

# Mathematical Appendix

---

## Methods

Temporal trends in health risks are often analyzed by considering age, period and cohort perspectives. Age, represented by  $a$ , captures how health risks vary by age. Period,  $p$ , or calendar year, captures changes over time, and is often the perspective of public health planning strategies. However, the range of ages in a population present in a given year represents different birth cohorts,  $c$ , that have different life experiences causing health risks patterns over their lifetime. In considering cigarette smoking exposure, cohort is a natural perspective because so much depends on what transpired at an earlier stage of life. An individual who starts to smoke cannot ever again be classified as a never smoker, and from the cohort perspective these people are part of the same group of individuals as they progress through life. Therefore, models developed here capture smoking history from the cohort perspective with potential modifications by period, e.g., due to a change in tobacco control strategy that affects all cohorts at a particular time point. However, a switch between perspectives is accomplished through the identity:

$$c = p - a .$$

### ***Number of premature deaths***

Estimates are obtained separately by gender and for simplicity of notation this distinction is not included here. Letting  $s$  represent smoking status (NS=never smoker, CS=current smoker and FS=former smoker) and  $v$  represent scenario (AC=actual tobacco control and NC=no tobacco control), the estimated prevalence is designated by  $r_{s,v}(a, p, c)$ . For the period perspective, we specify prevalences for all ages given  $p$ ,  $r_{s,v}(a, p, p - a)$  for all  $a$ . Likewise, the prevalences for cohort  $c$  would be  $r_{s,v}(a, a + c, c)$  for all  $a$ . The population for a given age group at a specified point period and cohort is  $N(a, p, c)$ .

Adapting the method of Rosenberg et al.<sup>1</sup>, we obtained estimates of the probability that an individual who was born in year  $c$  is dead at  $a + 1$  given that the individual is alive at  $a$ , designated by  $q_s(a, p, c)$  for  $a = 0, \dots, 99$  for smoking status  $s$ . Not only do we consider the separate smoking categories (CS, FS and NS), but also the smoking scenarios (AC and NC) that are derived as alternative mixtures of the smoking categories. The conditional death probability for each scenario ( $v = AC$  or  $NC$ ) is

$$q_v(a, p, c) = \sum_s r_{s,v}(a, p, c) q_s(a, p, c)$$

In the absence of exposure to cigarettes, the rates for never smokers would apply, thus the number of premature deaths in a given year for each scenario was determined by

$$\sum_{a=0}^{99} N(a, p, p - a) [q_v(a, p, p - a) - q_{NS}(a, p, p - a)] .$$

The expression in brackets is the excess death risk due to smoking (difference in death risk of current or former smokers from never smokers). Similarly, we obtain the number of premature deaths before age 65 by calculating

$$\sum_{a=0}^{64} N(a, p, p-a) [q_v(a, p, p-a) - q_{NS}(a, p, p-a)]$$

In this work, we are considering the actual tobacco control and the no tobacco control scenarios, each of which uses a different set of prevalences by age and period (and implicitly cohort). The difference in premature deaths between the actual tobacco control and the no tobacco control scenarios in a given year represents the number of premature deaths averted due to tobacco control.

### ***Years of life lost***

Years of life lost is derived from the expected number of years remaining at a given age for those with smoking status  $s$ . The conditional survival function represents the probability that an individual is alive at  $a$  given that the individual is alive at  $a^* < a$ . Considered from a cohort perspective, the conditional survival function is

$$\begin{aligned} P_s \{a, c+a, c | a^*, c+a^*, c\} &= \prod_{i=a^*}^{a-1} (1 - q_s(i, c+i, c)) \\ &= P_s \{a-1, c+a-1, c | a^*, c+a^*, c\} (1 - q_s(a-1, c+a-1, c)) \end{aligned}$$

This expression is also used to obtain the probability distribution of death at age  $a+1$ , given that an individual is alive at  $a^*$

$$\begin{aligned} f_s(a, c+a, c | a^*, c+a^*, c) &= P_s(a, c+a, c | a^*, c+a^*, c) q_s(a, c+a, c) \\ &= P_s(a, c+a, c | a^*, c+a^*, c) - P_s(a+1, c+a+1, c | a^*, c+a^*, c) \end{aligned}$$

where  $P_s(a, c+a, c | a, c+a, c) = 1$ . Assuming that deaths in a given year are uniformly distributed over the year, a death in the  $i$ -th year since  $a^*$  will have lived  $i+1/2$  years. The expected years of life remaining since  $a^*$  is given by

$$\begin{aligned} e_s(a^*, c+a^*, c) &= \sum_{i=0}^{99-a^*} (i+1/2) f_s(i+a^*, c+i+a^*, c | a^*, c+a^*, c) \\ &= \sum_{i=0}^{99-a^*} (i+1/2) [P_s(i+a^*, c+i+a^*, c | a^*, c+a^*, c) - P_s(i+a^*+1, c+i+a^*+1, c | a^*, c+a^*, c)] \\ &= \sum_{i=a^*}^{99} P_s(i, c+i, c | a^*, c+a^*, c) - 1/2 \end{aligned}$$

We note that  $P_s(a^*, c + a^*, c | a^*, c + a^*, c) = 1$  and  $P_s(100, c + 100, c | a^*, c + a^*, c) = 0$ . Total years of life lost was calculated by returning to the period perspective and cumulating the expected years of life remaining for the premature deaths, assuming these individuals had been never smokers.

$$\sum_{a=0}^{99} N(a, p, p-a) [q_v(a, p, p-a) - q_{NS}(a, p, p-a)] e_{NS}(a, p, p-a)$$

Similarly, we estimated the number of years lost before 65 by first determining expected years of life remaining before 65 in which the maximum age is constrained to be 64.5, given by

$$e_{s,65}(a^*, c + a^*, c) = \sum_{i=a^*}^{64} P_s(i, c + i, c | a^*, c + a^*, c) - 1/2.$$

The corresponding total years of life lost before 65 is

$$\sum_{a=0}^{64} N(a, p, p-a) [q_v(a, p, p-a) - q_{NS}(a, p, p-a)] e_{NS,65}(a, p, p-a).$$

The number of years lost is calculated for each tobacco control scenario,  $v=AC$  or  $NC$ . The difference in years lost between the actual tobacco control and the no tobacco control scenarios represents the number of years of life lost for premature deaths, or in positive terms the life years gained, due to tobacco control.

