"You've come a long way, baby": The convergence in age patterns of lung cancer mortality by sex, United States, 1959–2013

Natalie A. Rivadeneira*
Andrew Noymer†

6 April 2016

Abstract

We analyze lung cancer mortality by age and sex in the United States, 1959–2013. Mortality patterns by age are a very good fit to a quadratic-Gompertz model, i.e., log mortality rates are quadratic by age, peaking above age 70. These models bring sex differences in lung cancer into sharp relief. With a little additional historical data on sex differences in tobacco use, the models paint a clear picture of behavior-led convergence in lung cancer mortality by sex. In fact, it is uncanny just how well the changes in sex differences in tobacco use are reflected by the quadratic-Gompertz mortality models. While male lung cancer death rates, per se, are statistically-significantly higher than females throughout the data set, the shape of the mortality curves has converged dramatically. Since 1983, the sexes have had statistically-indistinguishable shapes of their quadratic-Gompertz mortality curves. Female lung cancer mortality patterns have shown a transformation from a non-smoking to a smoking pattern.

*Rollins School of Public Health, Emory University, Atlanta, Georgia.
†Department of Population Health and Disease Prevention, University of California, Irvine. To whom correspondence should be addressed. noymer@uci.edu

1 Background

The first goal of this paper is quantitatively to characterize patterns of lung cancer (LC) mortality by age and sex in the United States, from 1959 to the present, which is the period in which detailed cause-specific mortality data are available. We then relate these patterns, qualitatively, to historical sex-specific changes in tobacco use. This results in two principal findings. First, lung cancer mortality is, very clearly, quadratic-Gompertz — viz., when log mortality rates are plotted against age, the LC pattern is a quadratic curve with a negative-valued squared term. This pattern holds for both sexes, and over time, although there are some differences which we analyze. The quadratic pattern with a peak age plays an important role in our analysis.

Our second principal finding is vis-à-vis sex differences. The data since 1959 show three distinct patterns, as follows. In the first pattern (roughly, 1959–1964), male LC mortality is characteristic of tobacco use, while the female pattern, which peaked much older, was more representative of the non-smoking, or background, LC mortality rates. In the second period (roughly, 1965–1982), the female pattern converges to the male pattern, undergoing a transition from background to smoking-related LC mortality. Since 1983, the *pattern* (but not the level) of LC mortality has converged between the sexes, with both males and females experiencing a smoking-like pattern of LC mortality, with peak rates at age 80–89. However, both sexes are undergoing a slow transformation (as smoking prevalence slowly declines) back to background LC mortality patterns, with the peak age of mortality creeping upward. Needless to say, the connection between smoking and LC mortality

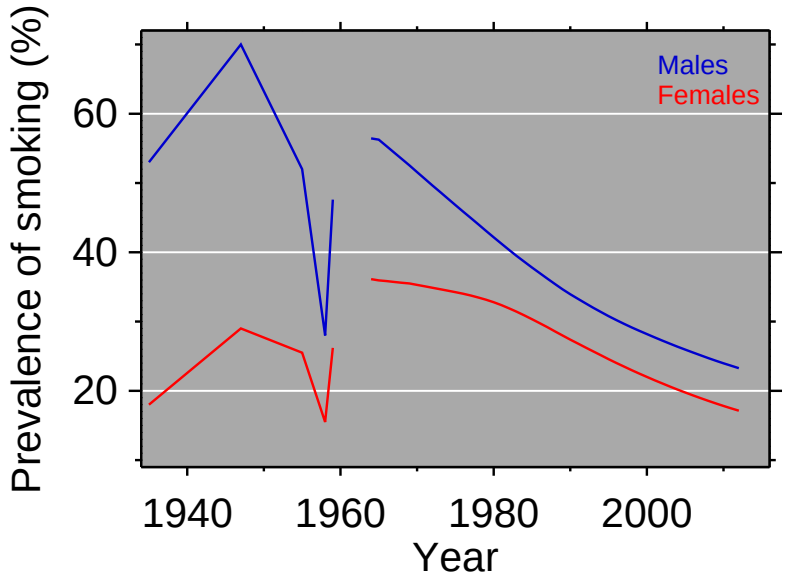


Figure 1: Prevalence of tobacco smoking, USA, 1933–2005. Data from Forey et al. (2012) and Holford et al. (2014).

is not novel (e.g., Hammond and Horn 1954). However, the present analysis of long-run changes in LC mortality sex differences casts new light on the relation between behavior and mortality.

2 Background

It is difficult to analyze lung cancer mortality without considering tobacco use (e.g., Siegel et al. 2015). We obtained sex-specific data on smoking prevalence in the United States, 1933–2005, from Forey et al. (2012) and Holford et al. (2014). These are presented in figure 1. The data for 1965 onward come from the National Health Interview Survey, and do not line-up perfectly with the earlier data, which are compiled from various sources.

