

## Chronic disease surveillance in Alberta's Tomorrow Project using Administrative Health Data

Ming Ye<sup>1</sup>, Jennifer E. Vena<sup>2</sup>, Jeffrey A. Johnson<sup>1</sup>, Grace Shen-Tu<sup>2</sup>, and Dean T. Eurich<sup>1,\*</sup>

### Submission History

Submitted:	09/04/2021
Accepted:	24/06/2021
Published:	21/10/2021

<sup>1</sup>School of Public Health, University of Alberta, Edmonton, Alberta, Canada, T6G 2E1

<sup>2</sup>Alberta's Tomorrow Project, Cancer Care Alberta, Alberta Health Services, Alberta, Canada, T2T 5C7

### Abstract

#### Introduction

Alberta's Tomorrow Project (ATP) is the largest population-based prospective cohort study of cancer and chronic diseases in Alberta, Canada. The ATP cohort data were primarily self-reported by participants on lifestyle behaviors and disease risk factors at the enrollment, which lacks sufficient and accurate data on chronic disease diagnosis for longer-term follow-up.

#### Objectives

To characterize the occurrence rate and trend of chronic diseases in the ATP cohort by linking with administrative healthcare data.

#### Methods

A set of validated algorithms using ICD codes were applied to Alberta Health (AH) administrative data (October 2000-March 2018) linked to the ATP cohort to determine the prevalence and incidence of common chronic diseases.

#### Results

There were 52,770 ATP participants ( $51.2 \pm 9.4$  years old at enrollment and 63.7% females) linked to the AH data with average follow-up of  $10.1 \pm 4.4$  years. In the ATP cohort, hypertension (18.5%), depression (18.1%), chronic pain (12.8%), osteoarthritis (10.1%) and cardiovascular diseases (8.7%) were the most prevalent chronic conditions. The incidence rates varied across diseases, with the highest rates for hypertension (22.1 per 1000 person-year), osteoarthritis (16.2 per 1000 person-year) and ischemic heart diseases (13.0 per 1000 person-year). All chronic conditions had increased prevalence over time ( $p < 0.001$  for trend tests), while incidence rates were relatively stable. The proportion of participants with two or more of these conditions (multi-morbidity) increased from 3.9% in 2001 to 40.3% in 2017.

#### Conclusions

This study shows an increasing trend of chronic diseases in the ATP cohort, particularly related to cardiovascular diseases and multi-morbidity. Using administrative health data to monitor chronic diseases for large population-based prospective cohort studies is feasible in Alberta, and our approach could be further applied in a broader research area, including health services research, to enhance research capacity of these population-based studies in Canada.

#### Keywords

chronic disease; surveillance; administrative health data; Alberta's tomorrow project

\*Corresponding Author:

Email Address: [deurich@ualberta.ca](mailto:deurich@ualberta.ca) (Dean T. Eurich)

## Introduction

Launched in October 2000, Alberta's Tomorrow Project (ATP), a province-wide population-based cohort study of cancer and chronic diseases in Alberta, Canada, started to recruit participants on a rolling basis until 2015 [1]. Alberta residents aged 35–69 years with no history of cancer [other than non-melanoma skin cancer] were eligible to participate and receive an invitation package, including a cover letter, a Study Information Booklet, a detailed consent to participating in ATP and allowing health data linkage [1]. A total of 52,810 participants joined the ATP as of March 2015. As a research platform, ATP aims to capture the natural course of cancer and chronic diseases and support research in understanding the underlying risk factors of chronic conditions by following the cohort for up to 50 years [1]. Other than biological markers and physical measurements, ATP cohort data are primarily self-reported information provided by participants on lifestyle behaviors and disease risk factors, which often bears the limitation of insufficient information on disease diagnoses and varied validity due to relatively high response bias (e.g., more socially acceptable answers) in self-report data. With these concerns, ATP has made great effort to link the cohort to other health-related data sources, especially the administrative health data in Alberta, to improve broadness and validity of the ATP data. Detailed information on the ATP recruitment and cohort profile can be found in our previous publications [1].

Administrative health data are healthcare information systematically collected by the publicly funded healthcare systems in each province and territory in Canada, often including data on health services, diagnosis, medications and costs. Many studies have shown that administrative health data, with diagnostic information coded by International Classification of Diseases (ICD) codes, is a promising and reliable tool to identify patients with chronic conditions [2–8]. Quan et al. (2005, 2008 and 2009) have shown that using ICD codes in Alberta Health (AH) administrative data has acceptable (56.2–86.5%) sensitivity and great (>90%) specificity in identifying major chronic conditions, including hypertension, cardiovascular diseases, diabetes and chronic respiratory diseases [2, 3, 8]. Tonelli et al. (2015) also demonstrated that 30 out of 40 chronic conditions can be identified in a local population-based cohort in Alberta, with positive predictive value and sensitivity over 70% using administrative health data [4]. Administrative health information has also been widely used to characterize health conditions in many other countries. For example, using primary care and hospital admission records, U.K has created a chronological map of the 50 most common health conditions for 3.8 million patients [9]. These studies suggest that the administrative health data is an important complementary data source for large population-based cohort studies such as ATP, to improve the data broadness and quality for cancer and chronic diseases research.

The purpose of this study is to characterize the occurrence rate (prevalence and incidence) and the trends of chronic diseases, including multi-morbid chronic conditions, in the ATP cohort, using individual-level administrative health data.

## Methods

### Study population

Alberta residents aged 35–69 years with no history of cancer (other than non-melanoma skin cancer) based on self-report were eligible to participate in ATP [10]. As of March 2015, ATP has established a cohort of 52,810 participants who had completed baseline questionnaires and consented to data linkage to health-related databases; from those, 52,770 (>99.0%) provided valid Personal Health Numbers (PHN) to facilitate individual-level linkage to AH administrative data [10, 11]. ATP participants are actively followed up by requesting additional questionnaire data every four years until March 2050 and passive follow-ups are possible for ATP participants by linking ATP cohort data to administrative healthcare databases [1]. The study period of this analysis is from October 01, 2000 to March 31, 2018 when the latest AH data were linked. By linking with administrative health data, including Alberta Population Registry and Vital statistics (death), data on migration and death can be obtained for epidemiology censoring.

### Data sources

The Alberta Health (AH) administrative data, including healthcare services utilization, diagnosis, medications and cost, are routinely collected by the publicly funded provincial healthcare systems in Alberta. Ministry of Health, i.e. AH, is the custodian of these provincial health databases in Alberta. Detailed descriptions of AH databases can be obtained from the AH website: <http://www.health.alberta.ca/initiatives/health-research.html>.

The AH administrative datasets utilized in this study were Ambulatory Care, Inpatient data, Practitioner Claims, Population Registry, Alberta Blue Cross and Pharmaceutical Information Network (PIN) and Vital Statistics. Supplementary Table 1 shows a list of AH data elements used in this study, including diagnostic information coded with ICD codes, medication information with Anatomic Therapeutic Chemical (ATC) codes, provincial migration-in/out data from the Population Registry and death data from Vital Statistics.

### Data linkage

Each resident in Alberta has a unique PHN recorded for administrative and billing purposes when they interact with the healthcare systems in Alberta [12]. In this study, AH administrative datasets (October 2000–March 2018) were linked by AH using PHNs to the ATP cohort to identify cases with chronic diseases. After data linkage, AH healthcare data were de-identified before release to the authors of this study. The data linkage and data analyses conducted for this study had been approved by the Health Research Ethics Board of the University of Alberta (project ID Pro00058561).

### Case definition

A set of validated algorithms using ICD codes and ATC codes were applied to the linked AH administrative data to identify cases of most prominent chronic diseases [13] other

than cancer (in Alberta, the provincial cancer registry is a more reliable data source for cancer cases than administrative healthcare data). The algorithms used to identify cases of chronic diseases were either validated in Alberta's residents or commonly applied in established chronic disease surveillance programs, including the Government of Alberta Interactive Health Data Application (IHDA) and the Canadian Chronic Disease Surveillance System (CCDSS) (Supplementary Table 2). For example, in this study diabetes cases were identified using a modified version of the Canadian National Diabetes Surveillance System (NDSS) algorithm [14]:

*one hospitalization record with an ICD code of diabetes (ICD-9: 250, ICD-10: E10-E14) OR two physician claims within two years with an ICD code of diabetes OR self-report by participants, plus any of the following conditions: i) one hospitalization with ICD code for diabetes, ii) one physician claim with ICD code for diabetes, or iii) one diabetes medication with Anatomical Therapeutic Chemical Classification (ATC) code for insulin (A10A) or glucose-lowering drugs (A10B).*

In addition, multi-morbidity, an indicator of increased clinical complexity and disease burden, was defined as two or more of the top common (prevalence  $\geq 0.1\%$ ) chronic diseases investigated in this study.

## Data analysis

The ATP participants were recruited on a rolling basis from 2000 to 2015. To characterize the baseline (at-enrollment) health profile of the ATP cohort, the baseline prevalence of chronic diseases in the ATP cohort was calculated as the percentage of ATP participants who had been diagnosed with certain chronic diseases prior to the enrollment or up to 6 months after the enrollment (Supplementary Figure 1). The cumulative incidence rate since enrollment was also calculated, as the number of incident cases per 1,000 person-year (PY), for each chronic disease between the baseline (up to 6 months after the ATP enrollment) and the time when the latest AH data were linked (March 31, 2018). In these two calculations, cases were identified as "non-prevalent" or "incident" only if the first diagnosis date (index date) was  $>6$  months after enrollment ("clearance time") and not within the first year (October 01 2000–October 01 2001) when AH data was linked to the ATP cohort ("wash-out time"), to ensure the true "non-prevalent" or "incident" cases were identified; otherwise, cases were considered as "prevalent" cases at the baseline (Supplementary Figure 1). Among chronic diseases investigated in this study, depression and chronic pain are the two chronic conditions that could have multiple episodes during the course of disease. The baseline prevalence of these two conditions was thus calculated as the percentage of participants who had any episode(s) of the condition (i.e. depression or chronic pain) during the period between 3 years before the ATP enrollment (especially for participants recruited 2004 onwards) and 6 months after the enrollment.

To characterize the trend of chronic diseases in the ATP cohort over time, the annual prevalence and incidence rate for each chronic condition was calculated between 2001 and

2017. The prevalence and incidence rates were not calculated for the year of 2000 and 2018, given AH data were only available for three months in each year (October–December in 2000 and January–March in 2018) at the time of data linkage. In this calculation, incident cases are the new cases in each calendar year (e.g. in the year of 2001 from January 01, 2001 to December 31, 2001) and prevalent cases are the total number of cases that were existing in each calendar year, which includes the incident cases. Annual prevalence was calculated as the percentage of ATP participants who had certain conditions identified in each calendar year, where the denominator was counted as the total number of ATP participants who were alive and residents in Alberta by the end of each calendar year (Equation 1). Annual incident rates were calculated as the number of new cases (incidence cases) per 1000 ATP participants at risk of developing this condition (i.e. those who had not developed the condition at the beginning of each year) in each calendar year (Equation 2 and 3). ATP participants who were not registered Alberta residents based on Population Registry records (e.g. moving out of the province or death) in each calendar year were excluded from each calculation.

$$\begin{aligned} \text{Annual Prevalence Rate} &= \frac{\text{Total \# of prevalent in each calendar year}}{\text{Total \# of ATP participants by the end of each calendar year}} * 100 \end{aligned} \quad (1)$$

$$\begin{aligned} \text{Annual Incidence Rate} &= \frac{\text{Total \# of incident cases in each calendar year}}{\text{Total \# participants at risk in each calendar year}} * 1000 \end{aligned} \quad (2)$$

$$\begin{aligned} \text{Total \# participants at risk} &= \text{Total \# of participants by the end of each calendar year} \\ &\quad - \text{Prevalent cases} + \text{incident cases} \end{aligned} \quad (3)$$

For chronic diseases with multiple episodes (depression and chronic pain), only the episode (incidence) rate was calculated for each calendar year, where in Equation 3 participants at risk (the denominator) were the total number of participants (i.e. each participant was at risk of developing new episodes in each calendar year). We also calculated the age- and sex- standardized prevalence and incidence rates using 2016 Census-Alberta data (reference population) [15] for standardization (indirect standardization) to minimize the potential impact of different age- and sex- distribution on rate calculations and to facilitate comparison with other similar studies [4, 5].

The cohort was characterized using standard descriptive statistics (means and standard deviations or percentage where appropriate). We also used generalized linear models with prevalence (binomial distribution) or incidence (Poisson distribution, natural logarithm of person-year as offset) as dependent variables, calendar years (2001–2017), age groups (35–85+ years grouped by every 5 years based on the age on January 1<sup>st</sup> of each calendar year) and sex (male vs. female) as independent variables, to test the trend of chronic disease occurrence over time and the relationship with age and sex. Statistical analyses, with significant level at  $\alpha = 0.05$ , were conducted with STATA 14 software.

Data access and analyses in this study complies with the provincial Health Information Act (HIA) in Alberta, Alberta Health Services (AHS) and ATP data access and disclosure guidelines. Accordingly, in this study, we only report data with  $>30$  subjects across important categories (e.g. in each calendar year).

## Results

### The prevalence of chronic disease at the enrollment

The average age of 52,770 ATP participants who consented to individual level data linkage was  $51.2 \pm 9.4$  years at the time of enrollment. More than half of the cohort (63.7%) were women and 81.8% self-reported as being of Caucasian ethnicity. The baseline characteristics of study participants are shown in Supplementary Table 3. Detailed description of this cohort can be found in our previous publications [1, 16]. The AH administrative data (2000–2018) was linked to this cohort, with an average follow-up time of  $10.1 \pm 4.4$  years (from the ATP enrollment up to March 31, 2018). In this study, we only reported results for the top 16 chronic diseases that had a prevalence  $\geq 0.1\%$  at the enrollment, based on the linked administrative data (Table 1).

At the time of ATP enrollment (baseline), 18.5% of the participants (17.4% of women and 20.1% of men) had previously been diagnosed with hypertension, 18.1% (22.1% of women and 11.0% of men) with depression, 12.8% (13.8% of women and 11.2% of men) with chronic pain, 10.1% (10.8% of women and 9.0% of men) with osteoarthritis, 8.7% (7.2% of women and 11.5% of men) with cardiovascular diseases, 6.9% (9.4% of women and 2.4% of men) with hypothyroidism and 5.7% (4.9% of women and 7.1% of men) with diabetes (Table 1). In addition, at the time of enrollment, 23.6% of the ATP participants (25.6% of women and 20.0% of men) had two or more co-morbid conditions from the 16 most prevalent

chronic diseases (cardiovascular diseases were grouped as one single condition) investigated (Table 1).

### Cumulative incidence rate of chronic diseases since enrollment

Of the 16 top prevalent chronic conditions, hypertension (22.1 in total, 19.6 in women and 26.5 in men, per 1000 PY), osteoarthritis (16.2 in total, 17.4 in women and 14.1 in men, per 1000 PY) and ischemic heart diseases (13.0 in total, 11.0 in women and 16.6 in men, per 1000 PY) had the highest cumulative incidence rate since the ATP enrollment until March 2018, followed by hypothyroidism (7.8 in total, 10.3 in women and 3.9 in men, per 1000 PY), diabetes (6.2 in total, 5.1 in women and 8.1 in men, per 1000 PY) and obstructive sleep apnea (5.6 in total, 5.9 in women and 5.2 in men, per 1000 PY) (Table 2). For diseases that clinically present with multiple episodes, the cumulative incidence (episode) rate was 133.2 (167.8 in women and 77.2 in men) per 1000 PY for depression and 33.8 (36.9 in women and 28.7 in men) per 1000 PY for chronic pain. In addition, similar to the baseline prevalence, the cumulative incidence rate was significantly higher ( $p < 0.001$ ) in men than in women for diabetes, hypertension, cardiovascular diseases (CVD) and chronic obstructive pulmonary disease (COPD), and it was significantly lower ( $p < 0.001$ ) in men than in women for asthma, osteoarthritis, osteoporosis, hypothyroidism, depression and chronic pain (Table 2).