### ***Life expectancy***

To summarize the overall mortality experience in a year we calculated life expectancy as normally computed in a demographic analysis. This calculation is very similar to the calculation of expected years of life remaining, but it is determined from the period and not the cohort perspective. Life expectancy at birth is calculated using

$$e_s(0, p, p) = \sum_{i=0}^{99} P_s(i, p, p-i | 0, p, p) - 1/2.$$

This summary includes death rates early in life, including infant deaths that have primarily been affected by things other than cigarette smoking. Hence, our primary focus is on life expectancy at age 40, which is estimated by

$$e_s(40, p, p-40) = \sum_{i=40}^{99} P_s(i, p, p-i | 40, p, p-40) - 1/2$$

## ***All-cause mortality rates by birth-cohort and smoking status***

Rosenberg et al developed a methodology for estimating cohort life tables by smoking status for causes other than lung cancer.<sup>1</sup> We adapted this approach to obtain all-cause mortality tables by sex, birth-cohort, and smoking status: never smoker, former smoker, and current smoker. Age-specific mortality rates were obtained for the birth-cohorts of 1864-2012 through calendar year 2012. Briefly, the methodology uses 1) mortality relative risk estimates by sex and smoking status based on the large prospective CPS-I, CPS-II and the Nutrition Follow-up studies and 2) estimates of the US prevalence of never, current and former smokers by sex, age, and birth-cohort based on the National Health Interview Survey.<sup>2</sup> They also partition US mortality tables (Human Mortality Database life tables) by sex and birth-cohort into specific all-cause mortality tables for never, former and current smokers.

### ***Adjusting population distribution***

The age distribution of the population would be changed by the modification of death rates due to cumulative effects of changing cigarette smoking behavior. For example, without tobacco control, there would be fewer people living to later ages, thereby reducing the overall population. Following a cohort in time, we have

$$N(a+1, a+c+1, c) = N(a, a+c, c) [1 - q_v(a, a+c, c)] + \Delta(a, a+c, c)$$

represents change in the population size that does not result from mortality, i.e., migration or unknown discrepancy between our death rates and those used by the US Census. We refer to this as migration but recognize that it may also encompass other factors that we cannot identify, and it is estimated by

$$\Delta(a, a+c, c) = N(a+1, a+c+1, c) - N(a, a+c, c) [1 - q_{AC}(a, a+c, c)]$$

Assuming that this aspect of change would not have been affected by tobacco control, we estimate population under the no control scenario by

$$N_{NC}(a+1, a+c+1, c) = N_{NC}(a, a+c, c) [1 - q_{NC}(a, a+c, c)] + \Delta(a, a+c, c)$$

Beginning with the population in 1964, we apply this approach sequentially for subsequent years.

## **Results Not Included in the Main Text**

Figure 2 in the main text shows trends for overall prevalence of current smokers by period. Figures A1-A3 show further detail on these results by providing trends for age-specific prevalence of current, former and never smokers, respectively. Gender-, smoking status- and age-specific death rates ( $q_s(a, p, c)$  for  $s=NS, FS, CS$ ) for selected calendar years are shown in Figure A4.

The age- and smoking status-specific death rates,  $q_s(a, p, c)$ , and the smoking prevalence distributions,  $r_{s,v}(a, p, c)$ , are used to estimate age-specific all cause-mortality rates,  $q_v(a, p, c)$ , under the actual and primary counterfactual scenarios, which consists of mixture of three smoking status, and the results are shown in Figure A5. The higher mortality rates in the primary counterfactual scenario for ages 40-85 are due to the greater prevalence of current and former smokers (Figures A1-A3), as well as their correspondingly higher mortality rates shown in Figure A4. Change in the population structure that would be expected since 1964 would have been gradual and cumulative. Figure A6 compares the population pyramid in 2012 with what would be expected under the primary counterfactual scenario. The textured bars represent individuals who would not have been alive in 2012 in the absence of tobacco control.

In Table 1 in the main text we show the number of premature deaths in intervals of years following 1964. Figure A7 provides a graphical display of the same set of results but for single calendar years. This figure also shows results for the alternative counterfactuals showing the range of values that correspond to the limits, and the estimates for never smokers. Part (a) shows the trends for all premature deaths and (b) gives the corresponding results for premature deaths before age 65.

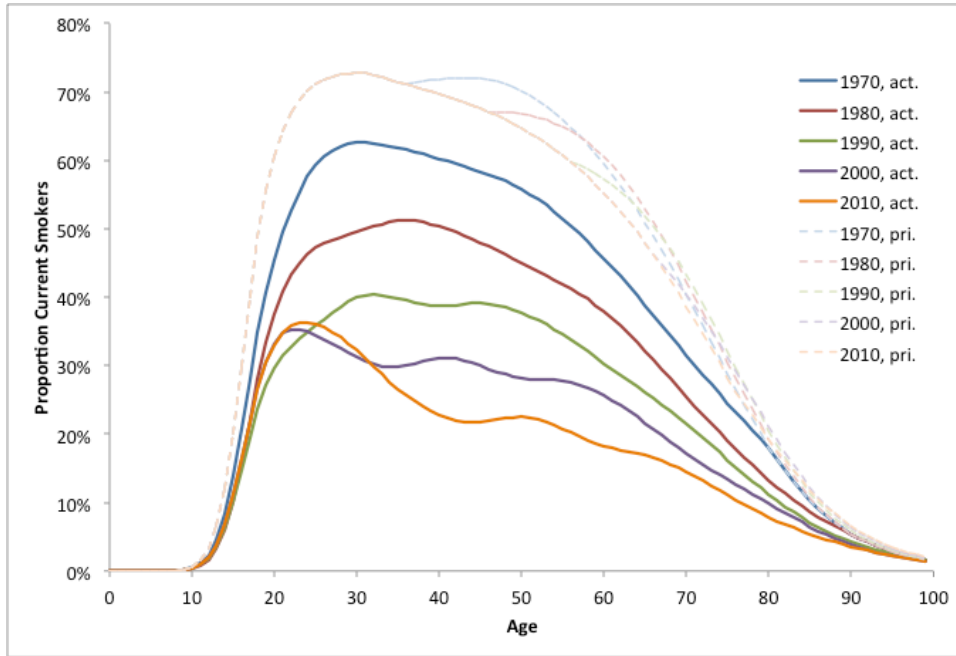
Figure A8 shows the estimated years of life lost each year due to premature deaths from cigarette smoking for the actual tobacco control and the counterfactual scenarios. The total of all years is shown in part (a) and the total for just the years before age 65 is shown in part (b).