However, a number of patterns are clear. In the United States, men smoke more than women, and always have — although the prevalence is converging. Male smoking rates peaked in mid-century, probably in the early 1950s, while female rates peaked later, in the early 1960s. Women adopted smoking later in time, and female cigarette use was still growing at the time of the Surgeon General's report (US Public Health Service 1964). Moreover, womens' smoking declined more slowly than did mens', in absolute terms.

3 Data & Methods

Using mortality multiple cause of death data from the US National Center for Health Statistics (2014), we extracted counts, by age and sex, of all deaths in which lung cancer was the underlying cause (the specific ICD codes are given in Appendix I). The time span is 1959 to 2013, which is the full extent of mortality microdata availability. Death counts were converted to rates by dividing by person-years at risk from the Human Mortality Database (2015). Five-year groups (40–44,45–49,...,95–99) were used to smooth heaping on preferential digits of age. Hereinafter, rate always refers to lung cancer age- and cause-specific death rates.

We estimated quadratic-Gompertz models, separately by sex, by regressing logged death rates on age and age-squared, with a constant. The quadratic specification was chosen after inspection of the data; the canonical Gompertz approach with an intercept and slope, only, is clearly not a fit to the

data. The regression equation is:

$$\log(M_x^{\mathrm{LC}}) = \alpha + \beta_1 x + \beta_2 x^2 + \epsilon, \tag{1}$$

where M_x^{LC} are lung cancer age-specific death rates, x is age, α, β_1, β_2 are parameters to be estimated, and ϵ is the error term. The age range for this analysis was 40 to 99. Below age 40 and above age 100, lung cancer deaths are very rare and thus subject to high variability, and are non-Gompertzian (even with the quadratic adjustment). The regressions are OLS weighted by the number of deaths. For example, in 2010, there were 14,094 lung cancer deaths among men age 70–74, and these were used as weights. Weighting has two advantages over unweighted regressions. Death rates at the age limits (i.e. 40–44, 95–99) are typically the poorest-fitting points in a Gompertzian pattern. Removing these points does not really solve the problem since the estimates then change, and 45–49 also becomes a poor fit, and so on. Weighting solves this problem because there are fewer deaths at the age limits, so these observations are down-weighted, reducing their leverage. The weighting by deaths also produces estimates that are closer to those that would be obtained by maximum likelihood (Abdullatif and Noymer 2016). Moreover, we do not see any disadvantage of using weights. All analyses were performed with Stata, version 13.1 (StataCorp LP, College Station, Texas).

4 Results and Discussion

Figure 2 presents lung cancer age-mortality profiles and fitted quadratic-Gompertz curves, decennially (1960–2010), for both sexes (see Appendix II for all years, 1959–2013). Plotting symbol sizes are proportional to the absolute number of deaths. In 1960, men and women had distinctly different age-mortality profiles for lung cancer. Male death rates were considerably higher than females at all ages. The fitted quadratic-Gompertz curves cross at age 95, implying convergence at that age, but the empirical data diverge from the fit at oldest ages, and males are always higher. The male death rates peak at ages 70–74, whereas the female death rates are nearly monotonic up to age 90–94 (to be precise, in 1960 the female lung cancer death rates peak in the 90–94 age group, but the lower rate in the 95–99 age group, is calculated from only 4 deaths).

Unlike all-cause mortality, or mortality from a number of specific causes (e.g., heart disease), lung cancer in the presence of tobacco smoking is non-Gompertzian (i.e., unless a quadratic term is introduced). Lung cancer mortality does not just keep going up with age, but reaches a peak and then declines (Horiuchi 1997). This refers to period data; Manton et al. (1986) present some analysis of smoking-related mortality by age, holding cohort constant. The reason for this peak and decline is thought to be frailty — i.e., that intense mortality selection in the 70s and 80s (by age) means that the heaviest smokers don't live into their 90s. In addition, cohort differences in tobacco use affect the observed patterns — for example, the male 1910 birth cohort were heavier smokers than men born in 1900 (Moolgavkar

et al. 2012). It is also worth remembering that smokers have higher all-cause mortality (Banks et al. 2015; Carter et al. 2015), so the forces of out-selection of smokers from the population at risk of lung cancer death are complex.

Given both the delayed effect of smoking on mortality, and the later adoption of smoking by women, the pattern of mortality seen in 1960 for women is much closer to a "background" mortality rate for lung cancer (Preston and Wang 2006). This is not only because the female rates are lower (and note that at peak ages male LC death rates in 1960 were about *ten times* higher than those of women), but because they peak at a higher age. Because female smoking was not zero, the female pattern cannot be said to be a true background mortality rate, but it is much closer to it than are males. Exposure to secondhand smoke among females is another reason it is not a true background rate, although the role of secondhand smoke in lung cancer is not clear (Boffetta et al. 1998, Jöckel et al. 1998).