Table 1: The baseline (at-enrollment) chronic disease status of the ATP cohort

	Women (n = 33,627)		Men (n = 19,143)		Total (n = 52,770)	
	Freq.	Prev. (%)	Freq.	Prev. (%)	Freq.	Prev. (%)
Diabetes <sup>†</sup>	1,645	4.9	1,366	7.1	3,011	5.7
Hypertension <sup>†</sup>	5,862	17.4	3,903	20.1	9,765	18.5
Major CVD (total) <sup>†</sup>	2,405	7.2	2,195	11.5	4,600	8.7
Unstable angina <sup>†</sup>	181	0.5	344	1.8	525	1.0
Myocardial infarction <sup>†</sup>	128	0.4	327	1.7	455	0.9
Ischemic heart disease <sup>†</sup>	2,167	6.4	2,083	10.9	4,250	8.1
Heart Failure <sup>†</sup>	478	1.4	440	2.3	918	1.7
Transient ischemic attack <sup>†</sup>	41	0.1	32	0.2	73	0.1
Acute ischemic stroke <sup>†</sup>	58	0.2	73	0.4	131	0.2
Chronic Obstructive Pulmonary Disease <sup>†</sup>	305	0.9	229	1.2	534	1.0
Asthma <sup>†</sup>	1,496	4.4	552	2.9	2,048	3.9
Obstructive Sleep Apnea	1,051	3.1	656	3.4	1,707	3.2
Osteoarthritis <sup>†</sup>	3,647	10.8	1,719	9.0	5,366	10.1
Osteoporosis <sup>†</sup>	2,033	6.0	88	0.5	2,121	4.0
Hypothyroidism <sup>†</sup>	3,175	9.4	519	2.7	3,694	6.9
Chronic pain <sup>*†</sup>	4,630	13.8	2,138	11.2	6,768	12.8
Depression <sup>*†</sup>	7,423	22.1	2,106	11.0	9,529	18.1
Multi-morbidity <sup>**†</sup>	8,609	25.6	3,825	20.0	12,434	23.6

\*for conditions with multiple episodes, including depression and chronic pain, the baseline was defined as the time frame between 3 years before the ATP enrollment and 6 months after the ATP enrollment.

\*\*multi-morbidity was defined as 2 or more comorbidities from 16 chronic diseases investigated in this study; cardiovascular diseases in the table were grouped as one single condition as major cardiovascular diseases (CVD).

<sup>†</sup>there are statistically significant difference ( $p < 0.001$ ) between women and men.

Table 2: Cumulative incidence rate of chronic diseases in the ATP cohort (up to March 2018)\*

N=52,770	Women (n = 33,627)			Men (n = 19,143)			Total (n = 52,770)		
	Freq.	Person-year (PY)	Rate (per 1,000 PY)	Freq.	Person-year (PY)	Rate (per 1,000 PY)	Freq.	Person-year (PY)	Rate (per 1,000 PY)
Diabetes <sup>†</sup>	1,546	304328.8	5.1	1,416	174966.0	8.1	2,962	479294.8	6.2
Hypertension <sup>†</sup>	4,805	244532.7	19.6	3,664	138084.7	26.5	8,469	382617.4	22.1
Major CVD (total)** <sup>†</sup>	3,714	287825.4	12.9	3,021	160366.5	18.8	6,735	448191.9	15.0
Unstable angina <sup>†</sup>	158	325038.1	0.5	283	190866.1	1.5	441	515904.2	0.9
Myocardial infarction <sup>†</sup>	271	325356.7	0.8	455	190328.8	2.4	726	515685.5	1.4
Ischemic heart disease <sup>†</sup>	3,202	292394.8	11.0	2,696	162672.5	16.6	5,898	455067.3	13.0
Heart Failure <sup>†</sup>	822	319686.2	2.6	600	189008.0	3.2	1,422	508694.2	2.8
Transient ischemic attack <sup>†</sup>	69	326888.1	0.2	76	194956.6	0.4	145	521844.7	0.3
Acute ischemic stroke <sup>†</sup>	161	326425.3	0.5	153	194418.4	0.8	314	520843.7	0.6
Chronic Obstructive Pulmonary Disease <sup>†</sup>	980	320779.6	3.1	717	190559.2	3.8	1,697	511338.8	3.3
Asthma <sup>†</sup>	1,259	308853.8	4.1	523	188143.5	2.8	1,782	496997.3	3.6
Obstructive Sleep Apnea	1,842	310747.1	5.9	959	185235.9	5.2	2,801	495983.0	5.6
Osteoarthritis <sup>†</sup>	4,740	271614.5	17.4	2,369	167789.0	14.1	7,109	439403.5	16.2
Osteoporosis <sup>†</sup>	1,982	297184.4	6.7	165	193970.1	0.9	2,147	491154.4	4.4
Hypothyroidism	2,925	284010.8	10.3	733	187554.8	3.9	3,658	471565.6	7.8
Chronic pain** <sup>†</sup>	10,932	296263.2	36.9	5,152	179776.9	28.7	16,084	476040.1	33.8
	(episodes)			(episodes)			(episodes)		
Depression** <sup>†</sup>	50,132	298805.3	167.8	14,257	184720.8	77.2	64,389	483526.1	133.2
	(episodes)			(episodes)			(episodes)		

\*Cases were identified as “incident” only if the diagnosis date was >6 months after enrollment (clearance time, 2 years for depression and chronic pain) AND not in the first year when Alberta Health data was linked (washout time, 2 years for depression and chronic pain); person-years were calculated for a period between 6 months after enrollment and the date of diagnosis (or March 31, 2018, the end of data linkage).

\*\*For major cardiovascular diseases (CVD), in addition to “clearance time” and “washout time”, cases with surgery (PCI or CABG) for major CVD or diagnosis with CVD before enrollment or within 6 months of enrollment were not considered as incident cases.

<sup>†</sup>there are statistically significant differences ( $p < 0.001$ ) between women and men.

## Trend of chronic diseases by calendar year

The annual prevalence and incidence rates of chronic diseases were summarized in Supplementary Figure 2–14. All chronic diseases investigated in this study had an increased trend in the prevalence between 2001 and 2017 ( $p < 0.001$  for trend tests), among which cardiovascular diseases (Supplementary Figures 4–8) had the highest level of increase in prevalence, while hypertension (Supplementary Figure 3a) had the lowest extent of increase in prevalence annually over time. In addition, the percentage of having multi-morbidities steadily increased over time from 3.9% in 2001 to 40.3% in 2017 ( $p < 0.001$  for trend test, Figure 1).

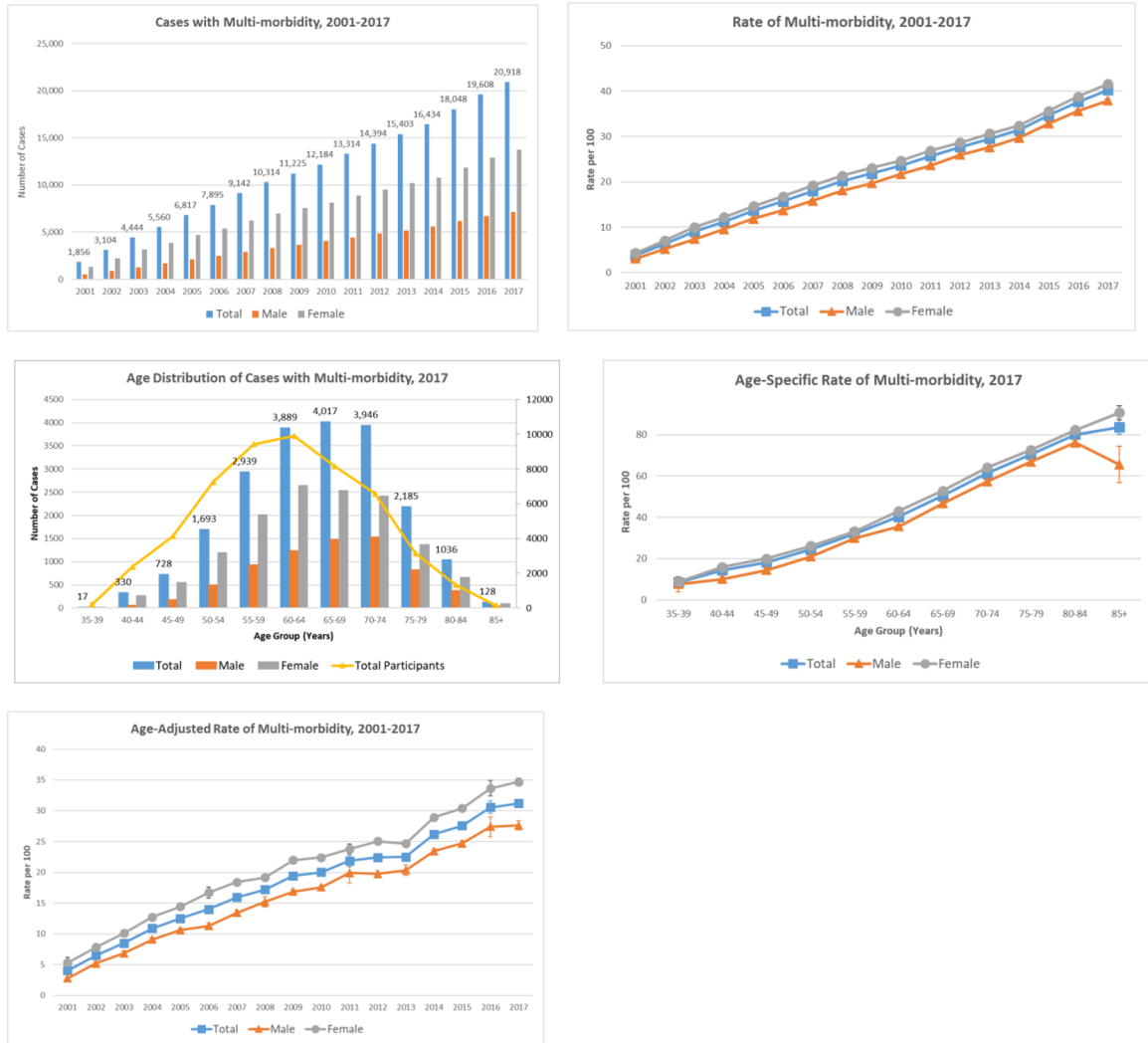
Between 2001 and 2017, the annual incidence rate (risk) of stroke (Supplementary Figure 6b) and COPD (Supplementary Figure 7b) was increased by 16.8% and 6.7% per year ( $p < 0.001$  for trend tests), respectively. In contrast, the annual incidence rate (risk) of osteoporosis was decreased by 3.2% per year over three waves between 2001–2017 (Supplementary Figure 10b). In addition, although the overall trend was decreasing, there was noticeable increase in the annual incidence rate (risk) of diabetes (Supplementary Figure 2b), hypertension (Supplementary

Figure 3b), asthma (Supplementary Figure 8b) and hypothyroidism (Supplementary Figure 11b) in 2014–2015. The annual incidence rate (risk) of acute coronary syndrome, heart failure, osteoarthritis and obstructive sleep apnea was relatively stable (changes <1% per year,  $p > 0.1$  for trend test) between 2001–2017. For conditions with multiple episodes, the annual incidence (episode) rate of depression increased by 3.3% per year ( $p < 0.001$  for trend test, Supplementary Figure 13) and decreased by 2.2% per year for chronic pain ( $p = 0.002$  for trend test, Supplementary Figure 14).

## Risk of chronic diseases, age and sex

The incidence rate (risk) of most chronic diseases was significantly increased with the age ( $p < 0.05$ ), except for asthma and depression, which either did not change with age ( $p = 0.13$ ) (asthma) or had lower risk in patients of older age ( $p < 0.001$ ) (depression) (detailed data available upon request). The regression analysis also showed that the incidence rate (risk) of diabetes, hypertension, CVD and COPD was significantly higher in men than in women ( $p < 0.01$ ), whilst it was lower in men than in women for asthma, osteoarthritis, osteoporosis, hypothyroidism, depression and

Figure 1: Rate of multi-morbidity in ATP, 2001–2017



chronic pain ( $p < 0.01$ ). In addition, the age- and sex-standardization with census data did not change the results of the trend of chronic disease rates over time, although there were noticeable differences between the actual crude and age- and sex- standardized rates (Supplementary Figure 2–14).

## Discussion

By linking provincial administrative health data to Alberta’s Tomorrow Project (ATP), a large population-based cohort study with sample size approximately 1.2% of the total population in Alberta (4.3 million in 2018), our study provides an overview of chronic disease occurrence rate and trend in prevalence and incidence in the ATP cohort. We found that of the 16 top prevalent chronic conditions, hypertension, depression, chronic pain, osteoarthritis and cardiovascular diseases were the most prevalent chronic diseases among the ATP participants when they enrolled into the cohort. All 16 chronic diseases had increased prevalence over time between 2001 and 2017, which is in part due to the positive relationship between chronic diseases and age [17–19]. Moreover, improved quality of care may relate to the increased prevalence of chronic diseases (e.g. by reducing the mortality rate of patients and/or increasing diagnosis of conditions due to better contact with healthcare systems) [20–23], which is consistent with

our observation that the prevalence rates of these 16 chronic diseases were still increased over time with adjustment of age. Nevertheless, the higher mortality rate in elderly aged 85+ years (the oldest age group in the study) may potentially lead to a lower percentage of cases surviving with the conditions of interest, which was indicated as a significant decrease in prevalence rate of major chronic conditions, including diabetes, cardiovascular diseases and COPD, in elderly aged 85+ years. The incidence rate (risk) of chronic diseases varied among conditions, with the highest rate for hypertension, osteoarthritis and ischemic heart disease. Interestingly, while the incidence rate of most chronic diseases was relatively stable over time, the percentage of multi-morbidity in the ATP cohort steadily increased, from 4% in 2001 to approximately 40% in 2017. In addition, for diabetes, hypertension, asthma and hypothyroidism, there was a noticeable increase in the incidence rate in 2015, which could be due to changes in healthcare systems (e.g. organizational structures, or clinical coding guidelines) or some unidentified factors related to the ATP cohort. However, we did not find any meaningful changes in Alberta’s healthcare systems in 2014–2015 that could potentially lead to noticeable increase in incidence rates of these conditions [24–29]. Future investigation on other factors, especially participants related factors (e.g. health seeking behaviors), is warranted.

Epidemiology studies and surveillance programs have shown the potential of using administrative health data to identify cases with chronic diseases in Canada [2–4, 8, 13, 30–34] and many other countries [35–44]. A few studies have also linked administrative health data to large population-based cohorts, to carry out disease surveillance beyond the primary self-report data possible for these cohorts [45–48]. Linking cohorts to administrative health data helps account for cohort participants over time even if they may withdraw from the study or been lost to follow-up. This may help to identify and understand how the participants who stay in the cohort might differ from those who withdraw/are lost to follow-up. It also enables a mechanism for passive data collection/follow-up overtime when administration of more direct data collection (e.g., interviews, questionnaires) over many time points is not logistically or economically feasible.

Although administrative health data in Canada are systemically collected for diagnosis, healthcare services and medication information [49], several studies have indicated that other than those prevalent chronic conditions, such as hypertension [8, 50] and diabetes [14, 51], the validity of using administrative healthcare data to identify chronic diseases varied across studies and cohorts [3, 49, 52]. In addition, given the organizational and operational nuances in the provincial healthcare systems in Canada [53], it will not be surprising to have some variations across jurisdictions in the types of healthcare data collected, the approaches to collect these data, and data available for research (e.g. cost data of drugs and medications) across jurisdictions. Therefore, in this study, we chose to use algorithms either from local studies of Alberta's populations or algorithms widely accepted by the established chronic disease surveillance programs in Alberta (or Canada if not available for Alberta) to identify 16 chronic diseases (Supplementary Figure 1). Our approach to use locally validated algorithms for chronic disease surveillance in the ATP cohort will increase the internal validity and possibly compatibility of our study with other local surveillance programs, such as the IHDA programs in Alberta, although it may not add in the external validity of our study.