Figure 3 of the paper shows life expectancy at age 40. Life expectancy at ages 50 and 60 are shown in Figure A9. The changes are gradual, although at older ages the differences become small and they approach never smokers because the prevalence of never smokers tends to become predominant as the smokers either quit or die.

## References

1. Rosenberg MA, Feuer EJ, Yu B, et al. Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Analysis*. 2012;32:S25-S38.
2. Holford TR, Levy DT, McKay LA, et al. Smoking histories for the United States, 1965-2009: Tobacco exposure parameters for risk estimation. *Submitted for publication*. 2013.

(a) Males



(b) Females

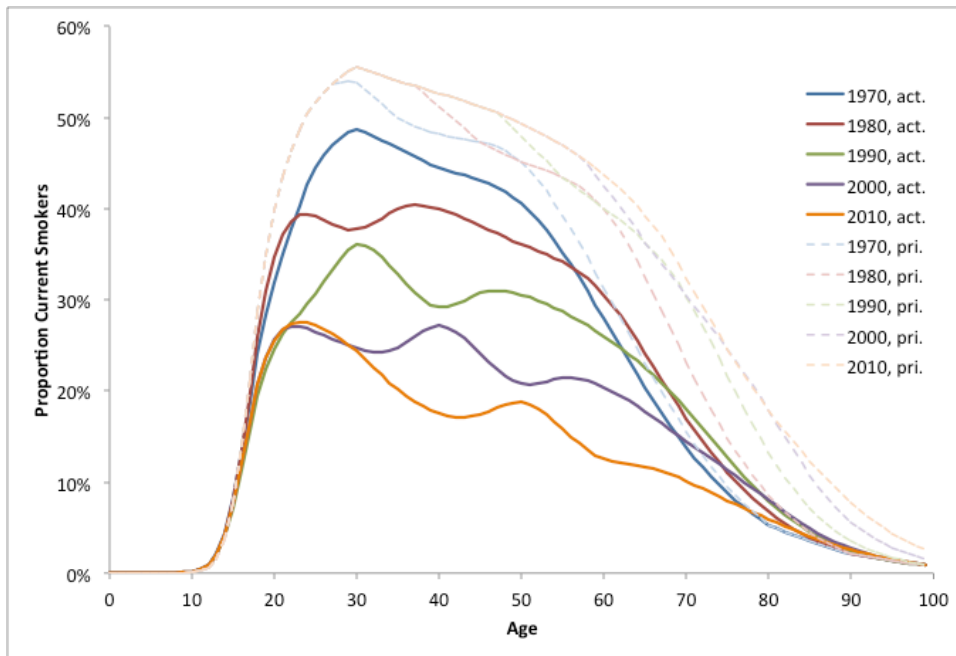
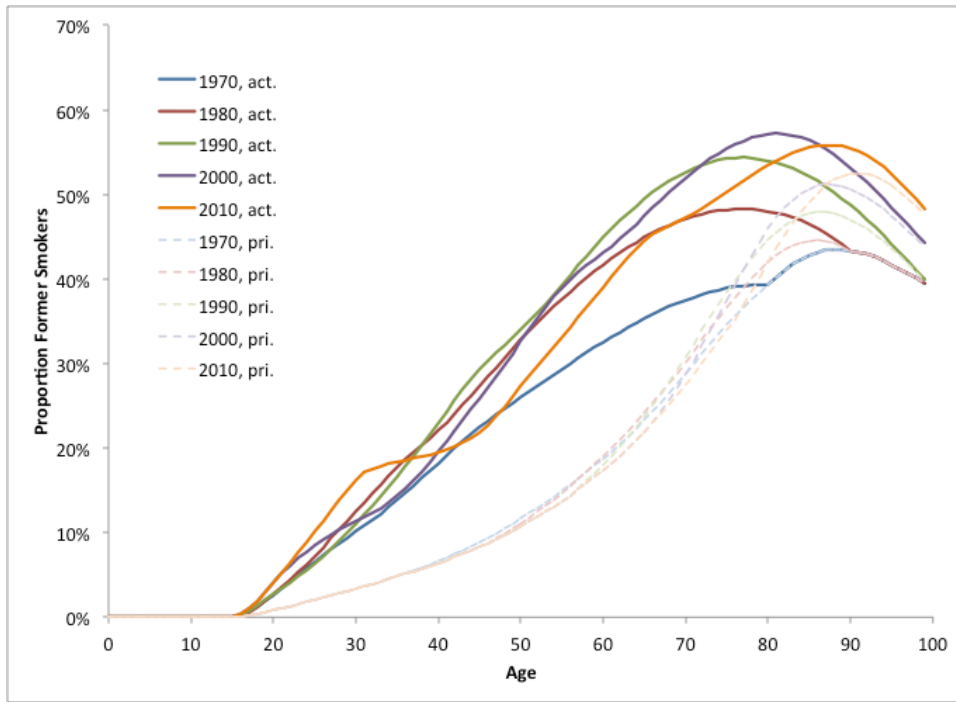


Figure A1. Observed (solid) and primary counterfactual (dashed) proportion of current smokers by age and calendar year.

