrollment using self-administered questionnaires.[15] The HLQ captured information related to personal and family health history, reproductive history, smoking habits, cancer screening behaviours, sun exposure, psychosocial factors, anthropometric variables (height, weight, waist circumference, hip circumference) and socio-demographic characteristics (Table 1). Non-respondents were sent reminder postcards at six and 16 weeks following the initial mailing, and were also mailed bi-annual newsletters for one year, after which no further contact occurred. Approximately 12 weeks after returning a completed HLQ and consent form, participants were mailed a second package containing a cover letter, Canadian Diet History Questionnaire (CDHQ),[16,17] and Past Year Total Physical Activity Questionnaire (PYTPAQ).[18] These two self-administered questionnaires had been either adapted for use in the ATP (CHDQ) or were specifically developed and tested for reliability and validity for use in this cohort (PYTPAQ). Non-respondents to the CDHQ and PYTPAQ were sent a reminder postcard 6 weeks post initial mailing and a replacement package at 16 weeks.

Figure 1 illustrates the flow of participant enrollment and questionnaire completion for each phase of ATP data collection. Enrollment continued between 2001 and 2009, and, depending on when they were enrolled, participants were sent follow-up questionnaires in 2004 and 2008. Individuals enrolled from 2001 to 2003 (n=11,631) were mailed Survey 2004, which was designed to update information, including anthropometric measures, cancer screening behaviours, smoking habits, health status, and sun exposure. Additionally, a number of new exposures were assessed including sleep habits, hormone therapy, cancer risk perception, health-related quality of life[22] and weight patterns in adulthood (Table 1). Non-respondents were sent a reminder post card at six weeks and a replacement package at 16 weeks following the original mail date. A second strategy, used only for Survey 2004, was to send a shorter version of the questionnaire (Survey 2004 Express) at 24 weeks, in an attempt to obtain critical information from participants who may have been deterred by the more comprehensive version. Individuals enrolled from 2001 to 2007 (n=28,888) were mailed Survey 2008, which was designed to collect updated information pertaining to personal and family health history, cancer screening tests, tobacco use and quitting status, consumption of fruits and vegetables, quality of life,[23] physical activity,[24] lifetime history of shift work, and anthropometric measures. In addition, there was a

focus on environmental exposures, capturing information on the built environment[25] and types of occupations in which participants were employed for six months or more. Information on current residence, and lifetime residential history which included street address, city/town, country, postal code, and dates of occupancy of all previous lifetime residences greater than one year, and birth place of participants, their parents, and grandparents was also captured (Table 1). A reminder post card was sent to non-respondent participants after six weeks and a replacement package sent at 16 weeks. Individuals enrolled between 2001 and 2003 were sent both Survey 2004 and Survey 2008.

Several strategies were used to minimize loss to follow-up including: a change of address form on the ATP website, collection of email addresses and alternative phone numbers, reminder post cards, collection of contact information for "secondary" contacts (i.e. friends, family members etc.) and rigorous follow-up procedures for returned mail and not-in-service telephone numbers. Additionally, bi-annual newsletters were mailed to all participants recruited to ATP. 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 enrollment.

Data Input, Cleaning and Analysis

Data entry, cleaning and verification was a multi-step process. The first stage involved manual review of the questionnaires by two individuals to identify missing, ambiguous or contradictory information. Following review and clarification with participants by phone, HLQs and CDHQs were scanned using TELE*form*® optical scanning software (Autonomy Company; Vista. California, USA; Versions 8.1 to 10.2). Blaise® software (Statistics Netherlands, version 4.8; Heerlen, Netherlands) was used initially to enter the PYTPAQ, but was transitioned to TELE*form*® in 2007. Several features were built into the TELE*form*® programming to reduce data entry errors, including intelligent mark and character recognition, checking data entries to ensure only valid marks and characters were captured and custom scripting to provide further error checking and verification of complex multi question skip patterns. Issues were resolved according to rules established *a priori*, or telephone follow-up with the participant. As each questionnaire was completed, a digital image of each page of the questionnaire was saved and

the data were exported to an MS Access® database for storage during data cleaning. Following cleaning, data were transferred to an MS SQL® server database.

The second quality control stage involved a graduated system of double data entry (DDE), where staff progressed through steps depending on quality of standards attained. At the start of data entry, either for a new questionnaire or a new staff member, each questionnaire was entered twice by different staff members (100% DDE). SAS software (SAS Institute Inc.; Cary, North Carolina, USA) was used to compare the initial data entry to the second entry, and discrepancies were checked against the hard copy of the questionnaire. Decisions about progression to the next step were made by the Research Manager, and were based on the results for each individual data entry clerk. Following a satisfactory 100% DDE, the rate dropped to 20% DDE and then to 10% DDE. At least 10% DDE was performed on all questionnaires. Errors were reviewed at regular meetings, and procedures were revised if necessary to reduce the frequency of common errors. The final data entry quality check ensured that questionnaires logged as received 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 cleaned using SAS programming (SAS Institute Inc.; Cary, North Carolina, USA) based on the rules for the initial manual review for that questionnaire. If information was contradictory or missing, and the participant could not be contacted, the affected variables were entered as missing data.