Our study characterized the overall occurrence rate and trend of chronic diseases in the ATP cohort, which will significantly support the ATP as a research platform for cancer and chronic disease studies. Nevertheless, there are several limitations in our study. First, compared to the general population in Alberta, the ATP cohort is older, with more women, and participants have a relative healthy lifestyle (less likely to smoke, consumed more fruits and vegetables and more physically active) [1]. Given the fact that the ATP cohort is not representative of general populations, the surveillance data reported in this study, even with the age- and sex- standardization, should not be applied to make inference on chronic diseases status of general populations in Alberta, especially for younger populations. Secondly, although administrative health data has widely been applied as a useful data source for disease diagnosis, drug prescription and health services utilization, these datasets are not particularly designed for disease surveillance and research purposes. However, cross-linking administrative health data with other data sources, such as large population-based cohort studies, which are rich in data on sociodemographic factors, lifestyles and other modifiable risk factors, will maximize the research potential of

administrative health data. Future studies on the relationship between chronic disease occurrence, health outcomes, health services utilization and a broader range of exposures and risk factors provided by the ATP cohort (e.g. socioeconomic status and lifestyle behaviors), would benefit significantly from the administrative health data linkage outlined in this study. Lastly, due to the defined period of the linked healthcare data (October 2000–March 2018) in this study, we were unable to track retroactively healthcare records for patients diagnosed with chronic diseases prior to the date of data linkage (i.e. before October 2000). Therefore, the incidence case identified with index date in the first year of data linkage (October 2000–March 2001) might have been a prevalent case, which may lead to overestimated incidence rates, particularly for the beginning year of data linkage (i.e. in 2000/01). Without availability of the data prior to the data linkage, this overestimation cannot be accounted for properly, especially for conditions with case definition using at least two or three years of health records (e.g. asthma and hypothyroidism). As a result, cautious interpretations are recommended for any decrease in incidence rates at the beginning years of data linkage. Nevertheless, in the cumulative incident rate calculation we applied a “washout time” (i.e. the first year of data linkage) and a “clearance time” (i.e. 6 months after the ATP enrollment) to ensure the true incidence was identified (Supplementary Figure 1). This approach is especially important to obtain accurate estimation of cumulative incidence rates after the ATP enrollment, given that the ATP had a dynamic enrollment on a rolling basis from 2000 to 2015.

## Conclusion

By linking the provincial administrative healthcare data to Alberta's Tomorrow Project, we characterized the occurrence rate and trend of chronic diseases in the ATP cohort and found hypertension and depression were the most prevalent chronic diseases among the participants. Our study proved the concept and feasibility of using provincial administrative data as a complementary data source beyond the primary self-report data to identify chronic diseases in large prospective cohort studies. More importantly, the diseases surveillance approach in our study by linking two data sources together could be further applied to other large population-based cohort studies to enhance their chronic disease research capacity.

## Acknowledgments

This study is supported by funding provided to Alberta's Tomorrow Project from the Canadian Partnership Against Cancer, Alberta Cancer Foundation, Alberta Cancer Prevention Legacy Fund (administered by the Government of Alberta), Alberta Health Services and School of Public Health Bridge Funding, University of Alberta. We also thank the contributions from participants and staff of Alberta's Tomorrow Project.

This study is based in part on data provided by Alberta Health. The interpretation and conclusions contained herein are those of the researchers and do not necessarily represent the views of the Government of Alberta. Neither the Government nor Alberta Health express any opinion in relation to this study.

## Statement on conflicts of interest

The authors declared no conflict of interest.

## Ethics statement

All data used in this study were de-identified before releasing to authors for analyses. Data access and analyses of this study complied with the provincial Health Information Act (HIA) in Alberta and Alberta health (AH) and Alberta Health Services (AHS) data access procedures and data disclosure guidelines.

This study is based in part on data provided by Alberta Health. The interpretation and conclusions contained herein are those of the researchers and do not necessarily represent the views of the Government of Alberta. Neither the Government nor Alberta Health express any opinion in relation to this study.

## References

1. Ye M, Robson PJ, Eurich DT, Vena JE, Xu JY, Johnson JA. Cohort Profile: Alberta's Tomorrow Project. *Int J Epidemiol*. 2016. <https://doi.org/10.1093/ije/dyw256>
2. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical care*. 2005;43(11):1130–9. <https://doi.org/10.1097/01.mlr.0000182534.19832.83>
3. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, Alibhai A, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health services research*. 2008;43(4):1424–41. <https://doi.org/10.1111/j.1475-6773.2007.00822.x>
4. Tonelli M, Wiebe N, Fortin M, Guthrie B, Hemmelgarn BR, James MT, et al. Methods for identifying 30 chronic conditions: application to administrative data. *BMC Med Inform Decis Mak*. 2015;15:31. <https://doi.org/10.1186/s12911-015-0155-5>
5. Yu AY, Holodinsky JK, Zerna C, Svenson LW, Jette N, Quan H, et al. Use and Utility of Administrative Health Data for Stroke Research and Surveillance. *Stroke*. 2016;47(7):1946–52. <https://doi.org/10.1161/STROKEAHA.116.012390>
6. Barnabe C, Jones CA, Bernatsky S, Peschken CA, Voaklander D, Homik J, et al. Inflammatory Arthritis Prevalence and Health Services Use in the First Nations and Non-First Nations Populations of Alberta, Canada. *Arthritis Care Res (Hoboken)*. 2017;69(4):467–74. <https://doi.org/10.1002/acr.22959>. Epub 2017 Mar 9.
7. Marshall DA, Vanderby S, Barnabe C, MacDonald KV, Maxwell C, Mosher D, et al. Estimating the Burden of Osteoarthritis to Plan for the Future. *Arthritis Care Res (Hoboken)*. 2015;67(10):1379–86. <https://doi.org/10.1002/acr.22612>
8. Quan H, Khan N, Hemmelgarn BR, Tu K, Chen G, Campbell N, et al. Validation of a case definition to define hypertension using administrative data. *Hypertension*. 2009;54(6):1423–8. <https://doi.org/10.1161/HYPERTENSIONAHA.109.139279>
9. Kuan V, Denaxas S, Gonzalez-Izquierdo A, Direk K, Bhatti O, Husain S, et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digit Health*. 2019;1(2):e63–e77. [https://doi.org/10.1016/S2589-7500\(19\)30012-3](https://doi.org/10.1016/S2589-7500(19)30012-3)
10. Bryant H, Robson PJ, Ullman R, Friedenreich C, Dawe U. Population-based cohort development in Alberta, Canada: a feasibility study. *Chronic diseases in Canada*. 2006;27(2):51–9. PMID: 16867239.
11. Paula J, Robson NMS, Tiffany R, Haig, Heather K, Whelan, Jennifer E, Vena, Alianu K, Akawung, William K, Rosner, Darren R, Brenner, Christine M, Friedenreich. Cohort Profile: Design, Methods, and Demographics from Phase I of Alberta's Tomorrow Project Cohort. *Canadian Medical Association Journal*. 2016; <https://doi.org/10.9778/cmajo.20160005>
12. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. *Health services research*. 2010;45(5 Pt 2):1468–88. <https://doi.org/10.1111/j.1475-6773.2010.01142.x>
13. Public Health Agency of Canada. The Canadian Chronic Disease Surveillance System – An Overview 2018 [Available from: <https://www.canada.ca/en/public-health/services/publications/canadian-chronic-disease-surveillance-system-factsheet.html>].
14. Public Health Agency of Canada. Report from the National Diabetes Surveillance System: Diabetes in Canada, 2009. 2009. <https://www.canada.ca/en/public-health/services/reports-publications/report-national-diabetes-surveillance-system-diabetes-canada-2009.html>.
15. Focus on Geography Series, 2016 Census - Province of Alberta. Statistics Canada; 2019. <https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Facts-pr-eng.cfm?Lang=Eng&GK=PR&GC=48&TOPIC=1>
16. Robson PJ, Solbak NM, Haig TR, Whelan HK, Vena JE, Akawung AK, et al. Design, methods and demographics from phase I of Alberta's Tomorrow Project cohort: a prospective cohort profile. *CMAJ Open*. 2016;4(3):E515–E27. <https://doi.org/10.9778/cmajo.20160005>
17. Pollack RL, Morse DR. Free radicals and antioxidants: relation to chronic diseases and aging. *Int J Psychosom*. 1988;35(1–4):43–8. PMID: 3066773
18. Kennedy BK, Berger SL, Brunet A, Campisi J, Cuervo AM, Epel ES, et al. Geroscience: linking



- aging to chronic disease. *Cell*. 2014;159(4):709–13. <https://doi.org/10.1016/j.cell.2014.10.039>
19. Tracy RP. Emerging relationships of inflammation, cardiovascular disease and chronic diseases of aging. *Int J Obes Relat Metab Disord*. 2003;27 Suppl 3:S29–34. <https://doi.org/10.1038/sj.jjo.0802497>
  20. Government of Alberta. Report on Chronic Disease Management. 2014.
  21. Nolte EM, M Caring for people with chronic conditions. A health system perspective. Maidenhead, UK: Open University Press; 2008. Prevention CfDCa. National Breast and Cervical Cancer Early Detection Program (NBCCEDP) 2015 <https://bit.ly/3xdWKVw>
  22. Centers for Disease Control and Prevention. National Breast and Cervical Cancer Early Detection Program (NBCCEDP) 2015 <https://www.cdc.gov/cancer/nbccedp/index.htm>
  23. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, et al. Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. *The New England journal of medicine*. 2007;356(23):2388–98. <https://doi.org/10.1056/NEJMsa053935>
  24. Government of Alberta. Health Annual Report 2017–18. In: Health A, editor. 2018. <https://www.albertahealthservices.ca/assets/about/publications/2017-18-annual-report-web-version.pdf>
  25. Government of Alberta. Health Annual Report 2015-16. In: Health A, editor. 2016. <https://www.canada.ca/en/health-canada/services/publications/health-system-services/canada-health-act-annual-report-2015-2016.html>
  26. Government of Alberta. Health Annual Report 2016-17. In: Health A, editor. 2017. <https://open.alberta.ca/dataset/4bb6bc99-ab59-47fd-a633-dfc27d7a049e/resource/7acc0ad0-bf1b-45dd-9a41-c64f3d85db04/download/Annual-Report-2017-Health.pdf>
  27. ALBERTA HEALTH SERVICES Annual Report 2017-18. 2018. <https://open.alberta.ca/dataset/4bb6bc99-ab59-47fd-a633-dfc27d7a049e/resource/94b95989-e03a-4acc-8bf9-f080f911abf0/download/health-annual-report-2017-2018.pdf>
  28. ALBERTA HEALTH SERVICES Annual Report 2015-16. 2016. <https://www.albertahealthservices.ca/assets/about/publications/ahs-pub-2015-2016-annual-report.pdf>
  29. ALBERTA HEALTH SERVICES Annual Report 2016-17. 2017. <https://www.albertahealthservices.ca/assets/about/publications/ahs-pub-2016-2017-annual-report.pdf>
  30. Biro S, Williamson T, Leggett JA, Barber D, Morkem R, Moore K, et al. Utility of linking primary care electronic medical records with Canadian census data to study the determinants of chronic disease: an example based on socioeconomic status and obesity. *BMC Med Inform Decis Mak*. 2016;16:32. <https://doi.org/10.1186/s12911-016-0272-9>
  31. Austin PC, Stanbrook MB, Anderson GM, Newman A, Gershon AS. Comparative ability of comorbidity classification methods for administrative data to predict outcomes in patients with chronic obstructive pulmonary disease. *Ann Epidemiol*. 2012;22(12):881–7. <https://doi.org/10.1016/j.annepidem.2012.09.011>
  32. Bello A, Padwal R, Lloyd A, Hemmelgarn B, Klarenbach S, Manns B, et al. Using linked administrative data to study periprocedural mortality in obesity and chronic kidney disease (CKD). *Nephrol Dial Transplant*. 2013;28 Suppl 4:iv57–64. <https://doi.org/10.1093/ndt/gft284>
  33. Fleet JL, Dixon SN, Shariff SZ, Quinn RR, Nash DM, Harel Z, et al. Detecting chronic kidney disease in population-based administrative databases using an algorithm of hospital encounter and physician claim codes. *BMC Nephrol*. 2013;14:81. <https://doi.org/10.1186/1471-2369-14-81>
  34. Naylor KL, Garg AX, Kim SJ, Knoll GA. Epidemiology of Fracture in Adults from Ontario, Canada, with Chronic Kidney Disease: An Examination of Fracture Burden Using Administrative Health Data. *Healthc Q*. 2016;19(2):6–9. <https://doi.org/10.12927/hcq.2016.24691>
  35. Ehsani-Moghaddam B, Martin K, Queenan JA. Data quality in healthcare: A report of practical experience with the Canadian Primary Care Sentinel Surveillance Network data. *Health Inf Manag*. 2019;1833358319887743. <https://doi.org/10.1177/1833358319887743>
  36. Mikkelsen KH, Knop FK, Frost M, Hallas J, Pottegard A. Use of Antibiotics and Risk of Type 2 Diabetes: A Population-Based Case-Control Study. *J Clin Endocrinol Metab*. 2015;100(10):3633–40. <https://doi.org/10.1210/jc.2015-2696>
  37. Boursi B, Mamtani R, Haynes K, Yang YX. The effect of past antibiotic exposure on diabetes risk. *Eur J Endocrinol*. 2015;172(6):639–48. <https://doi.org/10.1530/EJE-14-1163>
  38. Wu LT, Zhu H, Ghitza UE. Multicomorbidity of chronic diseases and substance use disorders and their association with hospitalization: Results from electronic health records data. *Drug Alcohol Depend*. 2018;192:316–23. <https://doi.org/10.1016/j.drugalcdep.2018.08.013>
  39. Wang J, Bao B, Shen P, Kong G, Yang Y, Sun X, et al. Using electronic health record data to establish a chronic kidney disease surveillance system in China: protocol for the China Kidney Disease Network (CK-NET)-Yinzhou Study. *BMJ open*. 2019;9(8):e030102. <https://doi.org/10.1136/bmjopen-2019-030102>