(a) Males



(b) Females

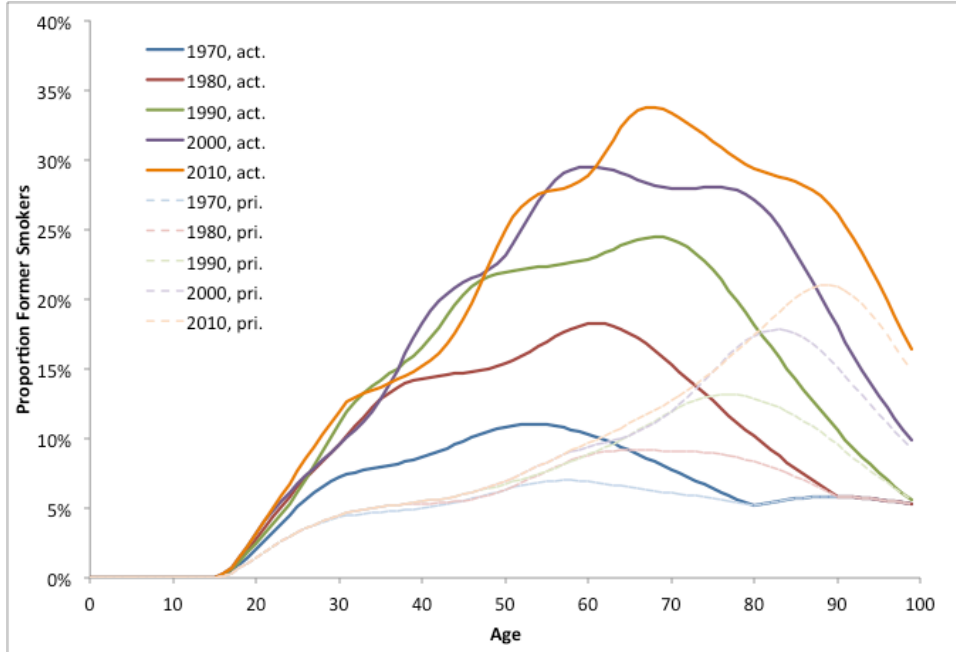
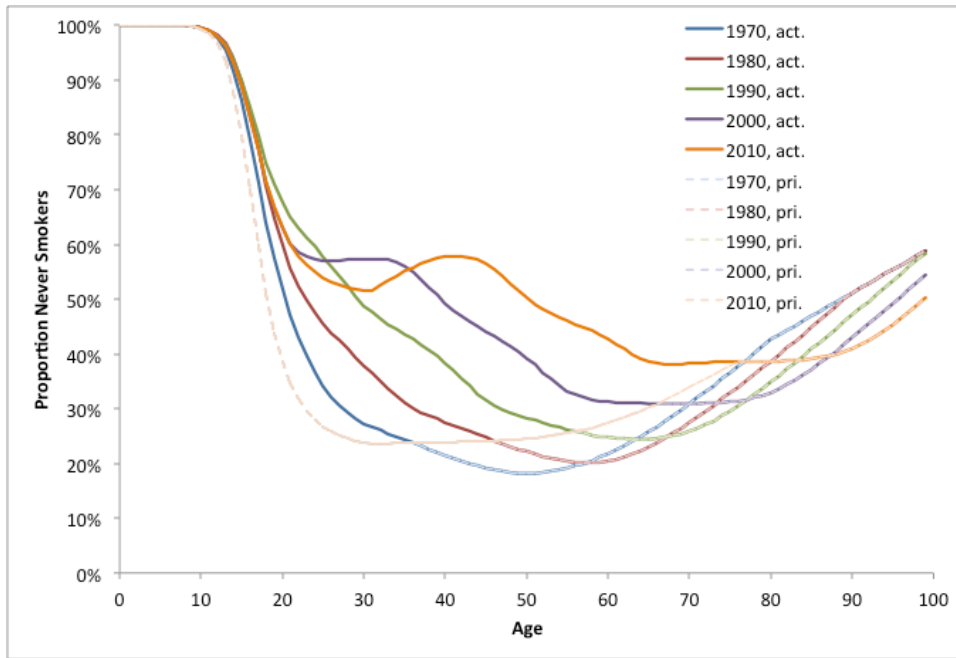


Figure A2. Observed (solid) and primary counterfactual (dashed) proportion of former smokers by age and calendar year.



(a) Males



(b) Females

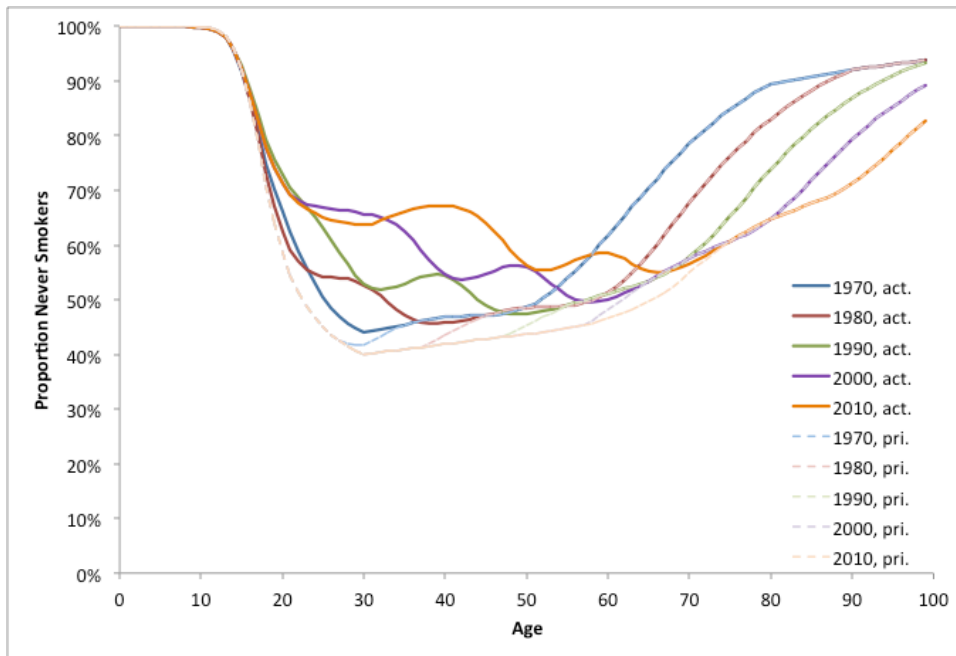
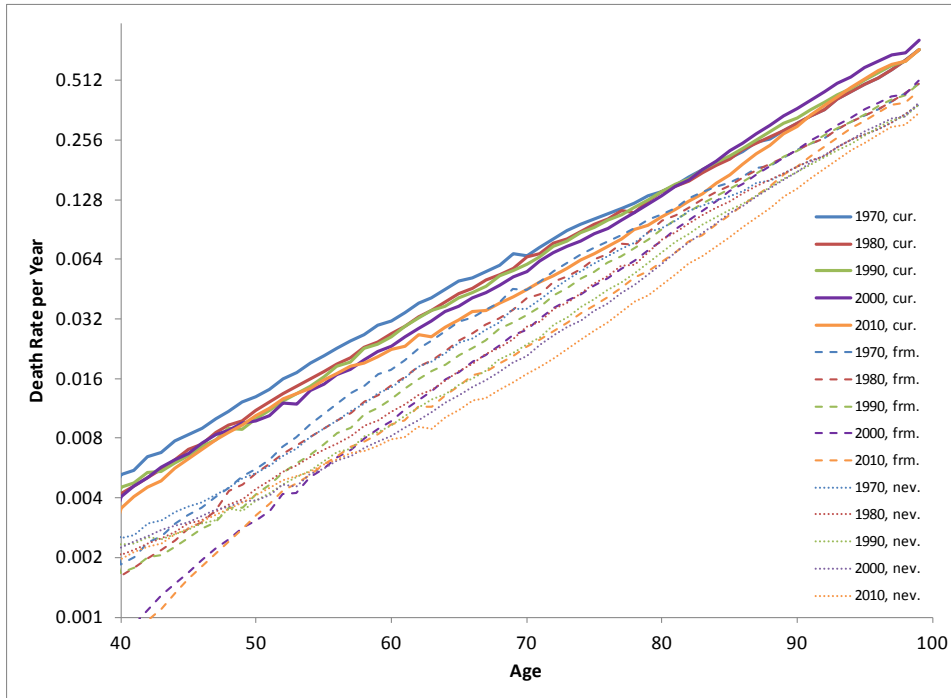