The panels of figure 2 show that, over time, the female pattern progressively looks more and more like that of males. By 1980, female death rates are very clearly peaking in the 70s (of age), and are in the neighborhood of 100 per 100,000. By 1990, the female rates are much higher than 100 per 100,000. The female pattern in 1990 looks somewhat parallel to that of males, at a lower level; it looks a lot like the male pattern in 1960. By 2010, the female LC mortality rates are nested neatly below those of males.

The peak age plays an important role in this analysis is of sex differences in LC mortality because it provides a unidimensional summary of the effect of past tobacco use and its effect on mortality. The older the peak age, the closer the LC pattern is to background mortality, and the younger the

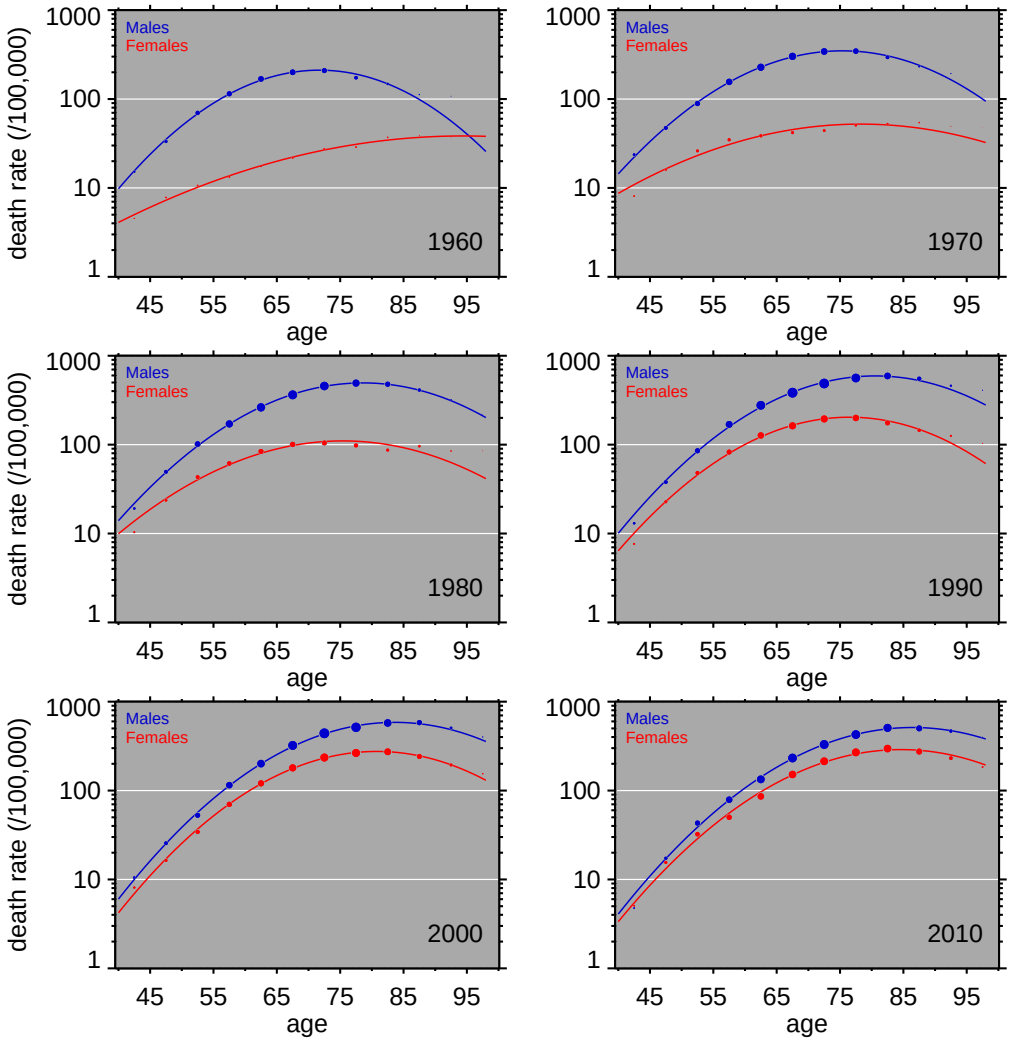


Figure 2: Age-mortality profiles, by sex. Lung cancer, USA, 1960–2010 (decennial). With quadratic-Gompertz fit.