40. Birkhead GS. Successes and Continued Challenges of Electronic Health Records for Chronic Disease Surveillance. *American journal of public health.* 2017;107(9):1365–7. <https://doi.org/10.2105/AJPH.2017.303938>
41. Figgatt M, Chen J, Capper G, Cohen S, Washington R. Chronic Disease Surveillance Using Electronic Health Records From Health Centers in a Large Urban Setting. *J Public Health Manag Pract.* 2019. <https://doi.org/10.1097/PHH.0000000000001097>
42. Klompas M, Cocoros NM, Menchaca JT, Erani D, Hafer E, Herrick B, et al. State and Local Chronic Disease Surveillance Using Electronic Health Record Systems. *American journal of public health.* 2017;107(9):1406–12. <https://doi.org/10.2105/AJPH.2017.303874>
43. Romo ML, Chan PY, Lurie-Moroni E, Perlman SE, Newton-Dame R, Thorpe LE, et al. Characterizing Adults Receiving Primary Medical Care in New York City: Implications for Using Electronic Health Records for Chronic Disease Surveillance. *Prev Chronic Dis.* 2016;13:E56. <https://doi.org/10.5888/pcd13.150500>
44. Perlman SE, McVeigh KH, Thorpe LE, Jacobson L, Greene CM, Gwynn RC. Innovations in Population Health Surveillance: Using Electronic Health Records for Chronic Disease Surveillance. *American journal of public health.* 2017;107(6):853–7. <https://doi.org/10.2105/AJPH.2017.303813>
45. Doiron D, Raina P, Fortier I, Linkage Between C, Health Care Utilization Data: Meeting of Canadian Stakeholders workshop p. Linking Canadian population health data: maximizing the potential of cohort and administrative data. *Canadian journal of public health = Revue canadienne de sante publique.* 2013;104(3):e258–61. <https://doi.org/10.17269/cjph.104.3775>
46. Mackay D, Mollard RC, Granger M, Bruce S, Blewett H, Carlberg J, et al. The Manitoba Personalized Lifestyle Research (TMPLR) study protocol: a multicentre bidirectional observational cohort study with administrative health record linkage investigating the interactions between lifestyle and health in Manitoba, Canada. *BMJ open.* 2019;9(10):e023318. <https://doi.org/10.1136/bmjopen-2018-023318>
47. Peacock A, Chiu V, Leung J, Dobbins T, Larney S, Gisev N, et al. Protocol for the Data-Linkage Alcohol Cohort Study (DACS): investigating mortality, morbidity and offending among people with an alcohol-related problem using linked administrative data. *BMJ open.* 2019;9(8):e030605. <https://doi.org/10.1136/bmjopen-2019-030605>
48. Stallmann C, Swart E, Robra BP, March S. Linking primary study data with administrative and claims data in a German cohort study on work, age, health and work participation: is there a consent bias? *Public Health.* 2017;150:9–16. <https://doi.org/10.1016/j.puhe.2017.05.001>
49. Lacasse Y, Montori VM, Maltais F. Administrative database: validity of recording vs. validity of diagnosis. *J Clin Epidemiol.* 2006;59(1):104; author reply -5. <https://doi.org/10.1016/j.jclinepi.2005.09.001>
50. Pace R, Peters T, Rahme E, Dasgupta K. Validity of Health Administrative Database Definitions for Hypertension: A Systematic Review. *The Canadian journal of cardiology.* 2017;33(8):1052–9. <https://doi.org/10.1016/j.cjca.2017.05.025>
51. Southern DA, Roberts B, Edwards A, Dean S, Norton P, Svenson LW, et al. Validity of administrative data claim-based methods for identifying individuals with diabetes at a population level. *Canadian journal of public health = Revue canadienne de sante publique.* 2010;101(1):61–4. <https://doi.org/10.1007/BF03405564>
52. Maringe C, Fowler H, Rachet B, Luque-Fernandez MA. Reproducibility, reliability and validity of population-based administrative health data for the assessment of cancer non-related comorbidities. *PLoS One.* 2017;12(3):e0172814. <https://doi.org/10.1371/journal.pone.0172814>
53. Canadian Institute for Health Information. Exploring the 70/30 Split: How Canada's Health Care System Is Financed. The Canadian Institute for Health Information; 2005. <https://www.ihe.ca/health-statistics-database/statistic/310>
54. Interactive Health Data Application Alberta Health. In: Alberta Health AaPRB, editor. 2016 [http://www.ahw.gov.ab.ca/IHDA\\_Retrieval/](http://www.ahw.gov.ab.ca/IHDA_Retrieval/)
55. Salas-Salvadó J, Martínez-González MÁ, Bulló M, Ros E. The role of diet in the prevention of type 2 diabetes. *Nutrition, metabolism, and cardiovascular diseases : NMCD.* 2011. <https://doi.org/10.1016/j.numecd.2011.03.009>
56. Muggah E, Graves E, Bennett C, Manuel DG. Ascertainment of chronic diseases using population health data: a comparison of health administrative data and patient self-report. *BMC public health.* 2013;13:16. <https://doi.org/10.1186/1471-2458-13-16>
57. Laratta CR, Tsai WH, Wick J, Pendharkar SR, Johannson KA, Ronksley PE. Validity of administrative data for identification of obstructive sleep apnea. *J Sleep Res.* 2017;26(2):132–8. <https://doi.org/10.1111/jsr.12465>
58. Leslie WD, Lix LM, Yogendran MS. Validation of a case definition for osteoporosis disease surveillance. *Osteoporos Int.* 2011;22(1):37-46. <https://doi.org/10.1007/s00198-010-1225-2>
59. Manitoba Center for Health Policy. Manitoba RHA Indicators Atlas In: Policy MCfH, editor. 2009. [http://mchp-appserv.cpe.umanitoba.ca/reference/RHA\\_Report\\_web.pdf](http://mchp-appserv.cpe.umanitoba.ca/reference/RHA_Report_web.pdf)

## Supplementary Materials

Supplementary table 1: Alberta Health (AH) data elements used to identify chronic diseases

AH dataset	Data elements	Start-end date
<b>Ambulatory Care</b>	<ul style="list-style-type: none"> <li>• Most Responsible Diagnosis (ICD-9/10) *</li> <li>• Service Start Date</li> </ul>	October 01, 2000–March 31, 2018
<b>Inpatient</b>	<ul style="list-style-type: none"> <li>• Most Responsible Health Diagnosis Code (ICD-9/10)*</li> <li>• Other Health Diagnosis Code (ICD-9/10)*</li> <li>• Services/Intervention Primary/Other (CCP/CCI)*</li> <li>• Service Start Date</li> </ul>	October 01, 2000–March 31, 2018
<b>Physician/Clinical Service Claim</b>	<ul style="list-style-type: none"> <li>• Health Diagnosis -Primary/Secondary/Tertiary (ICD9)*</li> <li>• Health Service Delivered (CCP-Alberta code)*</li> <li>• Service End Date</li> </ul>	October 01, 2000–March 31, 2018
<b>Population Registry</b>	<ul style="list-style-type: none"> <li>• Fiscal Year End Date</li> <li>• In/Out-Migration (Yes/No)</li> <li>• Age</li> <li>• Gender</li> </ul>	October 01, 2000–March 31, 2018
<b>Alberta Blue Cross</b>	<ul style="list-style-type: none"> <li>• Drug Identification Number</li> <li>• Drug Anatomical Therapeutic Chemical (ATC) Code</li> <li>• Claim Service Date</li> </ul>	October 01, 2000–March 31, 2018
<b>Pharmaceutical Information Network (PIN)</b>	<ul style="list-style-type: none"> <li>• Drug Identification Number</li> <li>• Drug Anatomical Therapeutic Chemical (ATC) Code</li> <li>• Dispensed Date</li> </ul>	April 01, 2008–March 31, 2018
<b>Vital Statistics</b>	<ul style="list-style-type: none"> <li>• Date of death</li> <li>• Cause of Death (ICD-10)</li> </ul>	October 01, 2000–March 31, 2018

\*ICD-9 and CCP codes used retroactively before 2002/2003.



Supplementary table 2: Case definition of 16 chronic conditions using Alberta Health administrative healthcare data

Conditions	Case definition algorithms		
	Case definition	Index date	ICD, CCI/CCP & ATC/DIN codes*
Hypertension	One or more hospitalizations <b>OR</b> two or more physician claims within two years (4, 54)	the earliest date of medical records that contribute to the case definition	ICD-9: 401-405 ICD-10: I10-I13, I15
Unstable angina	One or more hospitalizations (55)	the earliest date of medical records that contribute to the case definition	ICD-9: 411.X, 413.X ICD-10: I20.X, I24.X
Acute myocardial infarction	One or more hospitalizations (54, 55)	the earliest date of medical records that contribute to the case definition	ICD-9: 410.X ICD-10: I21.X, I22.X
Ischemic heart disease	One or more hospitalizations <b>OR</b> one or more physician claims (55)	the earliest date of medical records that contribute to the case definition	ICD-9: 410.X, 411.X, 412.X, 413.X, 414.X, ICD-10: I20-I25.X
Heart Failure	One or more hospitalizations <b>OR</b> one or more physician claims (54, 55)	the earliest date of medical records that contribute to the case definition	ICD-9: 428.X ICD-10: I50.X
Transient ischemic attack	One or more hospitalizations (55)	the earliest date of medical records that contribute to the case definition	ICD-9: 435.X ICD-10: G45, G45.0, G45.1, G45.2, G45.3, G45.8, G45.9, <b>NOT</b> G45.4
Acute ischemic stroke	One or more hospitalizations (55)	the earliest date of medical records that contribute to the case definition	ICD-9: 433.X, 434.X, 436.X, 362.3X ICD-10: I63.X, I64.X, H34.1; <b>NOT</b> 433.X0, 434.X0
Diabetes	One or more hospitalizations <b>OR</b> two or more physician claims within two years <b>OR</b> (ATP self-report <b>AND</b> (one or more hospitalizations <b>OR</b> one or more physician claims <b>OR</b> one or more drug prescriptions for diabetes)(14, 54, 55)	the earliest date of medical records that contribute to the case definition	ICD-9: 250 ICD-10: E10-E14 ATC: A10A, A10B
Hypothyroidism	One or more hospitalizations <b>OR</b> Two or more physician claims within two years (4)	the earliest date of medical records that contribute to the case definition	ICD-9: 240.9, 243, 244, 246.1, 246.8 ICD-10: E00-E03, E89.0
COPD	[One or more hospitalizations <b>OR</b> Two or more primary diagnoses in physician claims within two years ] <b>AND</b>	the earliest date of [1 <sup>st</sup> hospitalization date, 2 <sup>nd</sup> date of two claims] that contribute to the case definition	ICD-9: 491, 492, 496, ICD-10: J41-J44
Asthma	Age ≥ 35 years (54) One or more hospitalizations <b>OR</b> two physician claims at least 30 days apart within two years (54, 56)	the earliest date of [1 <sup>st</sup> hospitalization date, 2 <sup>nd</sup> date of two claims] that contribute to the case definition	ICD-9: 493 ICD-10: J45, J46

Continued

Supplementary table 2: Continued.

Conditions	Case Definition Algorithms		
	Case Definition	Index date	ICD, CCI/CCP & ATC/DIN codes*
Sleep apnea	One or more hospitalizations <b>OR</b> two or more physician claims within two years (57)	the earliest date of medical records that contribute to the case definition	ICD-9: 780.5 ICD-10: G47.3
Osteoarthritis	One or more hospitalizations <b>OR</b> two or more physician claims within two years <b>OR</b> two or more ambulatory care encounters within two years (7, 54)	the earliest date of medical records that contribute to the case definition	ICD-9: 715 ICD-10: M15-M19
Osteoporosis	[One or more hospitalizations <b>OR</b> One or more physician claims ] <b>AND</b> One or more BMD (Bone mineral density) tests in Claim Data within 3 years ] (54, 58)	the earliest BMD test date	ICD-9: 733.0 , or 733.0X, ICD-10: M80, M81, M82 CCP for BMD: X128
Depression	One or more hospitalizations (ICD-9/10-1) <b>OR</b> One or more physician claims (ICD-9-2) <b>OR</b> [One or hospitalizations (ICD-9-3, ICD-10-2) <b>AND</b> One or more drug prescriptions for depression within 6 months of hospitalization] <b>OR</b> [One or more physician claims (ICD-9-3) <b>AND</b> One or more drug prescriptions for depression within 6 months of claims] (59)	the earliest date of [hospitalization, claim date, drug dispense date] that contribute to the case definition within a period of 6 months (one episode) <b>AND</b> time between two index dates (episodes) $\geq 4$ months	ICD-9-1: 296.2-296.8, 300.4, 309, 311 ICD-9-2: 296, 309, 311 ICD-9-3: 300 ICD-10-1: F31, F32, F33, F34.1, F38.0, F38.1, F41.2, F43.1, F43.2, F43.8, F53.0, F93.0 ICD-10-2: F32.0, F34.1, F40, F41, F42, F44, F45.0, F45.1, F45.2, F48, F68.0, F99 ATC: N03AB02, N03AB52, N03AF01, N05AN01, N06A
Chronic pain	Two or more hospitalizations within 30 days <b>OR</b> Two or more physician claims within 30 days <b>OR</b> Two or more ambulatory care encounters within 30 days (4)	the earlier date of health records that contribute to the case definition with a period of 30 days (one episode) <b>AND</b> time between two index dates (episodes) $\geq 2$ years	ICD-9: 307.80, 307.89, 338.0,338.2, 338.4, 719.41,719.45 –719.47, 719.49,720.0, 720.2, 720.9, 721.0 –721.4, 721.6, 721.8, 721.9,722, 723.0, 723.1, 723.3 –723.9, 724.0 – 724.6, 724.70, 724.79, 724.8, 724.9, 729.0 –729.2, 729.4, 729.5 ICD-10: F45.4, M08.1, M25.50, M25.51, M25.55 – M25.57, M43.2 – M43.6, M45, M46.1, M46.3, M46.4, M46.9, M47, M48.0, M48.1, M48.8, M48.9, M50.8, M50.9, M51, M53.1 – M53.3, M53.8, M53.9, M54, M60.8, M60.9, M63.3, M79.0 – M79.2, M79.6, M79.7, M96.1

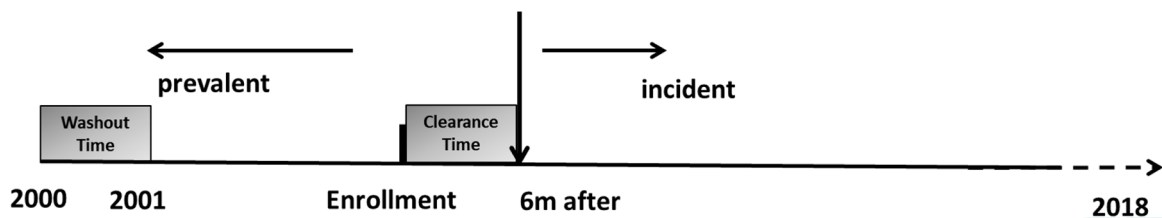
\*ICD-9 and CCP codes used retroactively before 2002/2003.

Supplementary table 3: Baseline characteristics of study participants by sex

		Women n = 33,627	Men n = 19,143	Total n = 52,770
		Mean (SD), %	Mean (SD), %	Mean (SD), %
Sex distribution		63.7	36.3	–
Age	years	50.9 (9.3)	51.7 (9.5)	51.2 (9.4)
Ethnicity	Caucasian	81.4	82.4	81.8
	Other	6.7	8.3	7.3
	Missing	11.9	9.3	10.9
Rural/urban	Rural	17.4	16.8	17.2
	Urban	82.6	83.2	82.8
Education level	High school or less	24.7	21.6	23.6
	Some post-secondary	44.0	43.1	43.7
	Post-secondary	31.2	35.2	32.6
	Missing	0.1	0.1	0.1
BMI category	Normal/under weight	27.4	14.5	22.7
	Overweight	26.5	39.0	31.0
	Obese	24.5	29.9	26.5
	Missing	21.6	16.6	19.8
Current Smoking	No	50.5	46.3	49.0
	Yes	49.2	53.4	50.7
	Missing	0.3	0.3	0.3
Physically active*	No	25.9	25.0	25.6
	Yes	22.2	25.6	23.5
	Missing	51.8	49.4	50.9
Healthy Eating Index-Canada (range 0–100)		n = 16,239	n = 9,714	n = 25,953
	Average	55.2 (9.5)	50.8 (8.9)	53.6 (9.5)

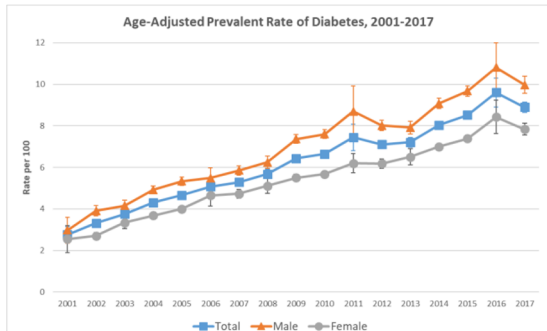
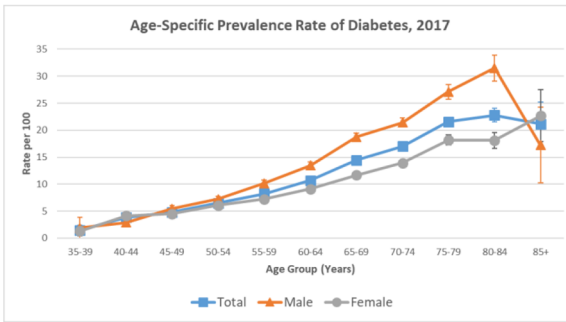
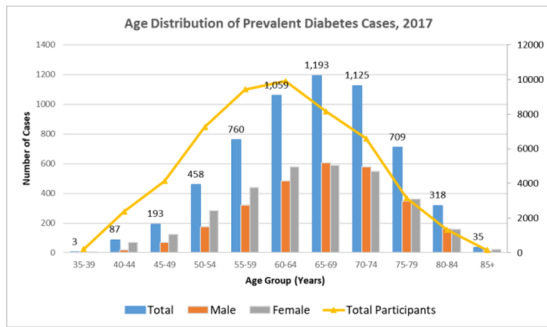
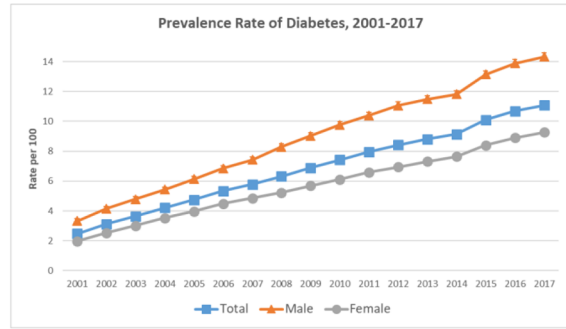
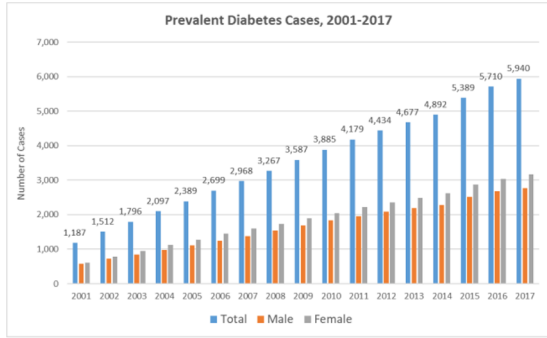
\*Categories (Yes/No) were created by whether or not participants reported accumulating at least 210 min moderate- to vigorous-intensity recreational physical activities per week in the past 12 month.