Figure A3. Observed (solid) and primary counterfactual (dashed) proportion of never smokers by age and calendar year.

a) Males



(b) Females

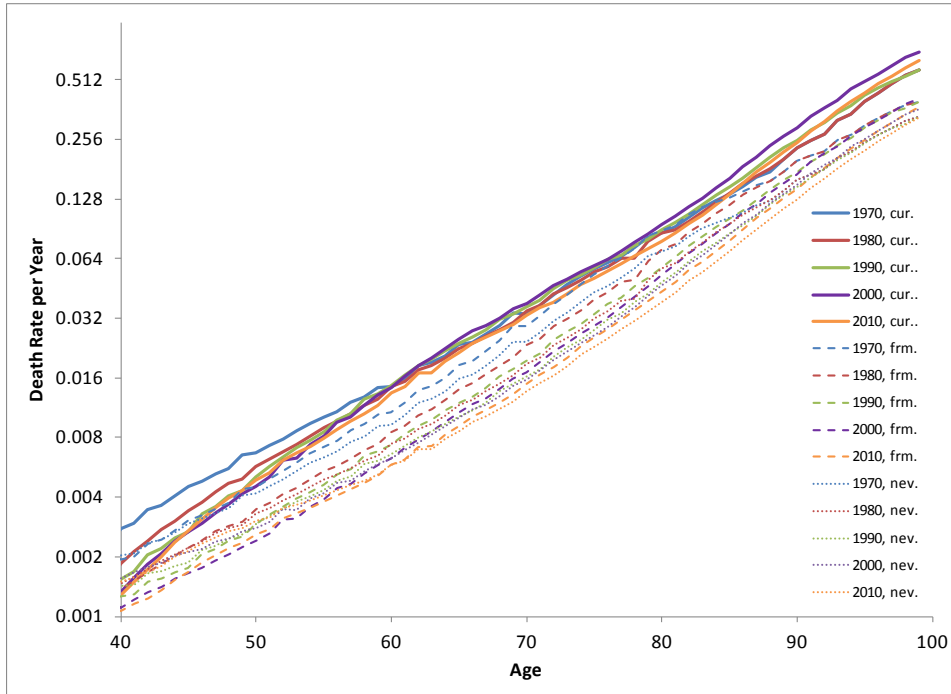
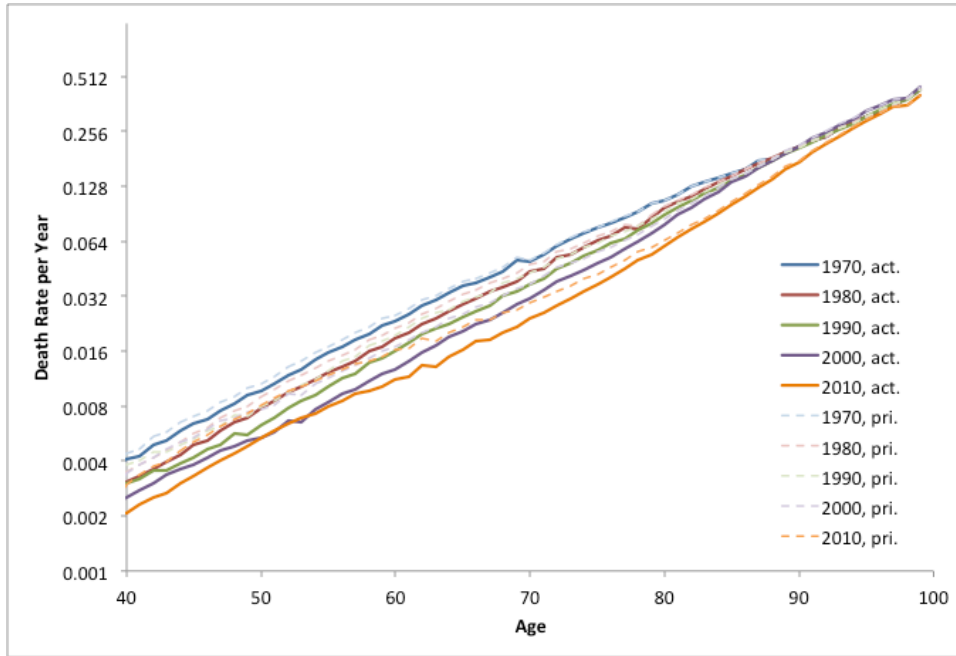


Figure A4. Estimated smoking status-specific mortality rates by year, and smoking status (current=solid, former=broken, and never=dotted).