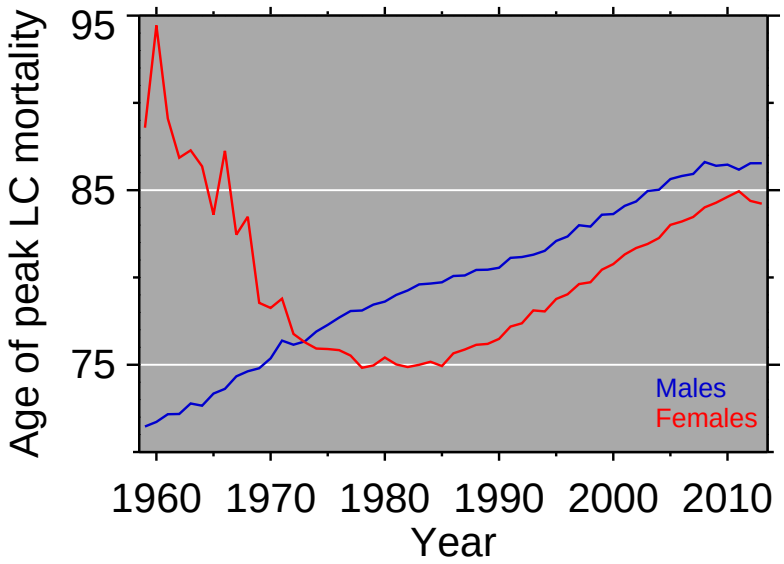


Figure 3: Maximum age of lung cancer mortality based on derivative of quadratic-Gompertz model.

peak age, the closer it is to a tobacco-influenced pattern. To quantify the quadratic-Gompertzian peak age, we solved the following equation:

$$\frac{d}{dx}M_x^{\mathrm{LC}} = \beta_1 + 2\beta_2 x^* = 0, \tag{2}$$

where x^* is the peak age of mortality (this is for the un-logged analog of equation 1, but the logarithm is a monotone transformation, so x^* is the same). Thus, $x^* = -0.5\beta_1/\beta_2$, which must be positive since $\beta_1 > 0$ and $\beta_2 < 0$ in the LC patterns.

Figure 3 shows the pattern of x^* for males and females. Consider males first. The mean age of male LC mortality has been rising, almost monotonically, for over half a century. An obvious question is, if increasing mean age

of mortality indicates a return to a background rates, why does it occur during a period when *peak* rates (i.e., $M_{x^*}^{\mathrm{LC}}$ in the prior notation) are increasing and then decreasing (cf. figure 2)? The answer is that smoking partly is a cohort phenomenon and as smoking prevalence declines, it does so on a cohort basis. It should be noted that all-cause mortality is falling during this period, so the increase in peak age of LC mortality is partly fed by reduced competing risks at younger ages.

The peak age is essentially increasing because smoking is going down and the pattern is (slowly) returning to the pattern of "background" (i.e., tobacco-free) lung cancer mortality, of which 1960 females is the best exemplar here. However, this pattern should be interpreted cautiously, because peak age is also affected by competing risks. As mortality for other causes (most notably, cardiovascular disease) has declined since the 1960s, more people are surviving longer, and must die of something [rephrase?] at older ages. Multiple cause mortality is a complex phenomenon since cardiovascular deaths averted at younger ages may be conspecific and simply delayed to later ages as opposed to necessarily transferring to another cause. Moreover, the declining use of heart-related conditions as a "garbage code" for unknown causes of death, especially at advanced ages (Preston 1976), may mean that some of the LC deaths were there before, so to say, but only became coded as malignant neoplasm as cause of death classification became refined. It should also be noted that the effects of tobacco on cardiovascular health have been noted almost as long its effects on lung cancer (Russek et al. 1955,Wald et al. 1973).

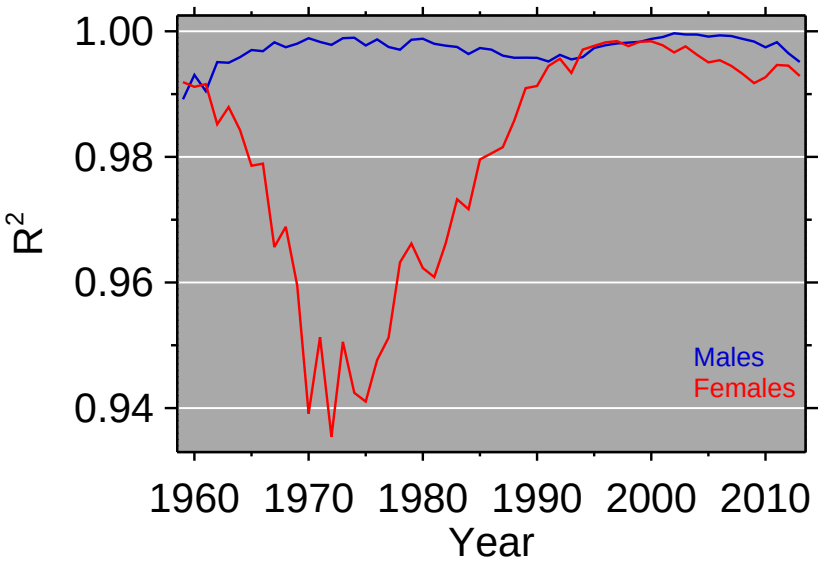


Figure 4: Goodness-of-fit (R^2) of quadratic-Gompertz models.