Supplementary Figure 1: The definition of “prevalent” and “incident” cases based on timing of enrollment in ATP

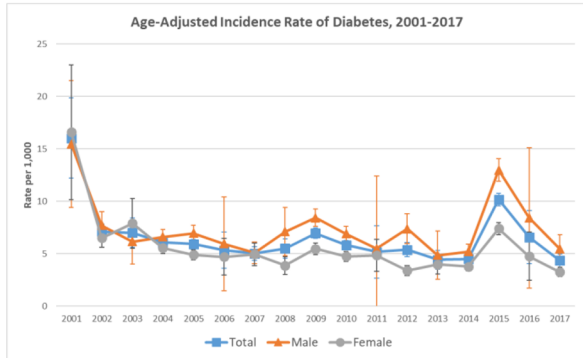
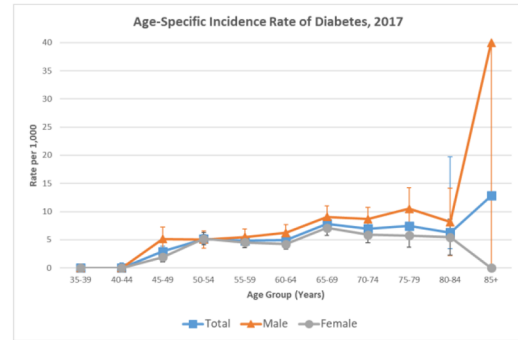
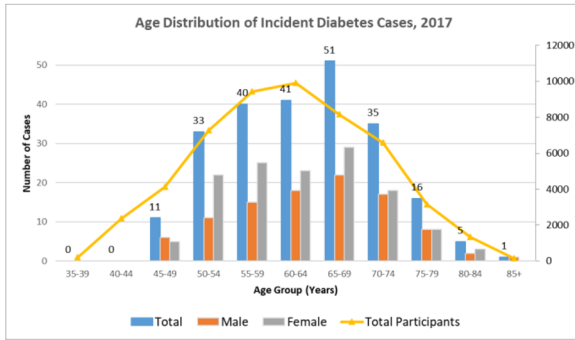
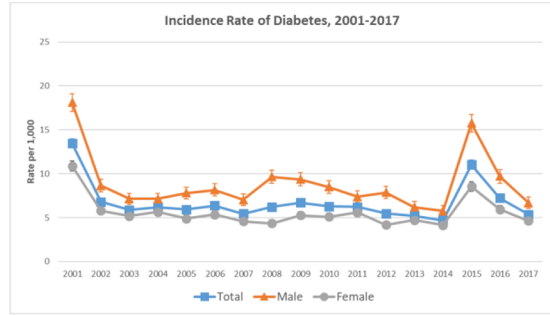
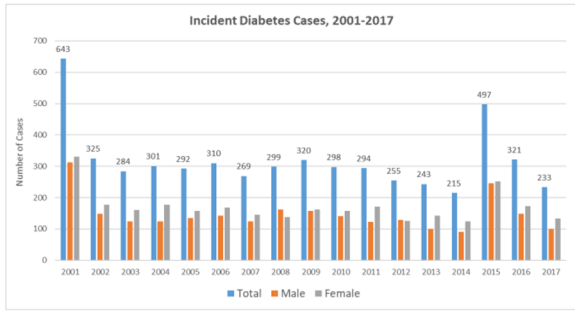


Alberta's Tomorrow Project (ATP) participants were enrolled on a rolling basis from 2000-2015. Cases with chronic diseases were identified as “non-prevalent” or “incident” cases only if the diagnosis date (index date) was >6 months after enrollment (“clearance time”) and not within the first year (October 01 2000–October 01 2001) when AH data was linked to the ATP cohort (“wash-out time”), to ensure the true “non-prevalent” or “incident” cases were identified; otherwise, cases were considered as the baseline “prevalent” cases.

Supplementary Figure 2a: Prevalence of diabetes in ATP, 2001–2017

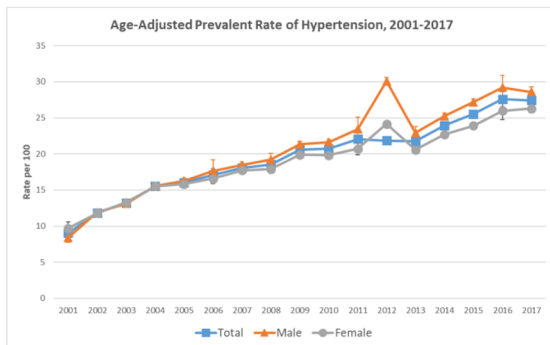
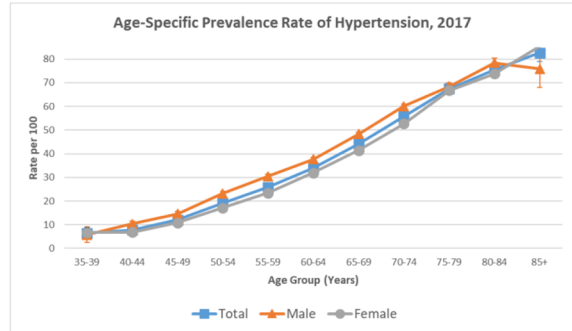
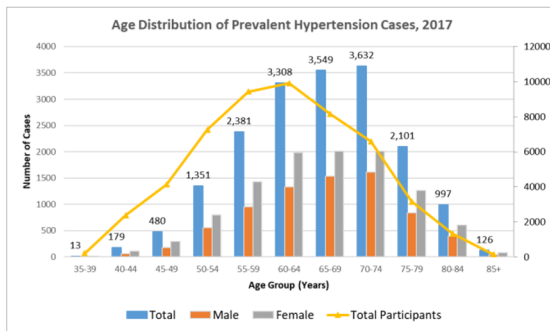
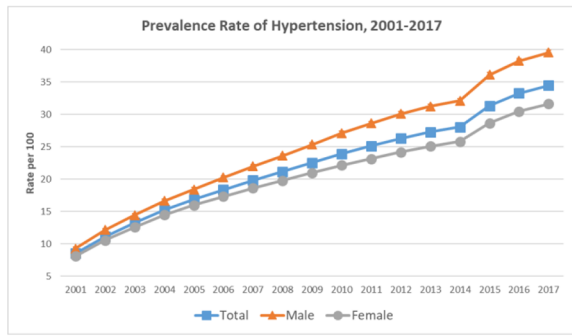
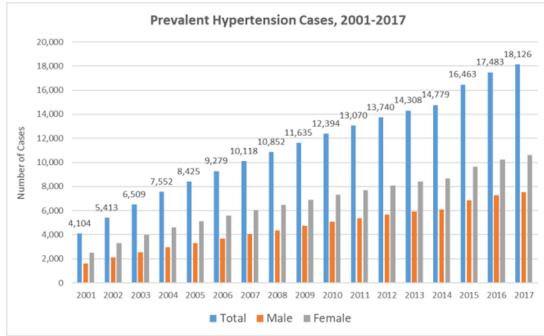


Supplementary Figure 2b: Incidence of diabetes in ATP, 2001–2017

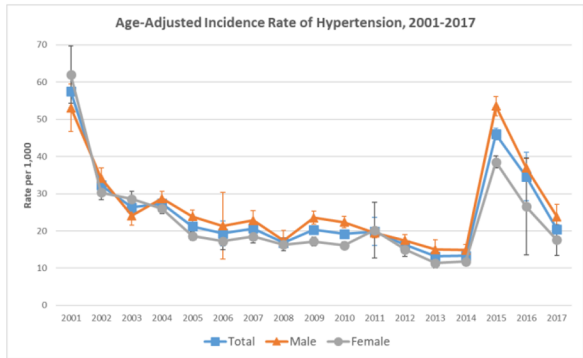
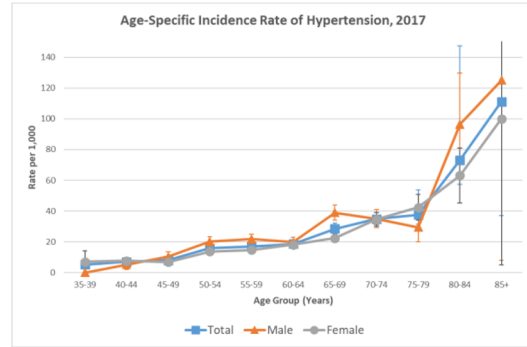
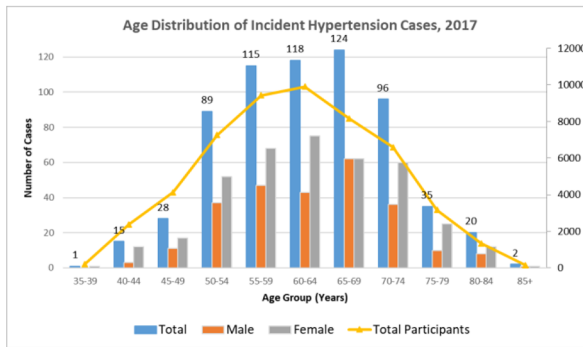
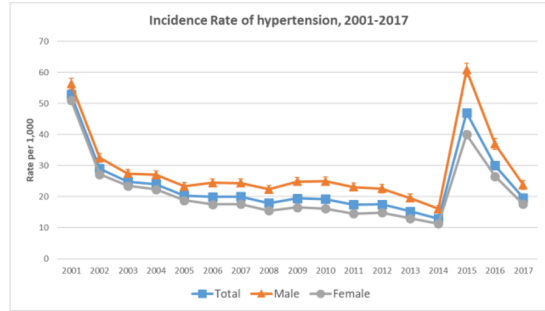
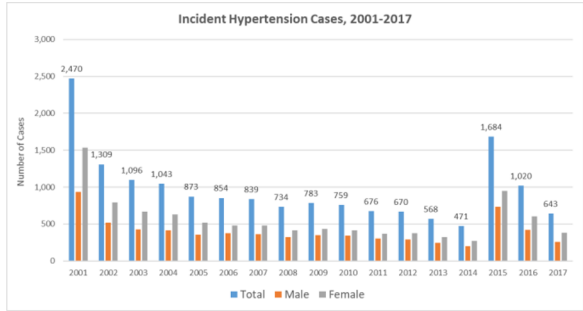




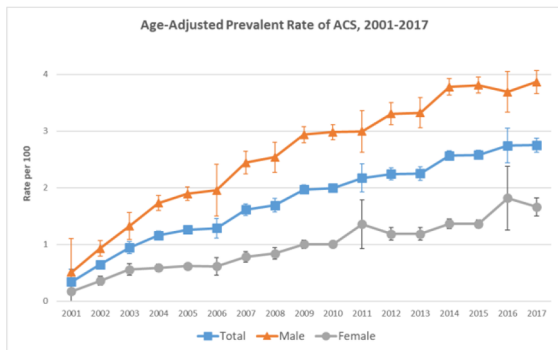
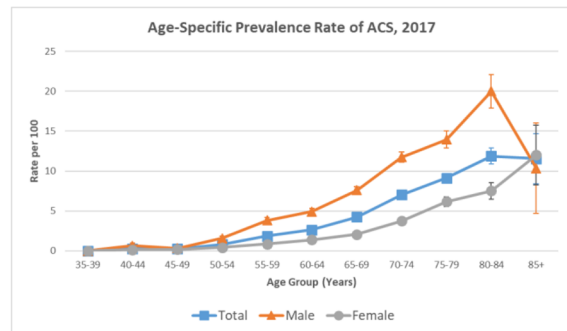
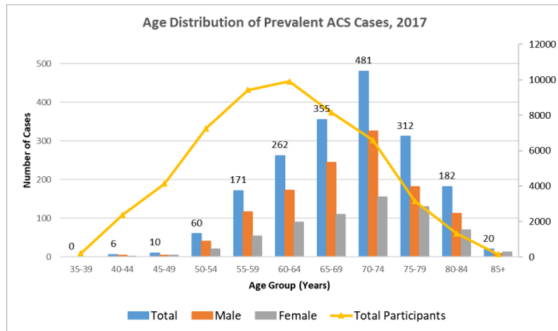
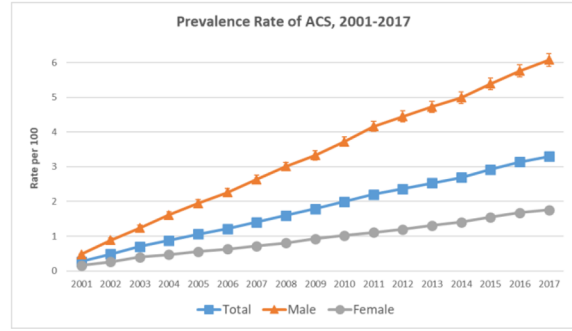
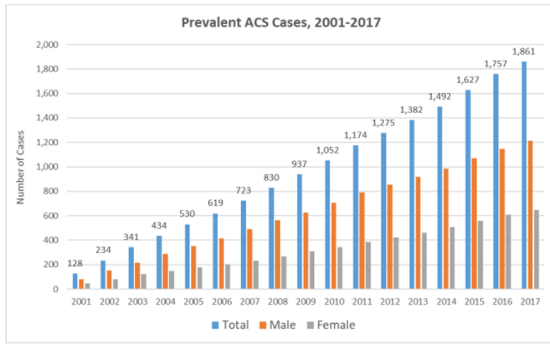
Supplementary Figure 3a: Prevalence of hypertension in ATP, 2001–2017



Supplementary Figure 3b: Incidence of hypertension in ATP, 2001–2017



Supplementary Figure 4a: Prevalence of acute coronary syndrome (ACS)\* in ATP, 2001–2017



\*Patients with unstable angina and myocardial infarction were considered as ACS cases (please refer to Supplementary Table 2 for detailed case definition using ICD codes).