(a) Males



(b) Females

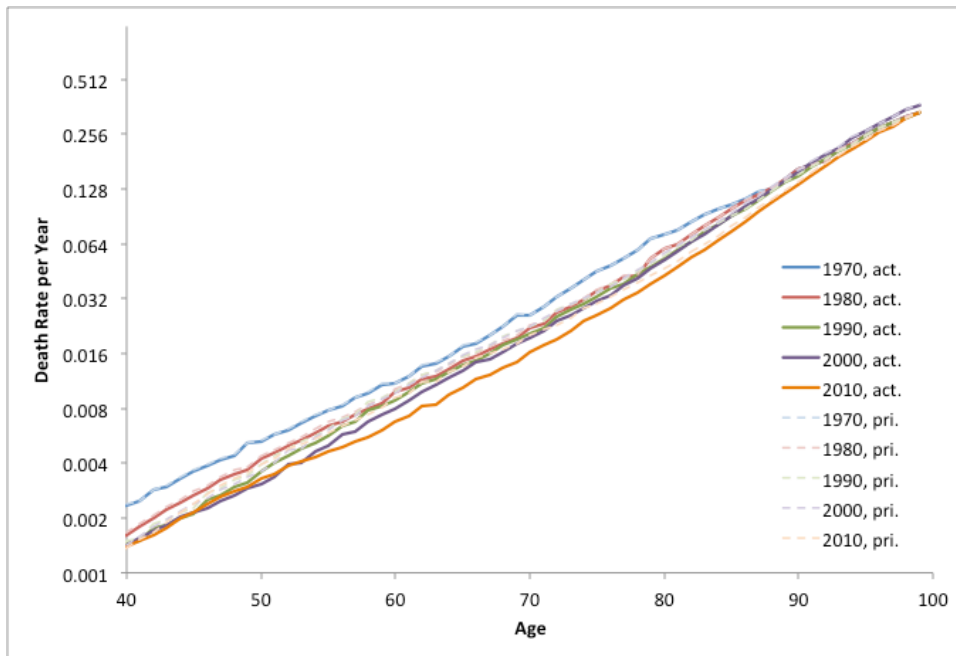


Figure A5. Observed (solid) and primary counterfactual (dashed) annual death rates by age, year and gender.

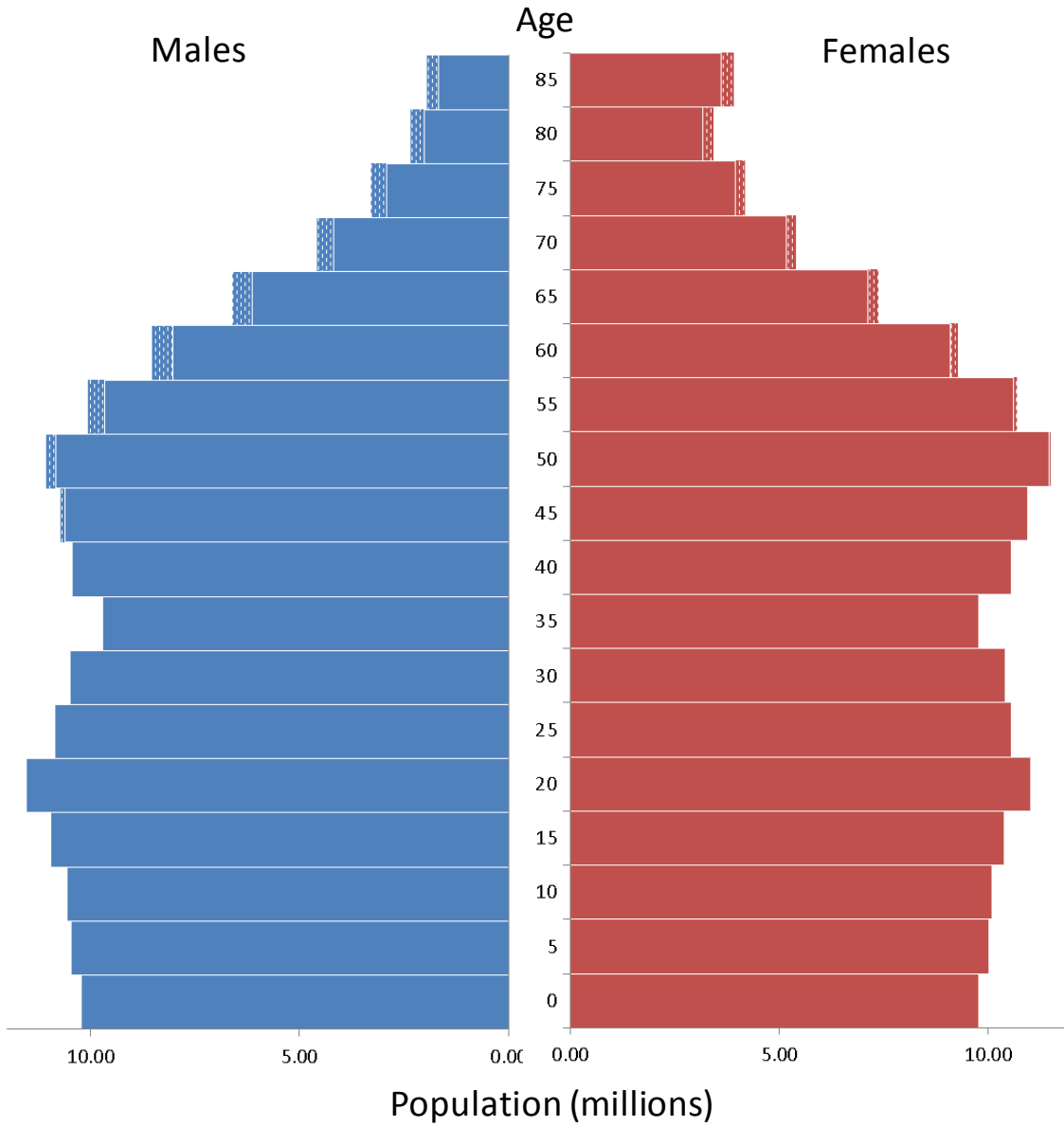
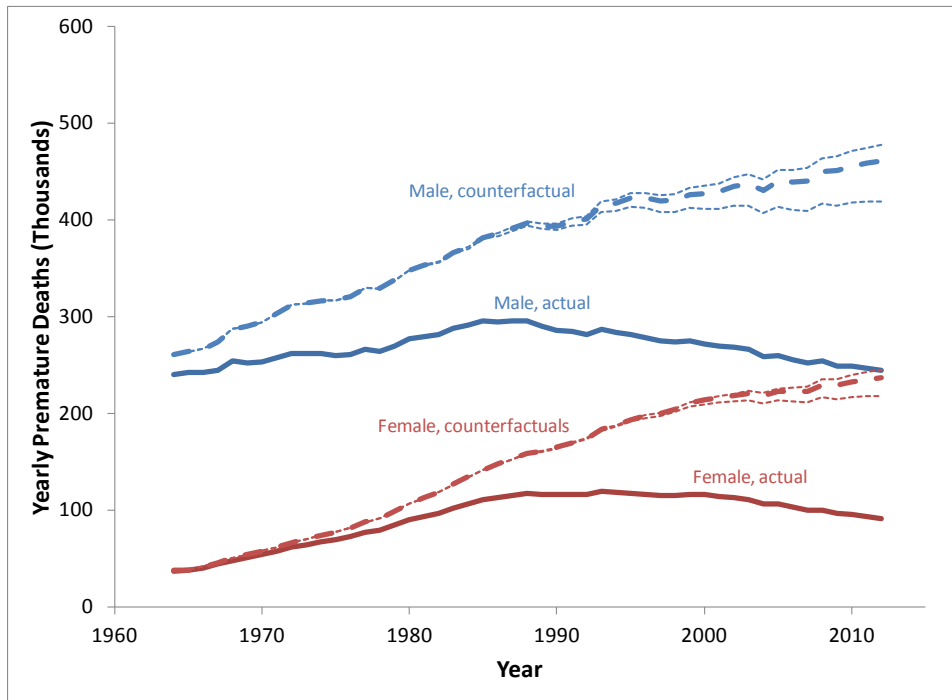