Statistical Analysis

Means and standard deviations (SD) were used to summarize continuous variables, while categorical variables were described using frequencies and percentages. Responses provided by participants within the same household were assessed for multi-co-linearity using intra-class correlation coefficients (ICCs). Postal codes were mapped with the Statistics Canada postal code conversion file (March 2009 postal codes) to derive urban/rural status and to determine geographic location for each participant at enrollment. Chi-square tests were used in the comparative analyses between the Canadian Community Health Survey (CCHS) Cycle 3.1[26] and the ATP cohort. Albertan respondents to CCHS Cycle 3.1 were selected for comparison as these data were collected in 2005 at the midpoint of the ATP recruitment period. Questionnaire completion rates were calculated and Pearson's chi-square test was used to assess proportional

differences in completion rates across socio-demographic domains. All statistical tests were performed at a 0.1% level of significance using SAS version 9.2 (SAS Institute, Cary, NC) on a Linux interface.

RESULTS

Objective 1. Describe the recruitment, enrollment, data collection and quality control methods used to create the ATP between 2001 and 2009 (Phase I)

The RDD process resulted in 63,547 people (42% men, 58% women) indicating that they would be willing to receive an enrollment package from ATP. Of the people sent an enrollment package, 49% enrolled, and the Phase I cohort consisted of 31,072 participants (39% men, 61% women; mean (SD) age 50.2 (9.2)). Participants were enrolled across the entire province of Alberta (Figure 2). Nearly 75% of participants lived in urban areas, defined using rural postal codes with a second digit of 0. A high proportion of individuals were of European descent (92%).

Approximately 99% (*n*= 30,658) of participants consented to allow ATP to access their healthcare utilization and outcome data by linking with administrative databases, and 30,431 (98%) provided a valid PHN. Additionally, 93% of the participants consented to be contacted in the future with an invitation to provide a biospecimen. Correlations between characteristics reported by participants recruited from the same household were evaluated for ICC ≥0.8. [11] Age, annual household income, body mass index (BMI), and Asian ethnicity had ICC ≥0.8. Educational status, occupational status, all other ethnicities, smoking status, second hand smoke exposure, and self-reported diagnosis of 13 different chronic conditions had ICC <0.8. Based on this analysis, it may not be necessary to remove 'second in household' participants from future studies, but the decision must be made on a case by case basis.

Objective 2. Describe characteristics reported by ATP participants at enrollment and compare with characteristics of Albertans reported in national surveys

ATP participants were distributed evenly across 10-year age categories. There was wide variation in total annual household income, and approximately half of the cohort reported

completion of some post-secondary education, while 10% had not completed high school. The proportions of men and women who reported heights and weights that categorized them as obese were similar (28.4% men; 26.4% women), however a greater proportion of men than women were overweight (48.4% vs 33.2%), and a greater proportion of women than men had a BMI in the normal range (39.4% vs 23.0%). Approximately four fifths of men and women reported being current non-smokers, and 15% were daily smokers. High cholesterol and high blood pressure were the most commonly reported chronic health conditions at enrollment to ATP. The proportion of missing values was very low (Table 2).

Compared to the CCHS Cycle 3.1 Alberta weighted data, a higher proportion of ATP participants had participated in a least some post secondary education. The distributions of frequencies for marital status and total annual household income were similar between ATP participants and the Alberta CCHS data. ATP participants had higher BMI, with lower proportions of participants with a BMI in the normal range (23% vs. 34% of men, 39% vs. 47% of women) and a higher proportion in the obese category (28% vs. 20% of men, 26% vs. 16% of women). ATP men reported a lower incidence of daily smoking (16%) compared to CCHS Alberta men (21%), while the distribution of smoking status was similar for women between ATP and CCHS (Table 2).

Objective 3. Explore characteristics of ATP participants who completed follow- up questionnaires compared to those who did not

Completion rates for the diet and physical activity questionnaires varied by sex, age, working status and smoking status (Table 3). A greater proportion of women (88%) than men (83%) completed HLQ, CDHQ and PYTPAQ. Similarly, all three questionnaires were completed by around 90% of adults 55 to 69 years, compared with 84% of adults 35 to 44 years. A greater proportion of retired participants (92%) and lower proportion of participants employed full time (84%) completed all three questionnaires. Finally, 87% of former/never smokers completed the HLQ, CDHQ and PYTPAQ, compared with 80% of current daily smokers (Table 3).

The first follow-up survey was completed by 9,197 (79%) participants. Survey 2004 Express was mailed to 2,431 people, and was completed by only 19%; 83% completed either Survey 2004 or Survey 2004 Express. Overall, the strategy of using a truncated version of the questionnaire to try to boost response rates was found to be inefficient and resource-intensive

and therefore not considered for future follow-ups. Characteristics of respondents and non-respondents to Survey 2004 showed the same patterns as observed for CDHQ and PYTPAQ completion. Relative to respondents, non-respondents were younger, more likely to work full time and to report being daily smokers. Survey 2008 was completed by 20,707 (72%) participants. In contrast to response patterns observed in Survey 2004, there were very few differences in socio-demographic characteristics reported by respondents and non-respondents (Table 3).

Objective 4. Describe changes over time in characteristics reported by ATP participants at enrollment and at follow-up

Changes in socio-demographic and health-related variables reported from enrollment to Survey 2008 were examined. The frequencies of self-reported physician diagnosis of high blood pressure (23% at enrollment, 31% at Survey 2008), high cholesterol (27% at enrollment, 37% at Survey 2008) and heart attack (2% at enrollment, 8% at Survey 2008) all increased from enrollment to Survey 2008. The proportion of participants with a family history of cancer increased from 54% at enrollment to 60% at Survey 2008.