Figure 4 shows the goodness-of-fit (R^2) of the annual quadratic-Gompertz models. The fit is excellent, with all models above 93%. The telling curve is that for females, which begins with a truly excellent fit (99%), then dips, and then rises again. What is happening is that during the early 1960s, the model is a good fit because it's mostly cohorts of women who weren't heavy smokers. The model fit declines as the heavier-smoking female cohorts move through the ages, and the fit returns to good later on when a new equlibrium of lower-smoking cohorts is in place. We miss this transition in the male series because the machine-readable detail mortality data in the US begins in 1959.

To test the sex differences in the shapes of the quadratic-Gompertz LC models, we pooled the male and female data and ran models with full-

year	α	β_1	β_2	year	α	β_1	β_2
1959	0.000	0.000	0.000	1987	0.534	0.356	0.885
1960	0.000	0.000	0.000	1988	0.729	0.523	0.882
1961	0.000	0.000	0.000	1989	0.749	0.994	0.430
1962	0.000	0.000	0.000	1990	0.813	0.943	0.490
1963	0.000	0.000	0.000	1991	0.771	0.974	0.477
1964	0.000	0.000	0.000	1992	0.532	0.776	0.296
1965	0.000	0.000	0.000	1993	0.868	0.862	0.660
1966	0.000	0.000	0.000	1994	0.452	0.676	0.272
1967	0.000	0.000	0.000	1995	0.653	0.933	0.364
1968	0.000	0.000	0.000	1996	0.338	0.587	0.176
1969	0.000	0.000	0.000	1997	0.381	0.618	0.172
1970	0.000	0.000	0.000	1998	0.454	0.686	0.231
1971	0.000	0.000	0.000	1999	0.394	0.648	0.205
1972	0.000	0.000	0.000	2000	0.453	0.758	0.252
1973	0.000	0.000	0.000	2001	0.735	0.945	0.437
1974	0.000	0.000	0.000	2002	0.825	0.898	0.509
1975	0.000	0.000	0.001	2003	0.223	0.403	0.113
1976	0.001	0.000	0.001	2004	0.573	0.798	0.363
1977	0.001	0.000	0.002	2005	0.909	0.898	0.660
1978	0.003	0.001	0.008	2006	0.959	0.867	0.669
1979	0.005	0.001	0.012	2007	0.823	0.714	0.855
1980	0.011	0.003	0.021	2008	0.728	0.879	0.576
1981	0.088	0.033	0.163	2009	0.950	0.821	0.897
1982	0.101	0.047	0.237	2010	0.885	0.750	0.995
1983	0.145	0.077	0.372	2011	0.507	0.380	0.604
1984	0.288	0.168	0.552	2012	0.899	0.807	0.899
1985	0.629	0.420	0.921	2013	0.975	0.953	0.795
1986	0.578	0.370	0.949				

Table 1: Annual sex difference test. These are p-values (not coefficients) for sex differences of each coefficient. Coefficients as in equation 1.

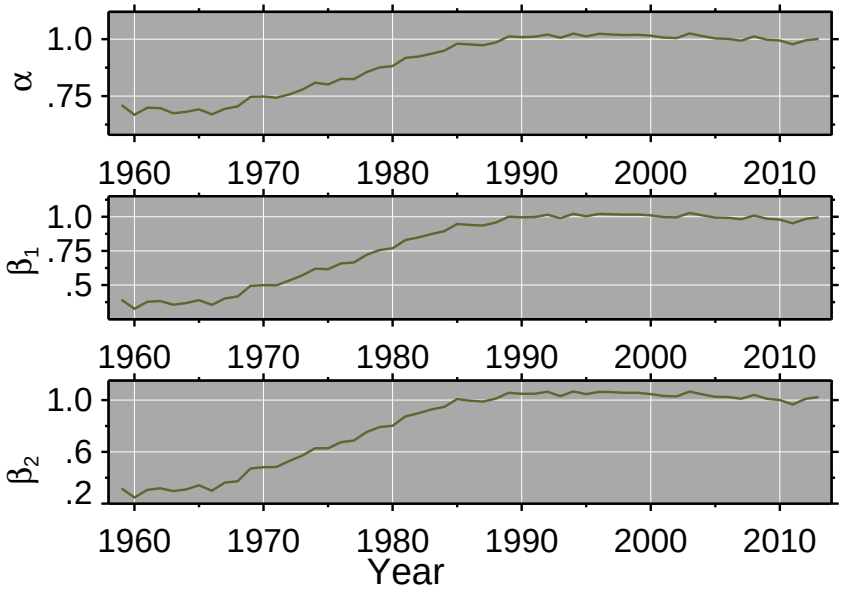


Figure 5: Female/male ratio, quadratic Gompertz coefficients over time.