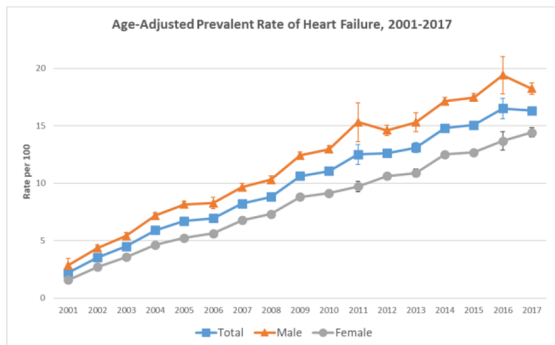
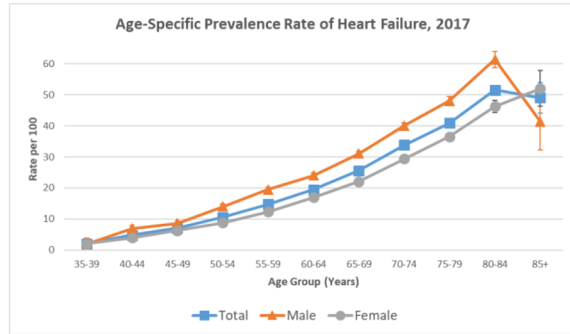
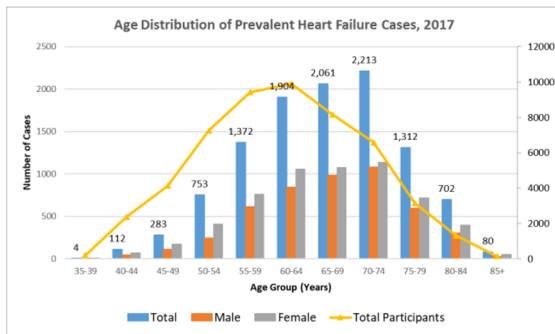
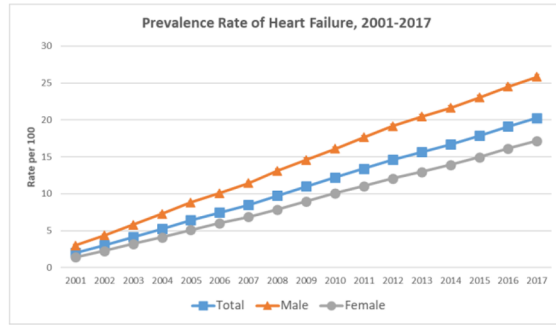
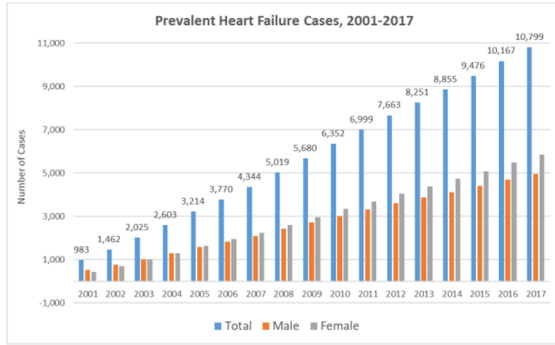
Supplementary Figure 4b: Incidence of acute coronary syndrome (ACS)\* in ATP, 2001–2017



\*Patients with unstable angina and myocardial infarction were considered as ACS cases (please refer to Supplementary Table 2 for detailed case definition using ICD codes).



Supplementary Figure 5a: Prevalence of heart failure\* in ATP, 2001–2017



\*Patients with ischemic heart disease and heart failure were considered as heart failure cases (please refer to Supplementary Table 2 for detailed case definition using ICD codes).



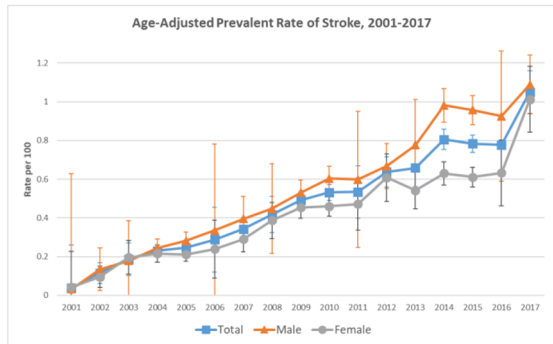
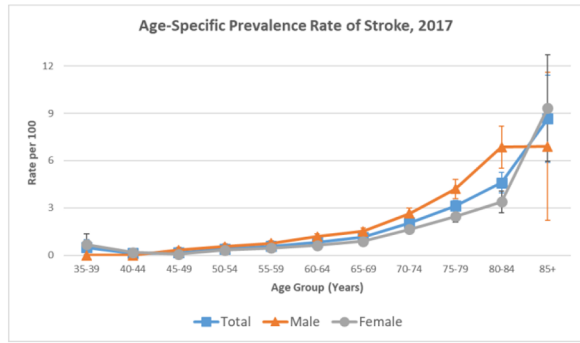
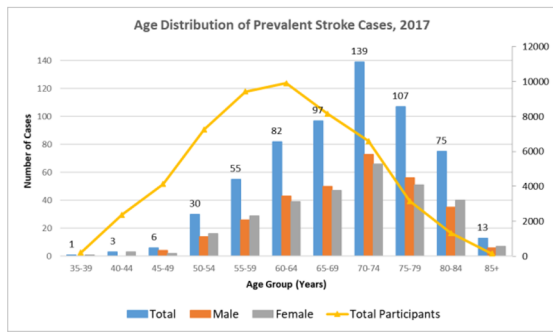
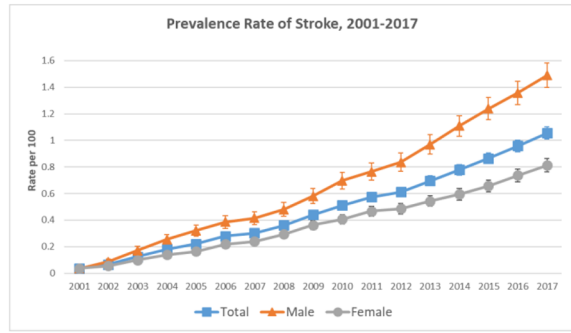
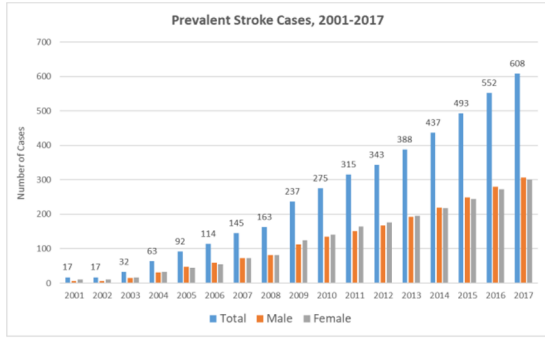
Supplementary Figure 5b: Incidence of heart failure\* in ATP, 2001–2017



\*Patients with ischemic heart disease and heart failure were considered as heart failure cases (please refer to Supplementary Table 2 for detailed case definition using ICD codes).



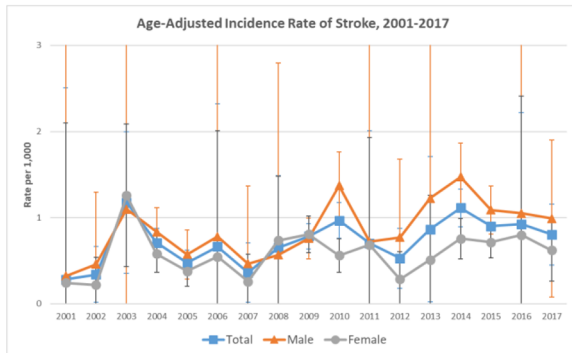
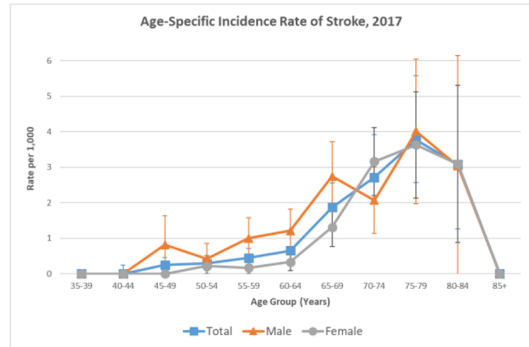
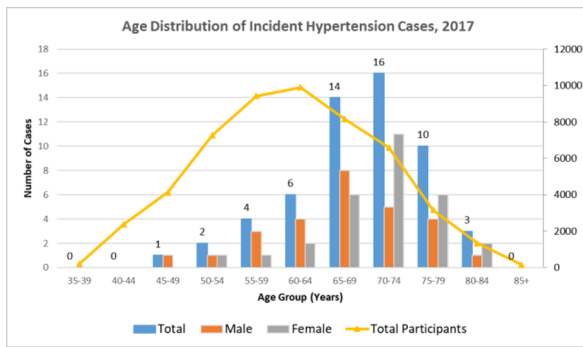
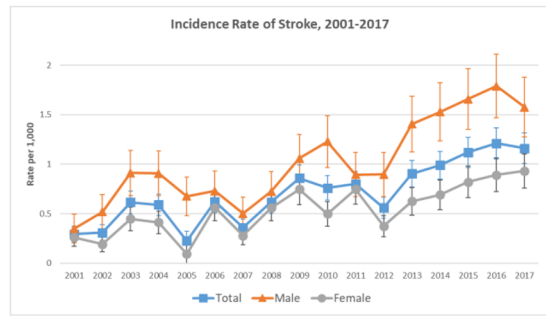
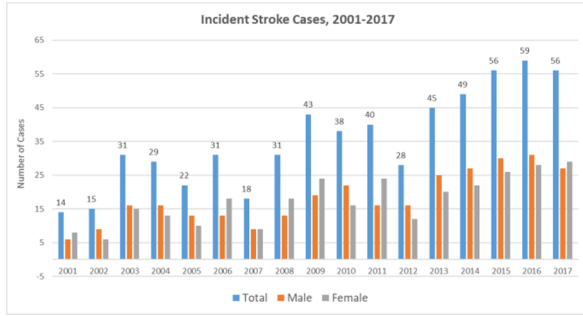
Supplementary Figure 6a: Prevalence of stroke\* in ATP, 2001–2017



\*Patients with transient ischemic attack (TIA) and acute ischemic attack (AIA) were considered as stroke cases (please refer to Supplementary Table 2 for detailed case definition using ICD codes).



Supplementary Figure 6b: Incidence of stroke in ATP\*, 2001–2017

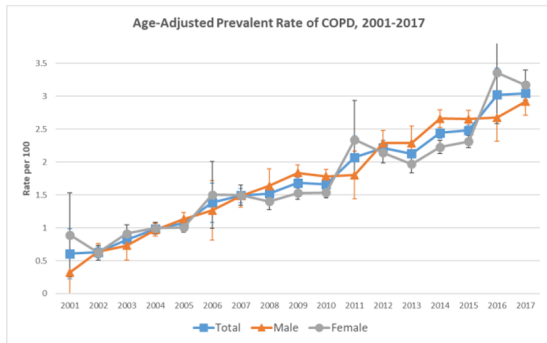
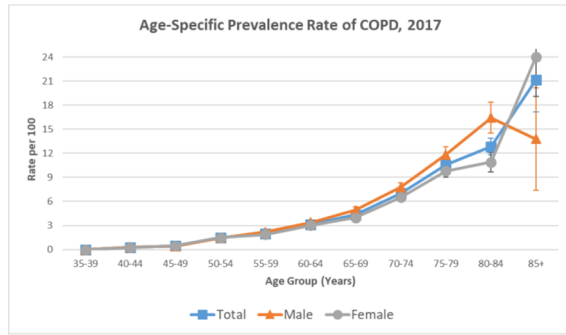
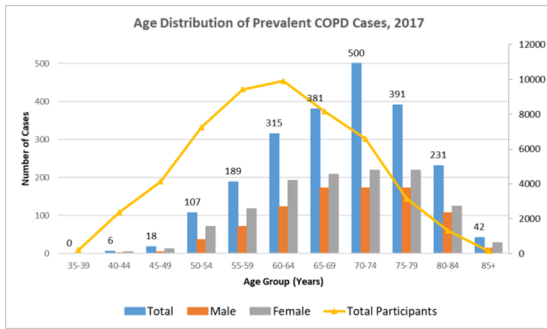
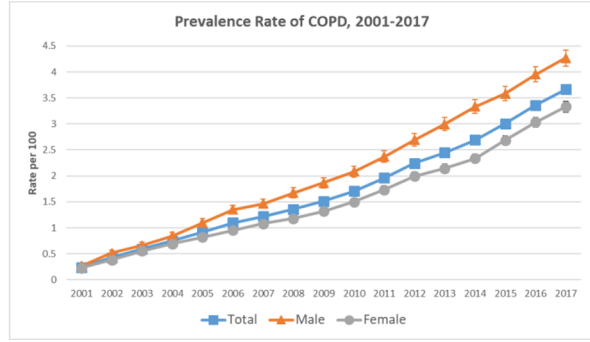
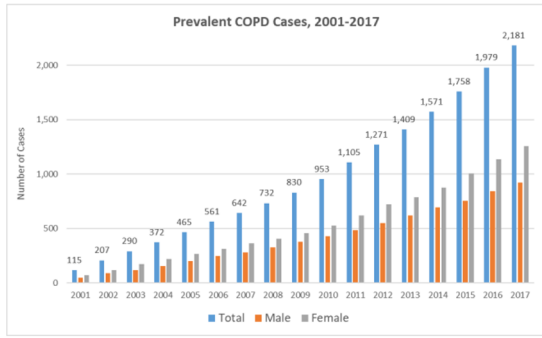


\*Patients with transient ischemic attack (TIA) and acute ischemic attack (AIA) were considered as stroke cases (please refer to Supplementary Table 2 for detailed case definition using ICD codes).

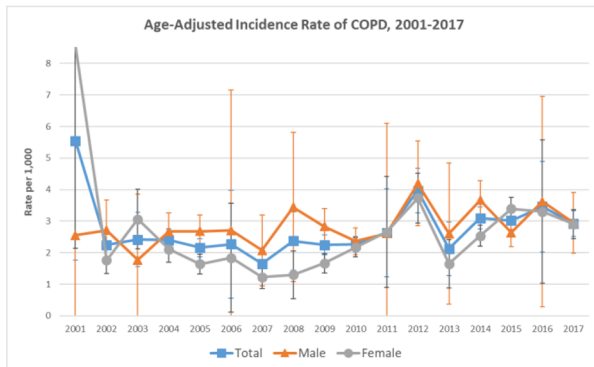
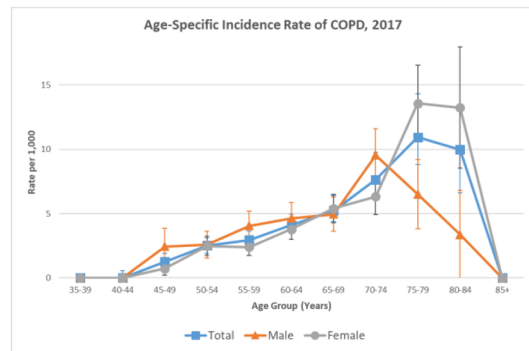
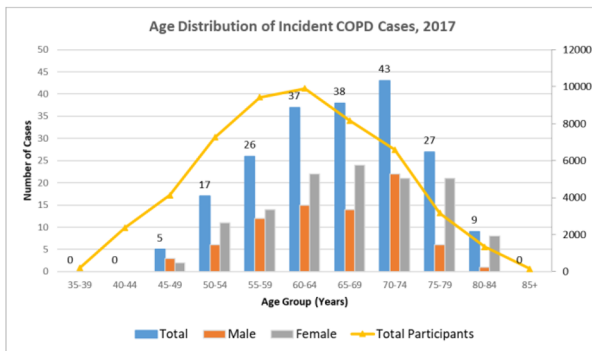
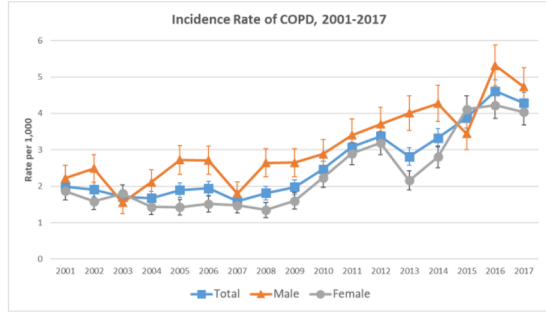
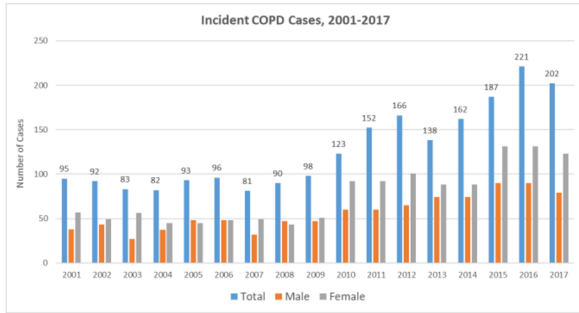