INTERPRETATION

The ATP Phase I cohort is well placed to support a broad range of health-related research initiatives. Comprehensive data related to socio-demographic, environmental and lifestyle domains have been obtained, and rigorous quality control procedures have resulted in well-documented databases with low frequencies for missing data. The very high proportion of participants who have consented to linkage with administrative databases is a strength, as such linkages will facilitate research on contextual factors that may be important in understanding how patterns of health services utilization are associated with health outcomes.[14,27] Another strength is the ability to link current residence, and potentially historical residences, with environmental data, using Geographic Information Systems technology based on either postal codes or street addresses to map different types of exposures against health outcomes.[28] Furthermore, many variables are harmonized to facilitate comparison or pooling with other large

cohorts,[29] which is particularly important for future studies that require larger sample sizes to explore how exposures interact to affect disease risk.[30]

ATP used RDD in an attempt to draw a large (>10,000) sample of the general population of Alberta aged 35-69 years. This approach is somewhat different from that used by many earlier North American cohorts which have either limited enrollment to people engaged in specific occupations (e.g. nurses,[31] physicians,[32,33] teachers[34]), those living in smaller geographic units such as particular towns[35] or listed on a registry that permitted random sampling.[36,37] During establishment of ATP, there were no population-based registries that could be used as a sampling frame, and as other approaches based on existing surveillance methods were unfeasible,[15] RDD was selected as the most viable option. Given recent trends pointing towards a decline in use of fixed land-lines,[38] it is unclear if RDD will continue to remain viable for recruitment to large cohorts. Indeed, there have been suggestions that approaches using newer technologies may help reduce costs associated with recruitment and enrollment,[39] but any approach, regardless of whether it is based on established or emerging technology, should be evaluated rigorously before full implementation.

One criticism levelled at cohorts is that they often comprise highly educated, health conscious participants with relatively low prevalence of chronic disease.[40–42] While it is true that prospective cohorts drawn from the general population rarely, if ever, comprise a representative sample of their source population, the use of RDD in ATP resulted in a cohort of participants from all over the province who reported a wide range of socio-demographic and health-related characteristics. While the proportions of men and women in ATP were somewhat unbalanced, it should be noted that this is not unusual for prospective cohorts. For example, the Genetic Epidemiology Research on Adult Health and Aging cohort (NIH accession number: phs000674.v1.p1)[37] established by Kaiser-Permanente in the USA comprises 42% men and the National Institutes of Health-American Association of Retired Persons Diet and Health study[36] comprises 60% men. Other cohorts have been limited only to one sex.[31,33,34] Despite these apparent limitations, all of these cohorts have made valuable contributions to our understanding of the antecedents of cancer and chronic disease, because like ATP, they have used rigorous approaches for collection and management of data obtained from the same people over time, and have had the capacity to link with administrative databases to obtain and analyze information on health outcomes.

With respect to other health-related characteristics, the ATP cohort appears broadly similar to its source population. Proportions of participants who reported conditions such as hypertension and diabetes are comparable to those reported in national surveillance data,[26] and are also similar to those reported recently by Quebec's CARTaGENE study[27] and the UK Biobank[43]; both of which drew samples from provincial and national health insurance databases. However, the ATP cohort seems to be somewhat different from the Alberta population in terms of overweight and obesity. Contrary to what might have been expected, the prevalence of obesity observed in men (28.4%) and women (26.4%) was substantially higher than reported by CCHS 3.1 (19.5%, 15.9%).[26] However, subsequent comparisons of measured and self-reported heights and weights in a sub-sample of CCHS participants aged 18-79 years resulted in adjusted estimates of obesity prevalence 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.[44] Despite the fact that heights and weights were self-reported by ATP participants, the prevalence of obesity in the cohort was somewhat closer to prevalence estimates based on measured heights and weights in Canadians aged ≥35 years.[45] We have speculated that providing detailed instructions for measurements (including diagrams) and including a tape measure in the package may have resulted in more accurate reporting of height and weight than would have been obtained by simply asking "how tall are you without shoes?" and "what weight are you?". However we have no evidence for this assumption, and more work is required to understand better how question wording and mode of administration are likely to affect the quality of self-reported anthropometric data.

Our exploration of response rates for different waves of questionnaire administration demonstrated that respondents to the second set of enrollment questionnaires and the first follow-up questionnaire were more likely to be women, older, retired and non-smokers. These characteristics are typical of those commonly reported for participants thought to be more engaged in research.[46] Intriguingly though, there were few differences observed in characteristics of respondents and non-respondents to the second follow-up questionnaire. 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 exploration of retention strategies continues to be a major priority for ATP. Although we have obtained consent for passive follow-up through linkage with administrative

databases, we plan to continue with active follow-up to enrich the databases and facilitate careful examinations of how changes over time in exposures or health status are likely to impact long term health outcomes. To date, we have been able to maintain a follow-up response rate of between 72% and 83% of participants who returned a completed HLQ. These rates are comparable to those reported by others such as the French E3N Cohort Study (80%)[47] and the Nord Trondelag Health Study (73-80%).[42] Nonetheless, we are becoming increasingly aware of the need to explore and implement strategies to engage participants effectively.

CONCLUSION

ATP Phase I is a robust platform that has and will continue to support a wide range of health-related research studies. ATP Phase I is currently being augmented by Phase II (the Canadian Partnership for Tomorrow Project protocol),[48] which includes the collection of additional health and lifestyle data, measured anthropometric variables, and collection of biospecimens. Both phases will result in the creation of rich repositories of data and biospecimens that may be accessed by *bona fide* researchers who have questions that will advance knowledge in cancer and chronic disease etiology and early detection.