interaction (sex, sex×age, sex×age-squared). These models recapitulate the coefficients in the single-sex models, but permit testing the sex differences between the suite of coefficients $(\alpha, \beta_1, \beta_2)$ in each year. These are tests of the shape of the quadratic curves, not tests of sex differences, per se — with thousands more male LC deaths, the overall male excess is always significant. The question is not, are male LC death rates higher (they are), but, are the male and female patterns (per se) distinguishable? Table 1 gives the results, as p-values. The evolution is clear: from 1959–1980 males and females have statistically-distinguishable differences in the shape of their quadratic-Gompertz LC mortality patterns. This is followed by two years of transition, and from 1983 to the end of the data set, males and females have no distinguishable shape differences in the LC mortality patterns.

A more direct approach is shown in figure 5. Rather than calculate *p*-values in a interactive model, it just plots the female to male ratio of the coefficient values. It is clear that since the mid 1980s, there has been enormous convergence, with ratios very close to 1.0.

5 Conclusion

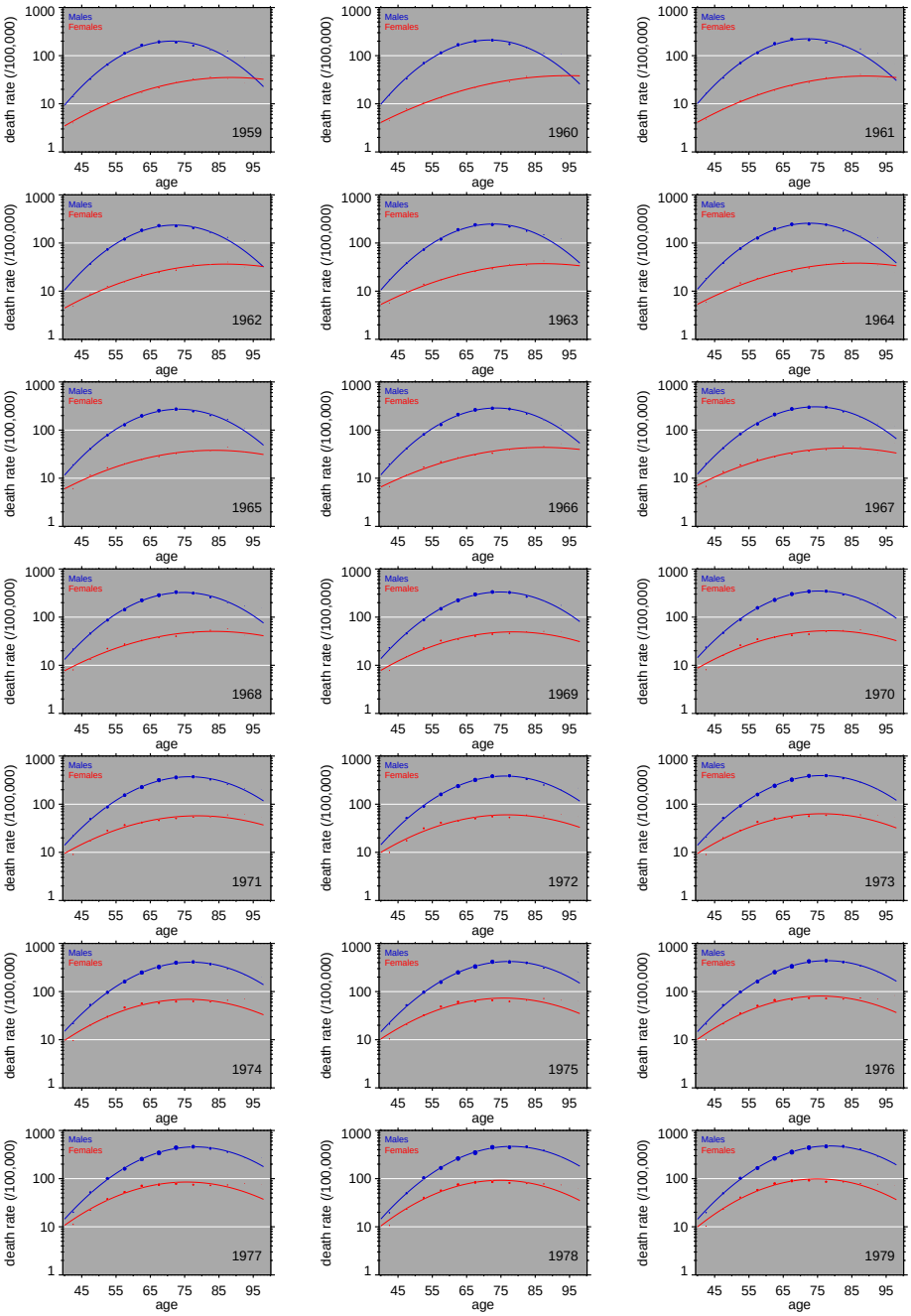
Period lung cancer mortality by age — especially in tobacco-smoking populations — is very firmly quadratic-Gompertz in pattern. Although this has been documented previously (for instance, Horiuchi and Wilmoth 1997), the present study is the most comprehensive treatment of which we are aware.