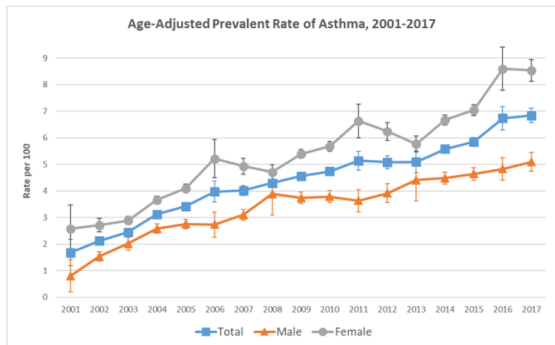
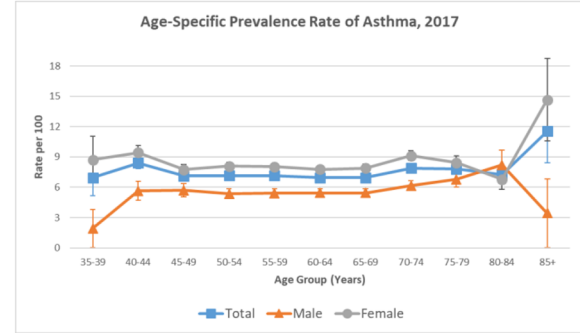
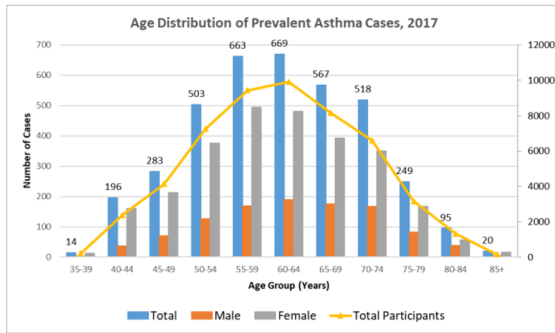
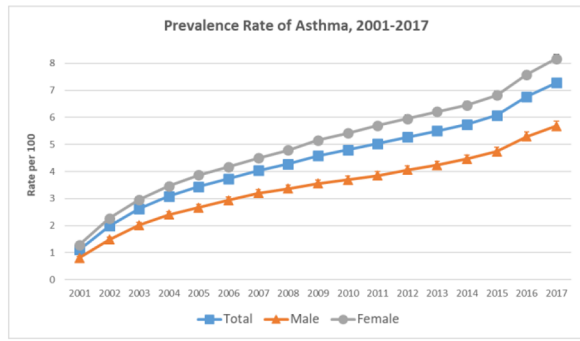
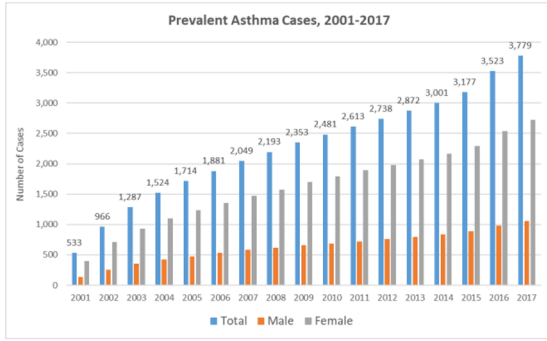
Supplementary Figure 7a: Prevalence of chronic obstructive pulmonary disease (COPD) in ATP, 2001–2017



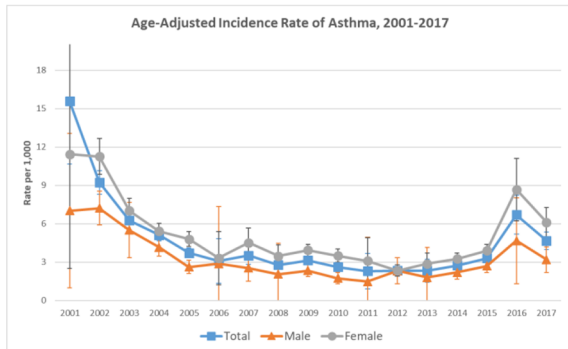
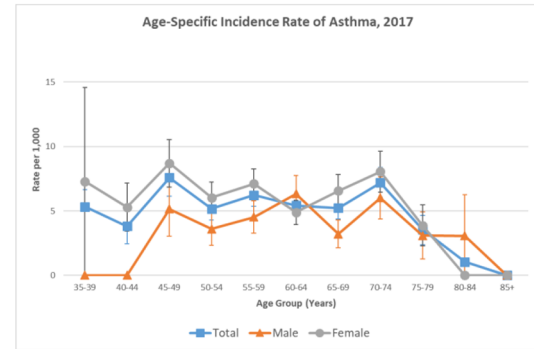
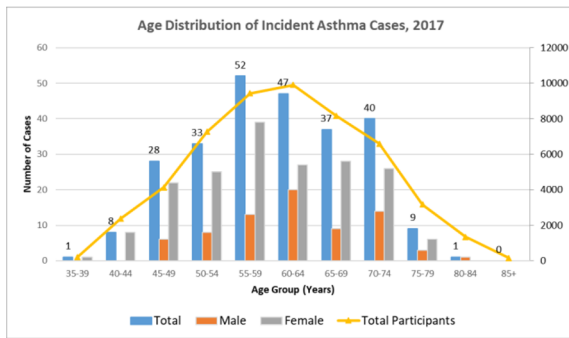
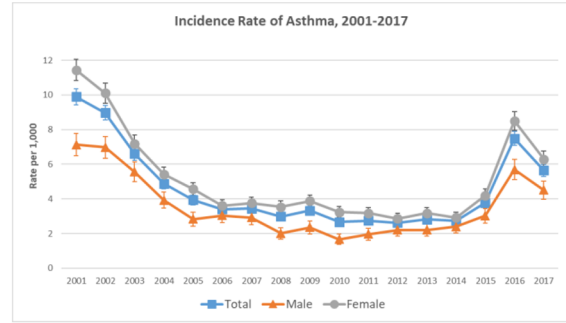
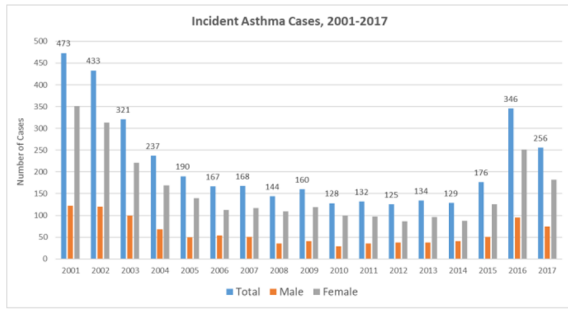
Supplementary Figure 7b: Incidence of chronic obstructive pulmonary disease (COPD) in ATP, 2001–2017



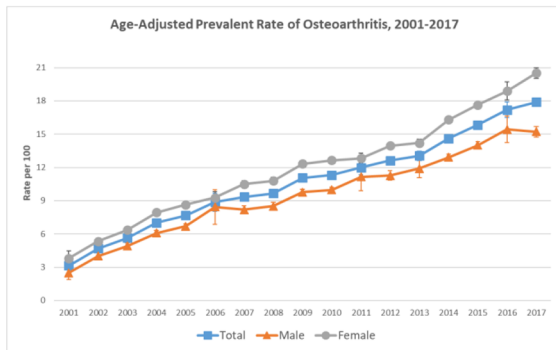
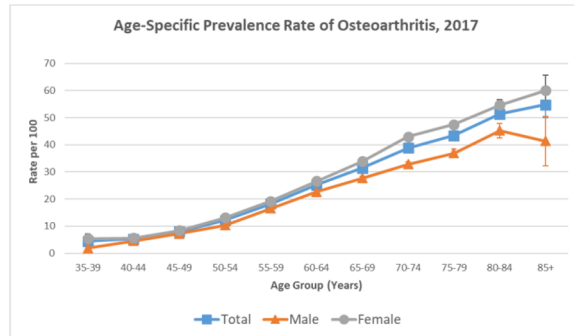
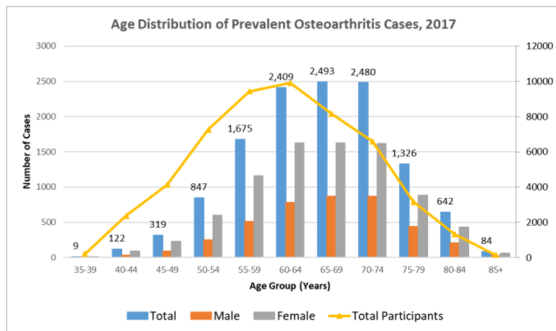
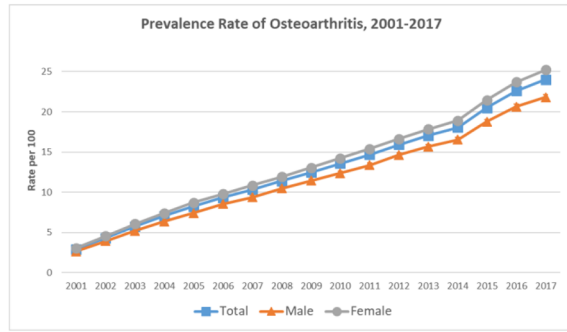
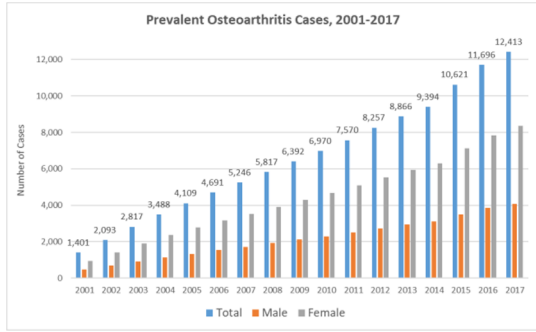
Supplementary Figure 8a: Prevalence of asthma in ATP, 2001–2017



Supplementary Figure 8b: Incidence of asthma in ATP, 2001–2017



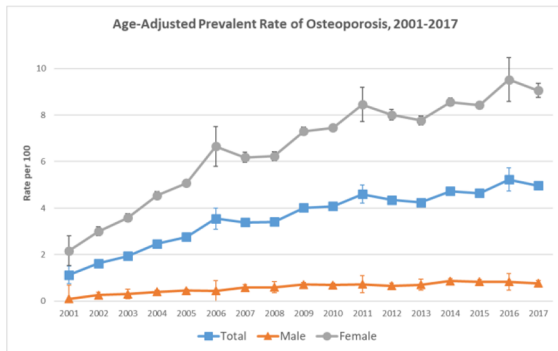
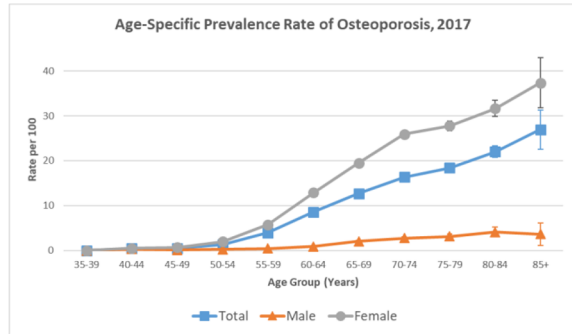
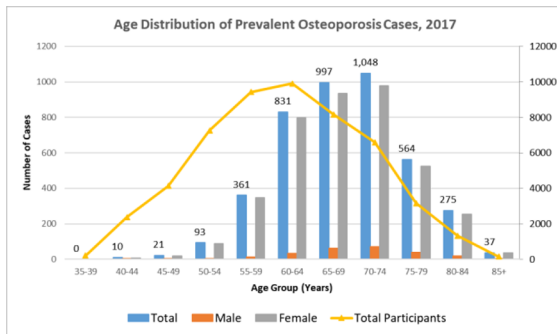
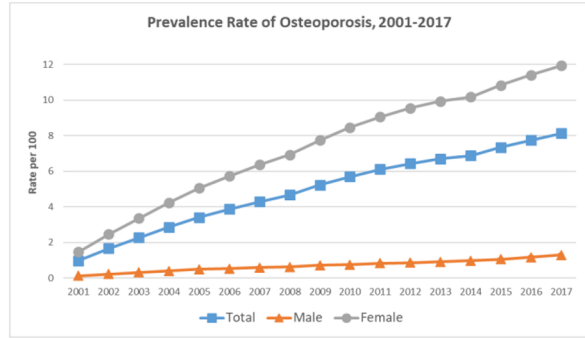
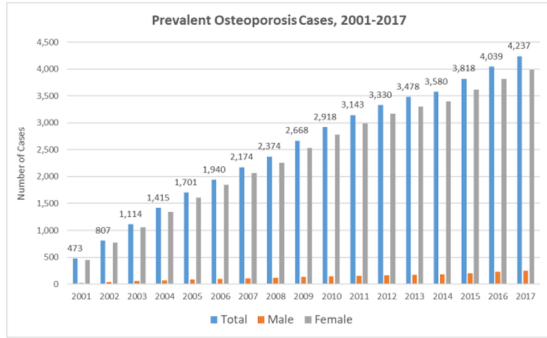
Supplementary Figure 9a: Prevalence of osteoarthritis in ATP, 2001–2017



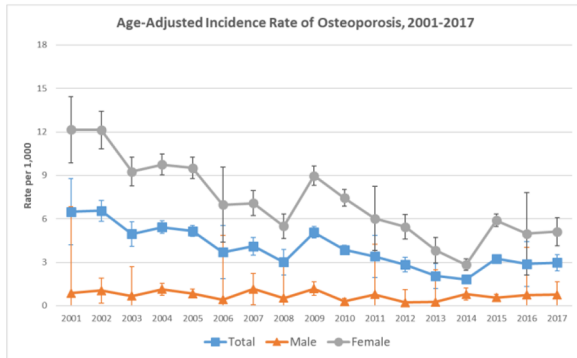
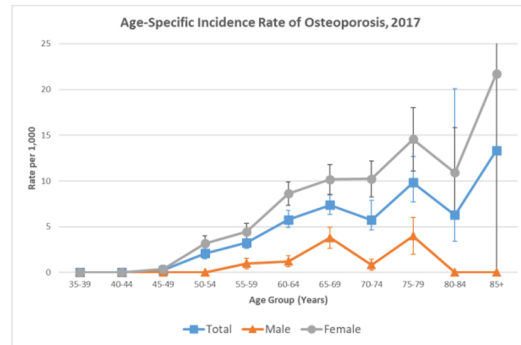
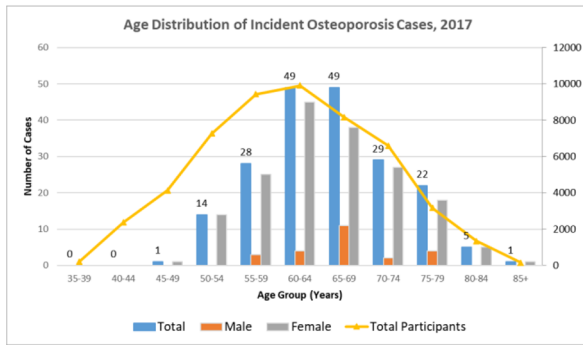
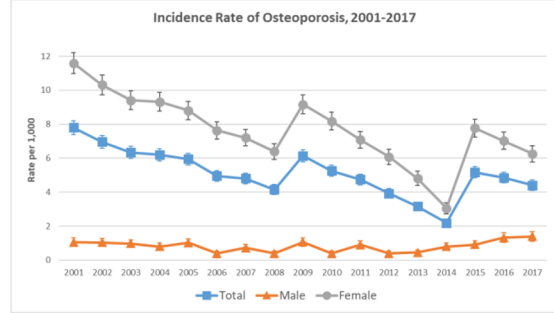
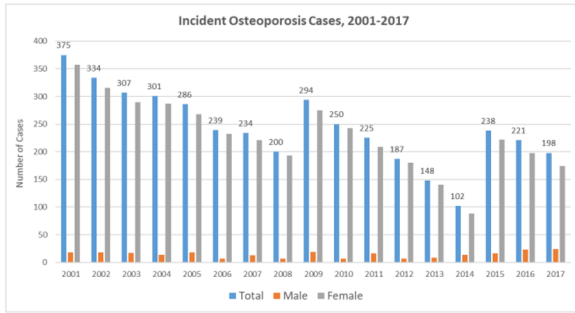
Supplementary Figure 9b: Incidence of osteoarthritis in ATP, 2001–2017