HOW TO ACCESS THE DATA

A list of previous publications, communications, and other information can be found at www.ABTP.ca. Access requests from national and international researchers are welcome. Standard operating procedures for data collection, processing, and storage protocols are available upon request. Access to data and samples from the questionnaires (HLQ, CDHQ, PYTPAQ, Survey 2004, Survey 2008) will be available to external researchers upon successful completion and approval of an access request. Information and details on how to request access to ATP data can be found at www.ABTP.ca, or by emailing ATP.Research@AHS.ca.

ACKNOWLEDGEMENTS AND 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 views expressed herein represent the views of the author(s) and not of Alberta's Tomorrow Project or any of its funders. The data product presented here from CCHS is provided 'as-is,' and Statistics Canada makes no warranty, either express or implied, including but not limited to, warranties of merchantability and fitness for a particular purpose. In no event will Statistics Canada be liable for any direct, special, indirect, consequential or other damages, however caused.

Conflict of Interest: None declared

APPENDIX

Figure 1. *Enrollment and follow-up of participants in Alberta's Tomorrow Project*. Bars in the Enrollment Measurements panel depict total number of participants enrolled to date and the proportion of participants who completed either only the HLQ (grey section) or all three enrollment questionnaires (HLQ, CDHQ, PYTPAQ; black section), such that by the end of enrollment in 2009 there were 31,072 participants and 86% had completed all three questionnaires. The Follow-Up Measurements panel describes the number of participants who completed the follow-up questionnaires in 2004 and 2008. For 2004, individuals who were enrolled between 2001 and 2003 were mailed Survey 2004 and/or Survey 2004 Express, and 83% completed a questionnaire. The bar shading represents the proportion of individuals who: completed questionnaires at enrollment but did not complete Survey 2004 (white area); completed only HLQ and Survey 2004 (grey area); and completed all 3 enrollment questionnaires and Survey 04 (black area). Please note that Survey 2004 includes the completion of either Survey 2004 or Survey 2004 Express. The follow-up questionnaire in 2008 was sent to individuals enrolled between 2001 and 2007, meaning that some individuals will have previously completed Survey 2004. The overall response rate for Survey 2008 was 72%. The bars represent the proportion of individuals who: did not return Survey 2008 (white area); completed only HLQ and Survey 2008 (grey area); completed all 3 enrollment questionnaires and Survey 2008 (black area); completed HLQ, Survey 2004 and Survey 08 (dark grey area); and completed all 3 enrollment questionnaires, Survey 2004, and Survey 2008 (hatched area).

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

Table 1. Questionnaire domains captured in Alberta's Tomorrow Project enrollment and follow-up questionnaires

Table 2. Select characteristics of Alberta's Tomorrow Project participants at enrollment compared with respondents in Alberta to the 2005 Canadian Community Health Survey (CCHS Cycle 3.1)

Table 3. Enrollment characteristics reported by participants who completed follow-up questionnaires compared with those reported by non-respondents to follow-up questionnaires

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

References

1 World Health Organization. Noncommunicable diseases country profiles 2014. 2014.

2 World Health Organization. Global status report on noncommunicable diseases 2014. 2014.

3 Bauer UE, Briss PA, Goodman RA, Bowman BA. 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, Maximova K, Paradis G, O'Loughlin JL. Chronicling changes to the chronic disease prevention landscape in Canada's public health system 2004-2010. *Public Health* 2014; **128**: 716–724.

5 Ashbury FD, Little J, Ioannidis JPA, Kreiger N, Palmer LJ, Relton C *et al.* A vision for chronic disease prevention and intervention research: Report from a workshop. *Can J Public Heal* 2014; **105**: e150–3.

6 Meschia JF, Bushnell C, Boden-Albala B, Braun LT, Bravata DM, Creager MA *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, Mesrine S, Clavel-Chapelon F. Association between five lifestyle habits and cancer risk: Results from the E3N cohort. *Cancer Prev Res*; **7**: 516–25.

8 American Cancer Society. American Cancer Society. Breast cancer facts & figures 2013-14. Atlanta, 2013.

9 World Cancer Research Fund, American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective. Washington, DC, 2007.

10 Hystad P, Brauer M, Demers PA, Johnson KC, Setton E, Cervantes-Larios A *et al.* Geographic variation in radon and associated lung cancer risk in Canada. *Can J Public Heal* 2014; **105**: e4–10.

11 Colditz GA, Wei EK, Edu C, Org W. 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, Bakker SJ, Dotinga A, Vonk JM *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol* 2015; **44**: 1172–1180.

14 Doiron D, Raina P, Fortier I. Linking Canadian population health data: maximizing the potential of cohort and administrative data. *Can J Public Heal* 2013; **104**: e258–61.

15 Bryant H, Robson PJ, Ullman R, Friedenreich C, Dawe U. Population-based cohort

development in Alberta, Canada: A feasibility study. *Chronic Dis Can* 2006; **27**: 51–9.

16 National Institutes of Health. Diet History Questionnaire. Washington, DC, 2007.

17 Csizmadi I, Kahle L, Ullman R, Dawe U, Zimmerman TP, Friedenreich CM *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.

18 Friedenreich CM, Courneya KS, Neilson HK, Matthews CE, Willis G, Irwin M *et al.* Reliability and validity of the Past Year Total Physical Activity Questionnaire. *Am J Epidemiol* 2006; **163**: 959–970.

19 Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, Mcnutt S *et al.* Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires. *Am J Epidemiol*; **154**: 1089–1099.