Figure A6. Population distribution in 2012 by age and gender [estimated change (textured) under primary counterfactual].

(a) Premature deaths for all ages



(b) Premature deaths <65

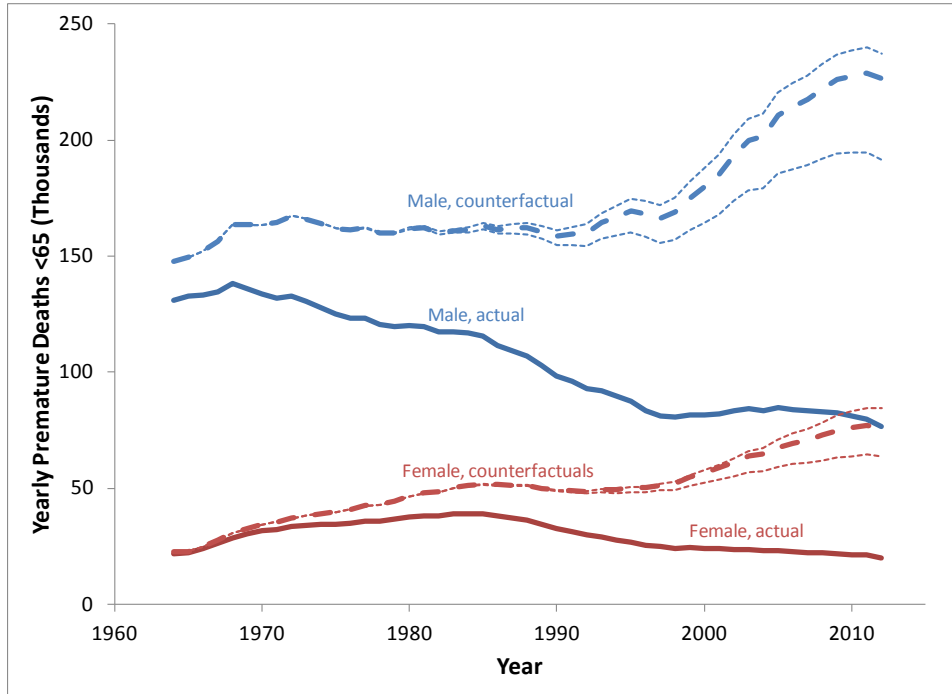
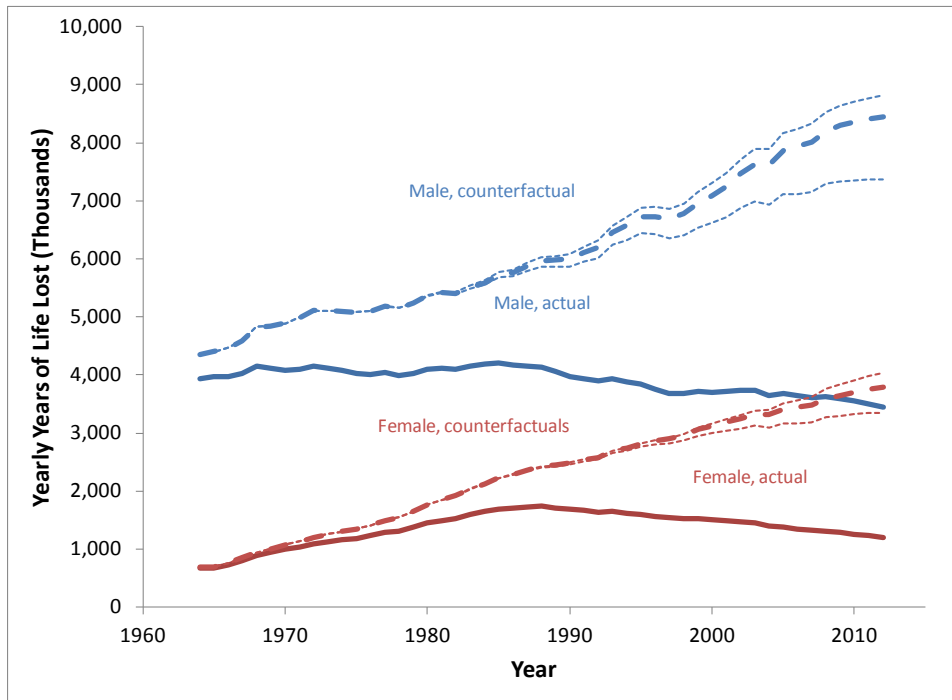


Figure A7. Premature deaths by tobacco control (solid=actual, dashed=primary counterfactual, and dotted=limits) , calendar year and gender.