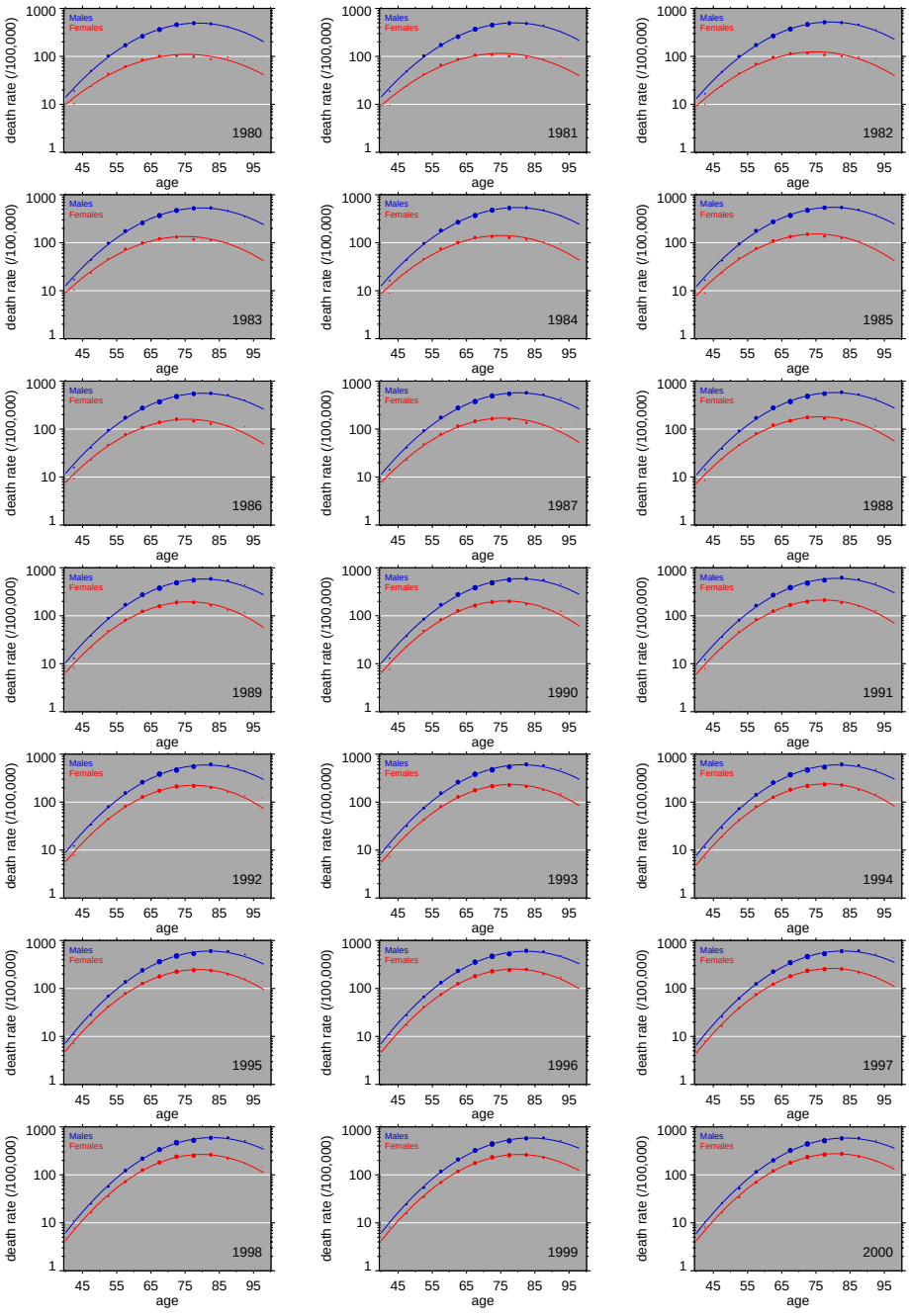
Past patterns of tobacco use in males and females have converged. Since 1983, there has been no statistically-distinguishable difference in the pattern of LC mortality between men an women. Male LC death rates exceed those of women, but the shapes of the age-mortality profiles have been the same since 1983. In the United States, there is no gendered *pattern* of lung cancer mortality, though there are sex differences in *level*. Mortality very often reflects behavior, never more so than with lung cancer and cigarette smoking. The shape of the lung cancer age-mortality profile reflects cohort histories of cigarette use, and these are now similar enough for both sexes that the pattern (but not the level) of lung cancer mortality is the same for both sexes.