20 Thompson F, Subar A, Brown C, Smith A, Sharbaugh C, Jobe J *et al.* Cognitive research enhances accuracy of food frequency questionnaire reports: Results of an experimental validation study. *J Am Diet Assoc* 2002; **102**: 212–218.

21 Friedenreich C, Courneya K, Bryant H. The lifetime total physical activity questionnaire: Development and reliability. *Med Sci Sports Exerc* 1998; **30**: 266–274.

22 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 (eds). *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach*. Duke University Press: Durham, NC, 1992, pp 277–290.

23 EuroQol Group. EuroQol* -a new facility for the measurement of health-related quality of life. *Health Policy (New York)* 1990; **16**: 199–206.

24 Craig C, Marshall A, Sjostrom M, Bauman A, Booth M, Ainsworth B *et al.* International physical activity questionnaire: 12-Country reliability and validity. *Med Sci Sports Exerc* 2003; **35**: 1381–1395.

25 Cerin E, Saelens B, Sallis J, Frank L. Neighborhood Environment Walkability Scale: validity and development of a short form. *Med Sci Sports Exerc* 2006; **38**: 1682–1691.

26 Statistics Canada. Canadian Community Health Survey 3.1 Final Questionnaire. 2005.

27 Awadalla P, Boileau C, Payette Y, Idaghdour Y, Goulet J-P, Knoppers B *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.

28 Edwards SE, Strauss B, Lynn Miranda M. Geocoding large population-level administrative datasets at highly resolved spatial scales. *Trans GIS* 2014; **18**: 586–603.

29 Fortier I, Burton PR, Robson PJ, Ferretti V, Little J, Deschênes M *et al.* Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. *Int J Epidemiol* 2010; **39**: 1383–1393.

30 Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J *et al.* Size matters: just how big is BIG? Quantifying realistic sample size requirements for human genome

epidemiology. *Int J Epidemiol* 2009; **38**: 263–273.

31 Belanger CF, Hennekens CH, Rosner B, Speizer FE. The Nurses' Health Study. *Am J Nurs* 1978; **78**: 1039–1040.

32 Hennekens C, Eberlein K. A randomized trial of aspirin and B-carotene among U.S. physicians. *Prev Med (Baltim)* 1985; **14**: 165–168.

33 Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. *Br Med J* 1956; **10**: 1071–81.

34 Bernstein L, Allen M, Anton-Culver H, Deapen D, Horn-Ross PL, Peel D *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.

35 Dawber TR, Meadors GF, Moore FE. Epidemiological Approaches to Heart Disease: The Framingham Study. *Am J Public Health Nations Health* 1951; **41**: 279–81.

36 Schatzkin A, Subar AF, Thompson FE, Harlan LC, Tangrea J, Hollenbeck AR *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.

37 dbGaP Study Accession: phs000674.v1.p1. Resource for Genetic Epidemiology Research on Adult Health and Aging (GERA). http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000674.v1.p1 (accessed 26 Nov2015).

38 Statistics Canada. Statistics Canada. Residential Telephone Service Survey. December 2010. 2010.http://www.statcan.gc.ca/daily-quotidien/110405/dq110405a-eng.htm (accessed 8 Dec2015).

39 Toledano MB, Smith RB, Brook JP, Douglass M, Elliott P. How to establish and follow up a large prospective cohort study in the 21st century - Lessons from UK COSMOS. *PLoS One* 2015; **10**: e0131521.

40 Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol* 2007; **17**: 643–653.

41 Banks E, Herbert N, Mather T, Rogers K, Jorm L. 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**: 1.

42 Langhammer A, Krokstad S, Romundstad P, Heggland J, Holmen J. The HUNT study: participation is associated with survival and depends on socioeconomic status, diseases and symptoms. *BMC Med Res Methodol* 2012; **12**: 143.

43 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**: 30123579.

44 Shields M, Connor-Gorbor S, Janssen I, Tremblay M. Bias in self-reported estimates of obesity in Canadian health surveys: An update on correction equations for adults. *Heal Reports* 2011; **22**.

45 Statistics Canada. Statistics Canada. Table 105-0407 - Measured adult body mass index (BMI), by age group and sex, household population aged 18 and over excluding pregnant females. Canadian Community Health Survey (CCHS 3.1), Canada, every 2 years. CANSIM. .

46 Littman AJ, Boyko EJ, Jacobson IG, Horton J, Gackstetter GD, Smith B *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.

47 Clavel-Chapelon F. Cohort Profile: The French E3N Cohort Study. *Int J Epidemiol* 2015; **44**: 801–9.

48 Borugian MJ, Robson P, Fortier I, Parker L, McLaughlin J, Knoppers BM *et al.* The Canadian Partnership for Tomorrow Project: Building a pan-Canadian research platform for disease prevention. *CMAJ* 2010; **182**: 1197–1201.