(a) Total years of life lost from premature deaths for all ages



(b) Total years of life lost from premature deaths before age 65

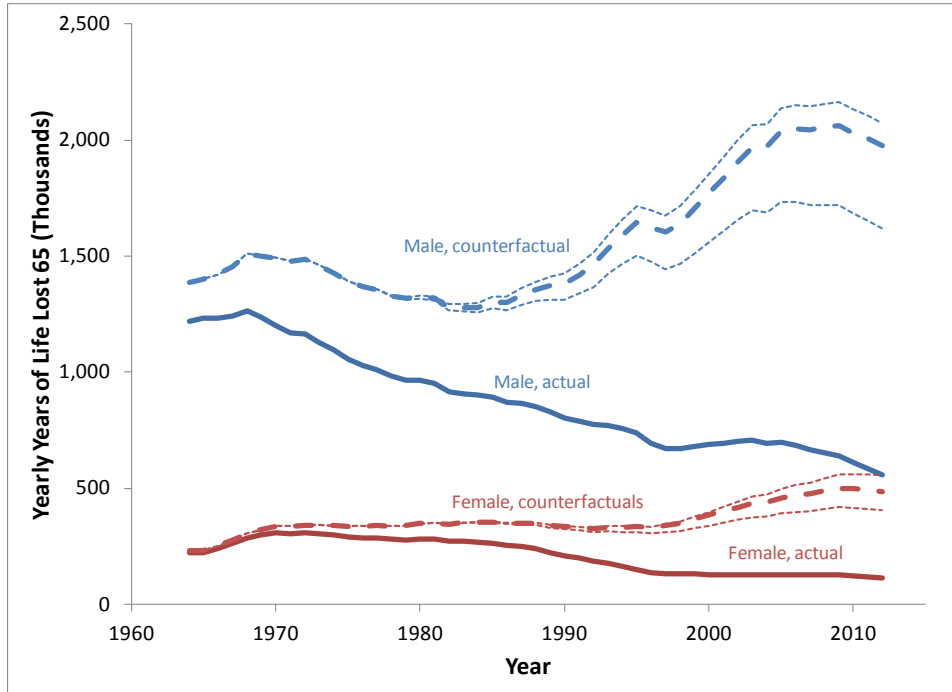
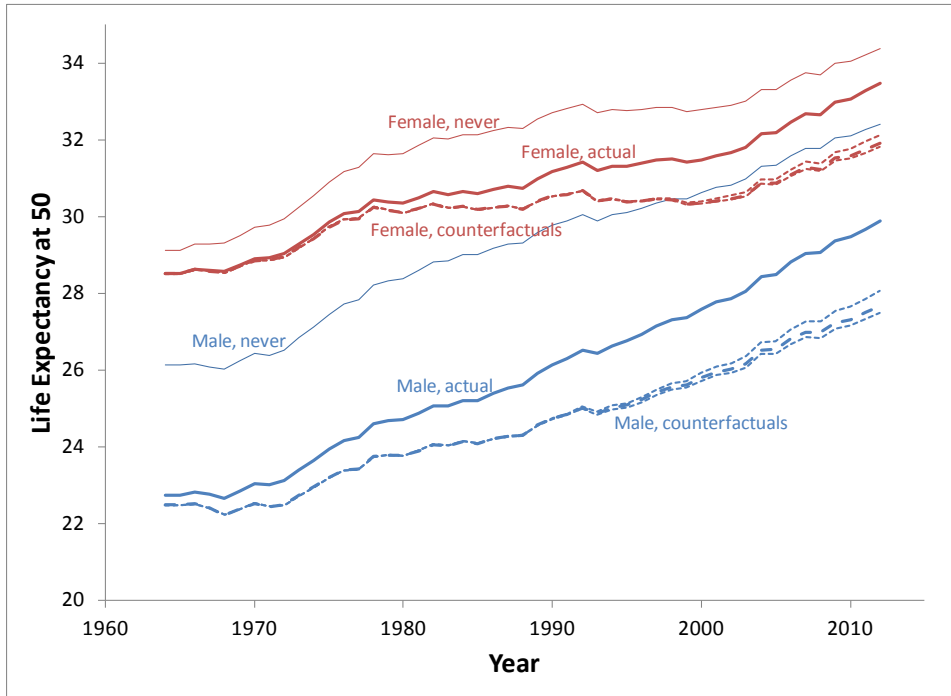


Figure A8. Years of life lost by tobacco control (solid=actual, dashed= primary counterfactual, and dotted=limits for counterfactual), calendar year and gender.

(a) Life expectancy at age 50



(b) Life expectancy at age 60

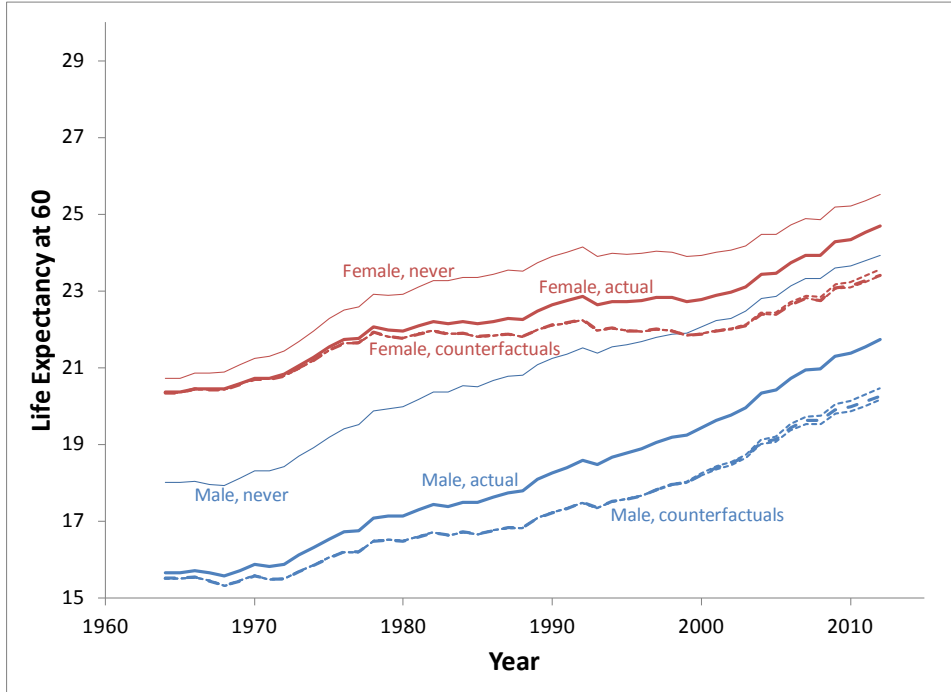


Figure A9. Life expectancy at ages 50 and 60 using actual mortality (solid), and counterfactuals (dashed).