Supplementary Figure 10a: Prevalence of osteoporosis in ATP, 2001–2017

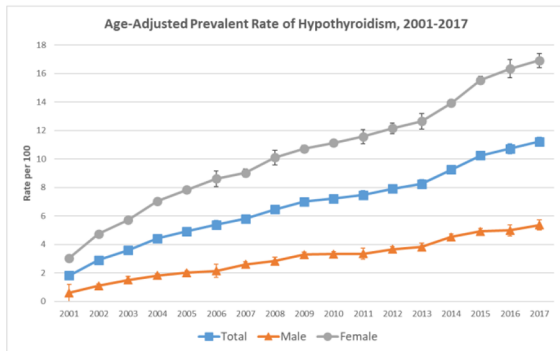
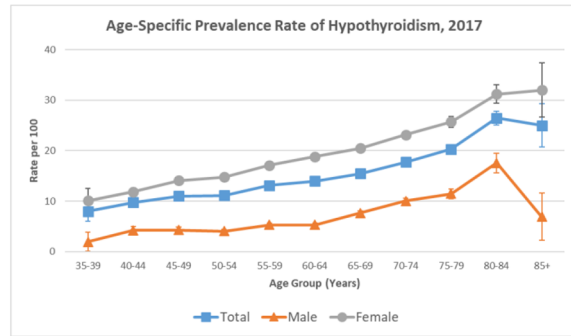
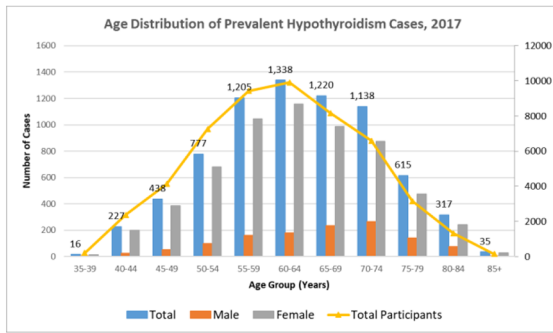
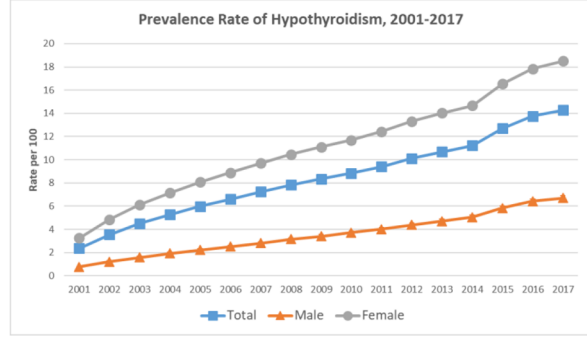
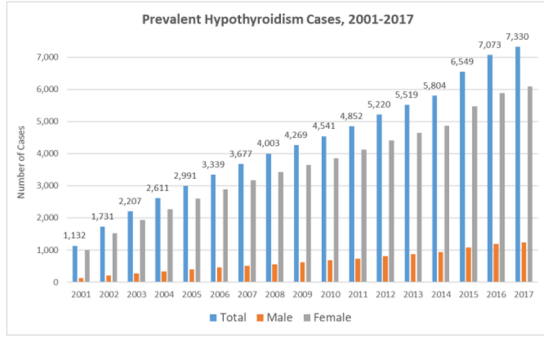


Supplementary Figure 10b: Incidence of osteoporosis in ATP, 2001–2017

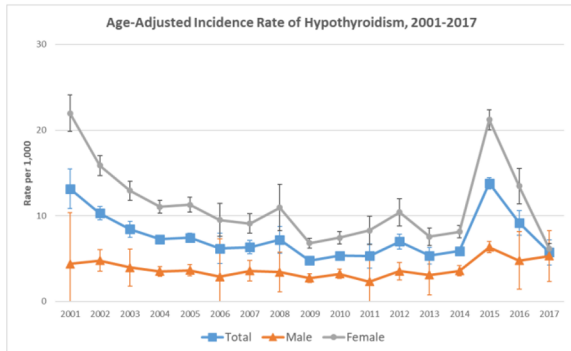
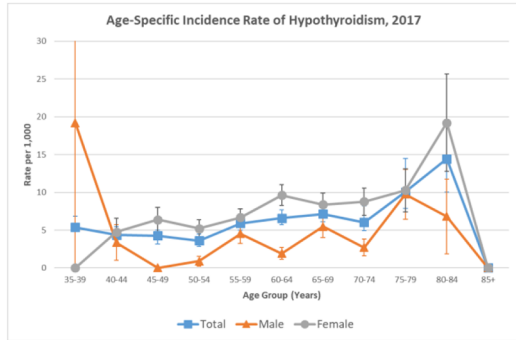
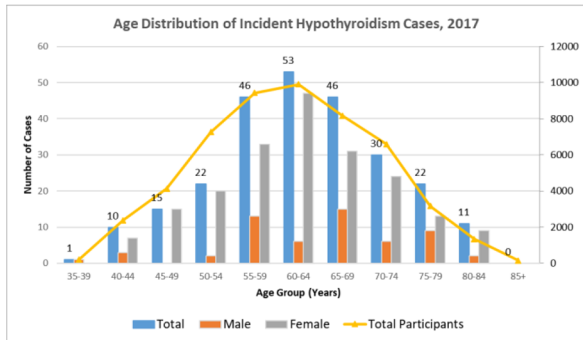
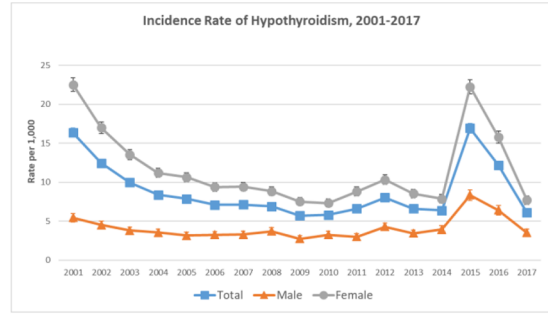
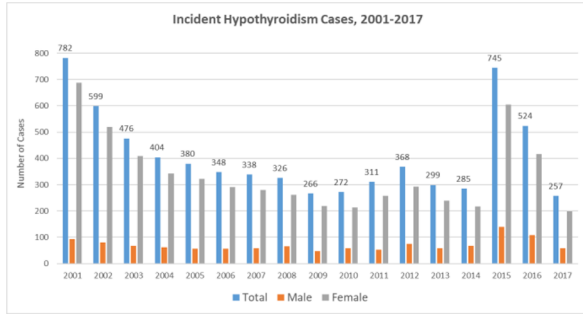




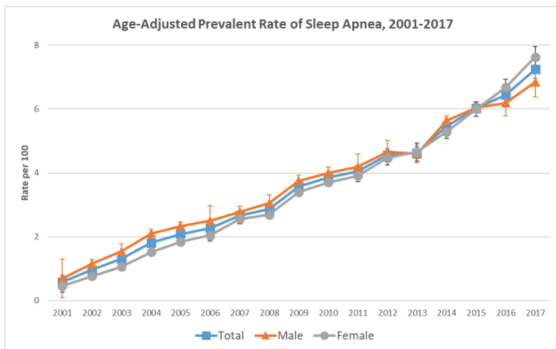
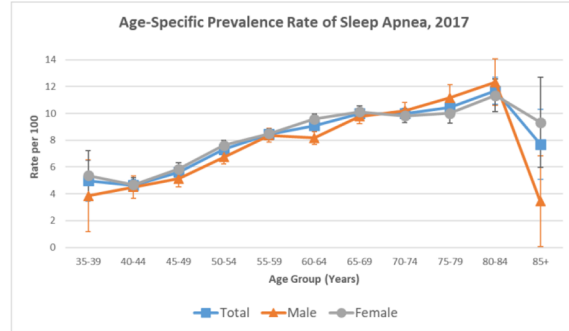
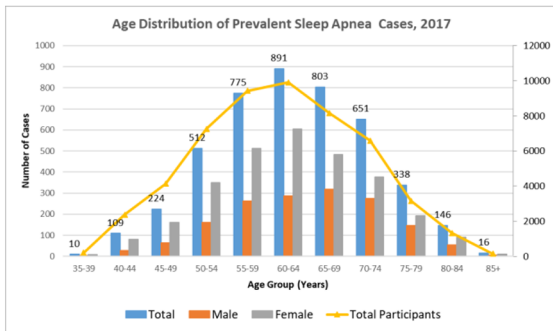
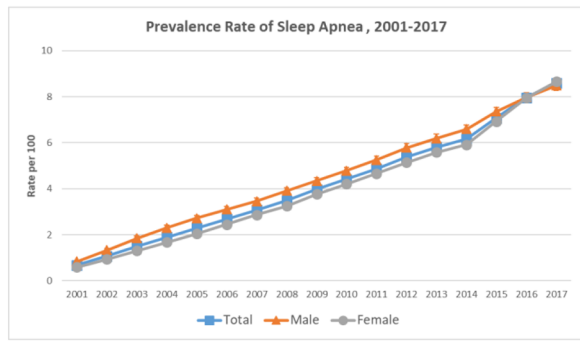
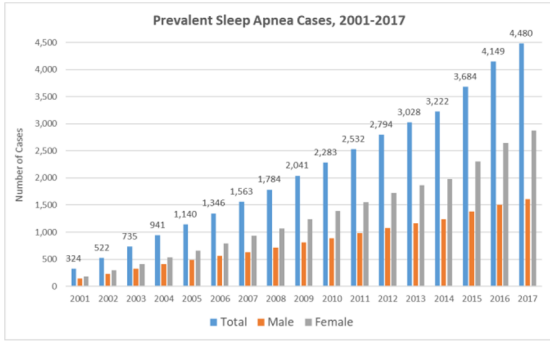
Supplementary Figure 11a: Prevalence of hypothyroidism in ATP, 2001–2017



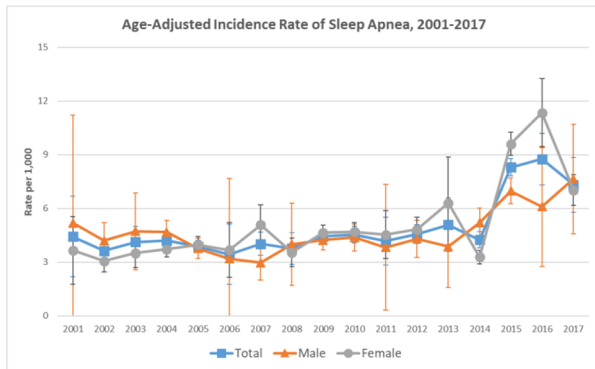
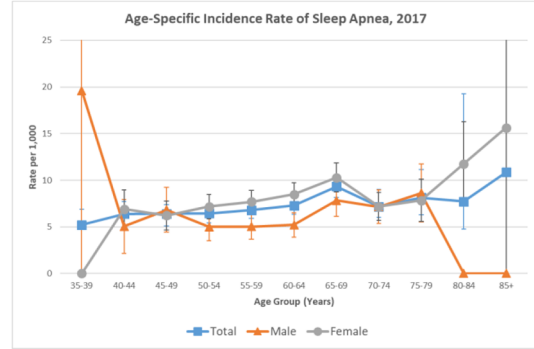
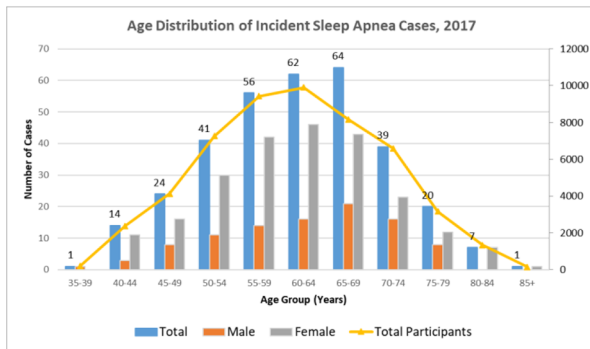
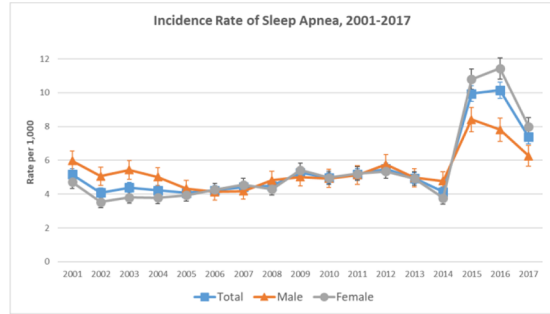
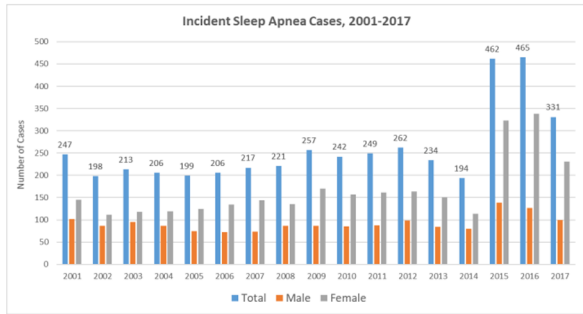
Supplementary Figure 11b: Incidence of hypothyroidism in ATP, 2001–2017



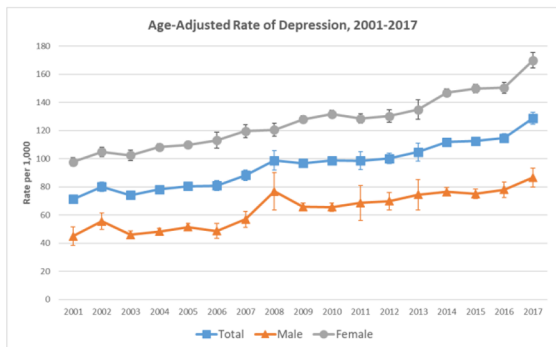
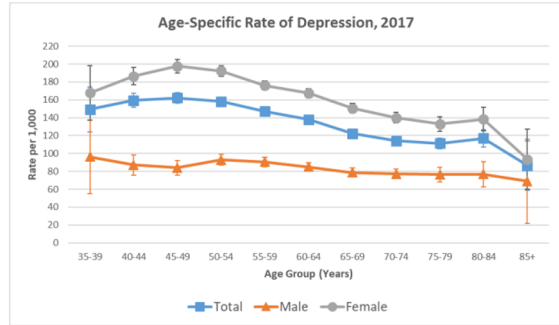
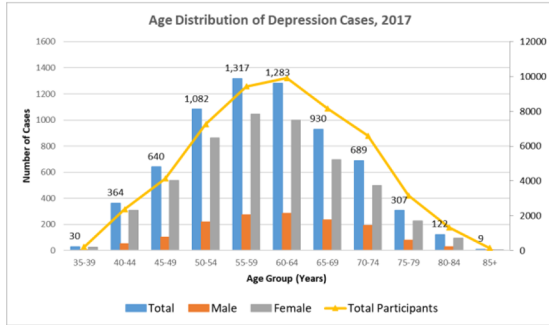
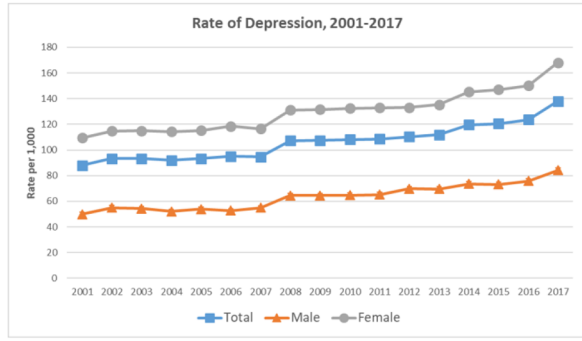
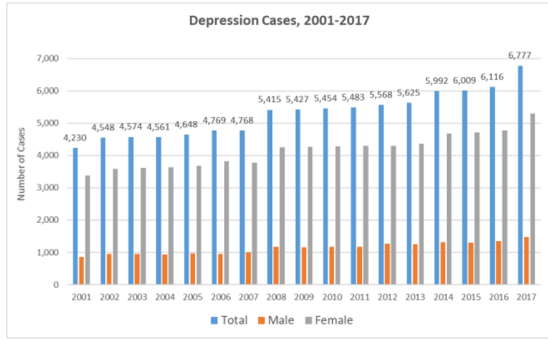
Supplementary Figure 12a: Prevalence of obstructive sleep apnea in ATP, 2001–2017



Supplementary Figure 12b: Incidence of obstructive sleep apnea in ATP, 2001–2017



Supplementary Figure 13: Rate of depression in ATP, 2001–2017



Supplementary Figure 14: Rate of chronic pain in ATP, 2001–2017