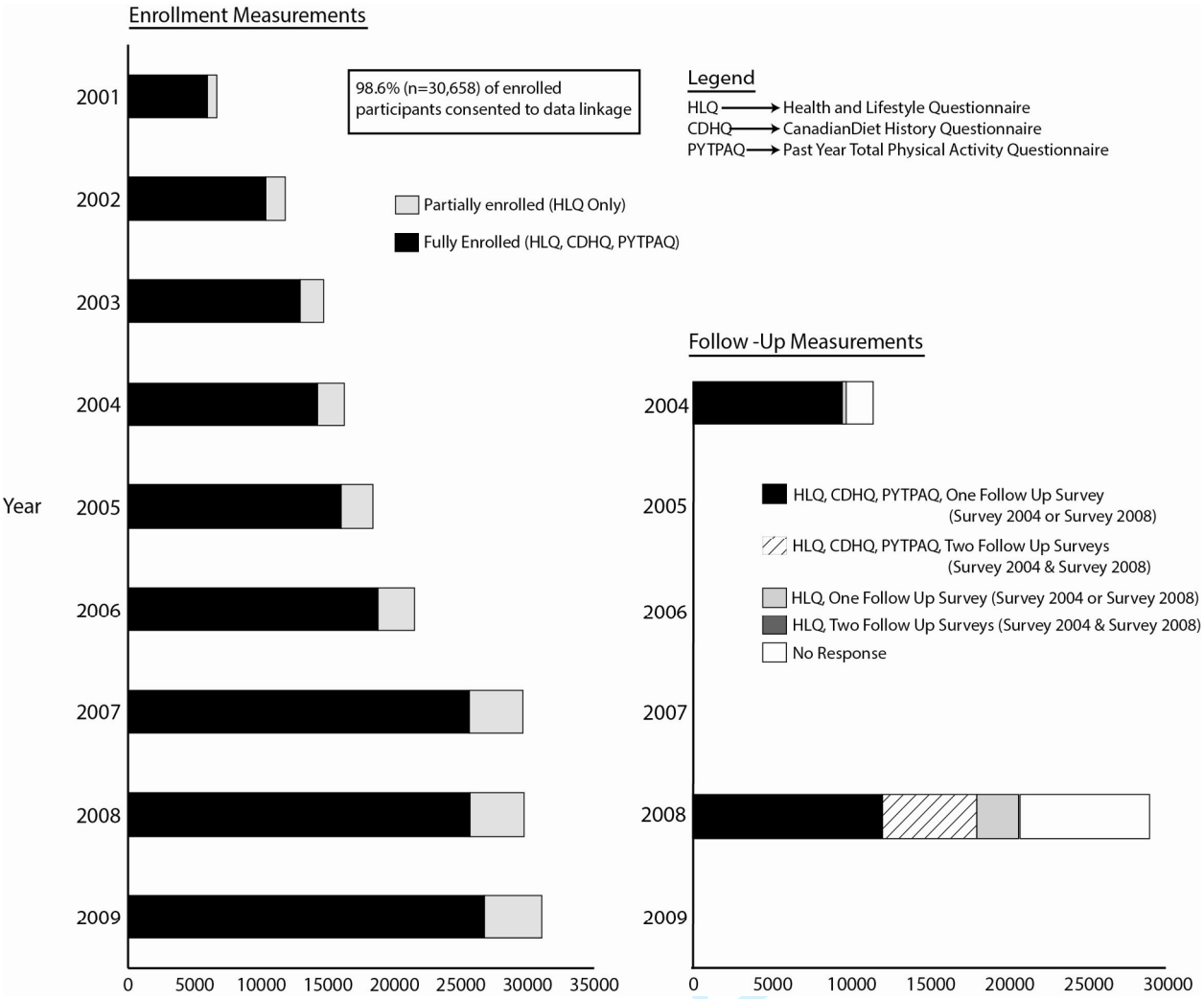

Figure 1. Enrollment and follow-up of participants in Alberta's Tomorrow Project

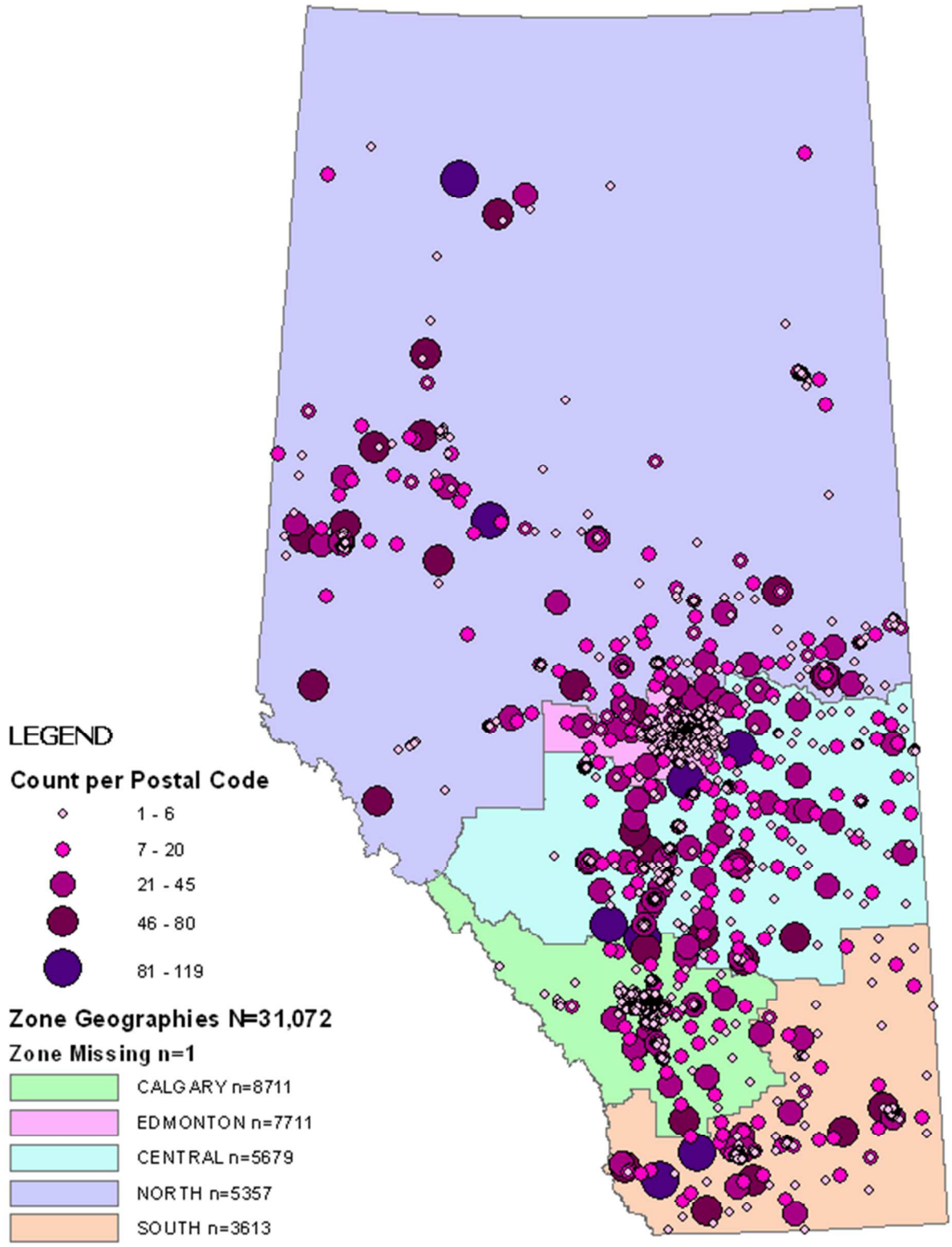

Figure 2. Geographic postal code region coverage, location of participants in Alberta's Tomorrow Project at enrollment, Alberta, Canada

Table 1. Questionnaire domains captured in Alberta's Tomorrow Project enrollment and follow-up questionnaires, 2000-2009

Measurements	Enrollment (2001-2009)	Follow-up (2004 and 2008)	
	Questionnaires[a] (*n*=31,072)	Survey 2004[b] (*n*=9,660)	Survey 2008 (*n*=20,707)
Demographic Information			
Current employment status	✓	✓	✓
Occupational history			✓
Shift work			✓
Marital status	✓	✓	✓
Education level	✓		✓
Annual household income	✓		✓
Ethnic background	✓	✓	✓
Cancer and Chronic Disease			
Personal health history[c]	✓	✓	✓
Personal cancer diagnosis	✓	✓	✓
Family structure (number of siblings, age of parents, cause of parental death)	✓		
Family history of cancer	✓		✓
Family history of health conditions[d]	✓		
Anthropometrics[e]			
Height	✓	✓	✓
Weight	✓	✓	✓
Waist circumference	✓	✓	✓
Hip circumference	✓	✓	✓
Lifetime weight patterns		✓	
Diet			
Food frequency questionnaire (including use of supplements)	✓		
Past 7 day intake of fruit and vegetables			✓
Physical Activity[f]			
Employment/volunteer activities	✓		✓
Household activities	✓		✓
Recreation/leisure activities	✓		✓
Sedentary behaviours			✓
Smoking and Tobacco			
Current and past use of tobacco	✓	✓	✓
Second hand smoke exposure	✓		
Quitting behaviors			✓

Alcohol			
Alcohol use	✓	✓	
Beverage type and amount	✓	✓	
Sleep			
Sleep pattern		✓	✓
Screening and Risk Behaviors			
Colon cancer screening (fecal occult blood test, colonoscopy, sigmoidoscopy)	✓	✓	✓
Sun exposure – sunburn history	✓	✓	
Sun exposure – sunscreen use, tanning, risk of sunburn		✓	
Primary care service utilization			✓
Men's Reproductive Health			
PSA screening	✓	✓	✓
Enlarged prostate	✓		
Vasectomy	✓		
Women's Reproductive Health			
Pap test screening	✓	✓	✓
Mammogram screening	✓	✓	✓
Breast exam	✓		
Menstruation (age at onset)	✓		
History of pregnancy and breastfeeding	✓		
Oral contraceptive use	✓		
Menopause (age, use of hormone replacement and alternative therapies)	✓	✓	
Oophrectomy or hysterectomy	✓	✓	✓
Perceived Health and Quality of Life			
General health rating	✓	✓	✓
Stress and emotional state (anxiety, depression)	✓	✓	
Social support	✓		
Spirituality	✓		
Quality of life (mobility, self-care, pain)		✓	✓
Perception of risk for cancer and diabetes		✓	
Built Environment			
Built environment			✓
Residential history			✓
Postal code	✓	✓	✓

a- Enrollment measurements included the Health and Lifestyle Questionnaire (HLQ; n=31,072), the Canadian Diet History Questionnaire (CDHQ; n=26,843), and the Past Year Total Physical Activity Questionnaire (PYTPAQ; n=26,769). n=26,761 completed all 3 (HLQ, CDHQ and PYTPAQ).
b- An abbreviated version of Survey 2004 (Survey 2004 Express) was mailed to participants who did not return Survey 2004 in order to try to capture partial data on these individuals. For the Express version, fewer questions

were asked under each topic but the questions that were asked were the same between questionnaires. n=9,197 individuals completed Survey 2004 and n=463 individuals completed Survey 2004 Express.

c- Personal health history included self-reported doctor diagnoses of chronic health conditions including:

- All Questionnaires: angina, chronic bronchitis, Crohn's disease, cirrhosis of the liver, diabetes, emphysema, heart attack, hepatitis, high blood pressure, high cholesterol, polyps in colon and rectum, stroke, ulcerative colitis
- Survey 2004 additional: arthritis, depression, high blood sugar, osteoporosis, thyroid problems
- Survey 2008 additional: asthma, acid reflux, arthritis, heart problems, irritable bowel syndrome, osteoporosis, thyroid problems

d- Family history of health conditions included diabetes, heart attack and stroke

e- Anthropometric measurements are self-reported

f- Physical activity was measured over past year (PYTPAQ) at enrollment and over past week (International Physical Activity Questionnaire) at Survey 2008

Table 2. Select characteristics of Alberta's Tomorrow Project participants at enrollment compared with respondents in Alberta to the 2005 Canadian Community Health Survey (CCHS Cycle 3.1)

Self-reported domains	Men		Women	
	ATP %	CCHS Alberta† %	ATP %	CCHS Alberta† %
Gender	39.0	49.8	61.0	50.2
Age (years) **				
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 post secondary[a]	18.7	5.7	22.3	6.4
Post-secondary completed[b]	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
BMI (kg/m^2)[c] **				
<18.5	0.2	0.4	1.1	2.8
18.5-24.9	23.0	34.0	39.4	46.8
25.0-29.9	48.4	45.3	33.2	29.9
≥30.0	28.4	19.5	26.4	15.9
Missing	0.0	0.8	0.0	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[d]	80.5	74.4	81.9	79.3
Missing	0.1	0.3	0.0	0.1
Chronic health condition[e]				
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's disease	0.5	0.2	0.7	0.3

† CCHS 3.1 Alberta weighted data, restricted to CCHS 3.1 Albertan respondents 35-69 years old as per ATP inclusion criteria

a- Some post-secondary includes combined responses to: some technical school/college training completed, some part of university degree completed
b- Post secondary completed includes combined responses to: completed technical school/college training, completed university degree, some part of post-graduate university degree completed, completed university post-graduate degree
c- BMI derived from participant self-reported height and weight
d- "Not at all" smoking status includes never smokers and former smokers
e- 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's disease. Similar statistic available from CCHS Cycle 3.1

** P< 0.001 ATP Enrollment versus CCHS3.1 Alberta weighted data

**/ P<0.001 ATP Enrollment versus CCHS3.1 Alberta weighted data men only

/** P<0.001 ATP Enrollment vs CCHS3.1 Alberta weighted data women only

Table 3. Enrollment characteristics reported by participants who completed follow-up questionnaires compared with those reported by non-respondents to follow-up questionnaires

Response	Baseline		Survey 2004‡		Survey 2008	
Total n	Fully enrolled[a] (*n*=26,761)	**Partially enrolled[b] (*n*=4,311)**	**Returned (*n*=9,660)**	**No response (*n*=1,971)**	**Returned (*n*=20,707)**	**No response (*n*=8,181)**
Gender		**		**		
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 (years)		**		**		
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[c]	20.6	22.9	21.2	22.9	20.7	21.2
Post-secondary completed[d]	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[e]						
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
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
BMI (kg/m^2)[f]	**		**			
<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[g]	**		**			

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[h]						
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

a- Fully enrolled: Participants who completed HLQ, CDHQ and PYTPAQ

b- Partially enrolled: Participants who completed HLQ and did not return CDHQ or PYTPAQ

c- Some post-secondary includes combined responses to: some technical school/college training completed, completed technical school/college training, some part of university degree completed

d- Post-secondary completed includes combined responses to: completed university degree, some part of post-graduate university degree completed, completed university post-graduate degree

e- Geographical location defined according to rural postal code, where "0" as the second digit indicates rural residence

f- BMI derived from participant self-reported height and weight

g- 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, other cancer

h -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, stroke

** $p<0.001$ Fully enrolled vs partially enrolled or Survey 2004 returned vs Survey 2004 no response or Survey 2008 returned vs Survey 2008 no response

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

	Men (*n*=7,788)		Women (*n*=12,919)	
	Enrollment†	**Survey 2008**	**Enrollment†**	**Survey 2008**
[a]BMI **				
<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
[b]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's 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
[c]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

† Enrollment data presented only for the participants who completed Survey 2008 (total n=20,707)

a- BMI was derived from participant self-reported height and weight

b- 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's disease, angina, high cholesterol, heart attack, stroke, hepatitis, cirrhosis of the liver.

c- Family history of cancer: 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

** - P<0.001 from enrollment to follow-up

Reporting Guideline Checklist

STROBE Statement—checklist of items that should be included in reports of *observational studies*

	Item No	Recommendation	Met? Yes/No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Yes
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Yes
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Yes
Objectives	3	State specific objectives, including any prespecified hypotheses	Yes
Methods			
Study design	4	Present key elements of study design early in the paper	Yes
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Yes
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	Yes
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Yes
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Yes
Bias	9	Describe any efforts to address potential sources of bias	Yes
Study size	10	Explain how the study size was arrived at	Yes
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were	Yes

		chosen and why	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Yes
		(*b*) Describe any methods used to examine subgroups and interactions	Yes
		(*c*) Explain how missing data were addressed	Yes
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	Yes
		(*e*) Describe any sensitivity analyses	N/A

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Yes
		(b) Give reasons for non-participation at each stage	Yes
		(c) Consider use of a flow diagram	Yes
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Yes
		(b) Indicate number of participants with missing data for each variable of interest	Yes
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	Yes
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	Yes
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Yes
		(*b*) Report category boundaries when continuous variables were categorized	Yes
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Yes
Discussion			
Key results	18	Summarise key results with reference to study objectives	Yes
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Yes
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Yes
Generalisability	21	Discuss the generalisability (external validity) of the study results	Yes
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Yes

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.