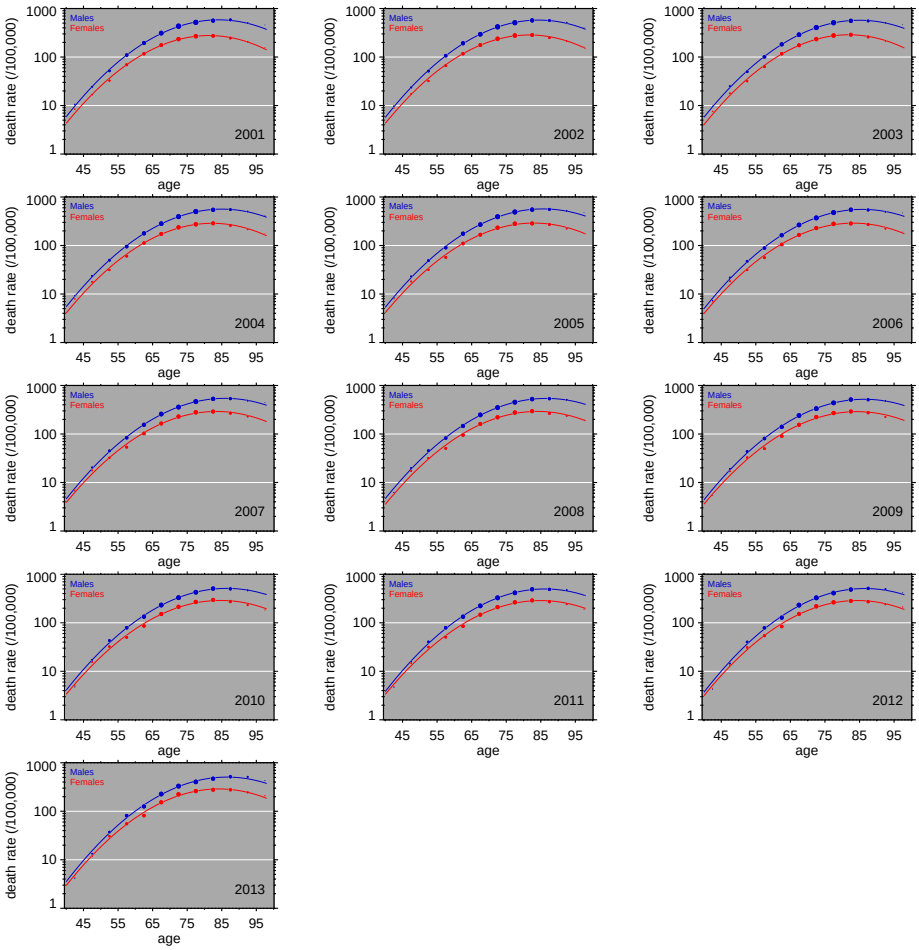
Appendix I: ICD codes for lung cancer mortality

	Cause of death	
Years	**ICD code**	**Description** (m.n., malignant neoplasm)
1959–1967 (ICD 7)	162	m.n. of bronchus and trachea, and of lung specified as primary
	163	m.n. of lung, unspecified as to whether primary or secondary
1968–1978 (ICD 8)	162	m.n. of trachea, bronchus and lung
1979–1998 (ICD 9)	162	*ibid.*
1999–2013 (ICD 10)	C33	m.n. of trachea
	C34	m.n. of bronchus and lung

Appendix II: Graphs for all years




Males
Females
death rate (/100,000)
1000
100
10
1
45 55 65 75 85 95
age
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000



Works Cited

Abdullatif, Viytta N. and Andrew Noymer. 2016. "*Clostridium difficile* infection: An emerging cause of death in the twenty-first century." *Biodemography and Social Biology* accepted for publication.

Banks, Emily, Grace Joshy, Marianne F. Weber, Bette Liu, Robert Grenfell, Sam Egger, Ellie Paige, Alan D. Lopez, Freddy Sitas, and Valerie Beral. 2015. "Tobacco smoking and all-cause mortality in a large Australian cohort study: Findings from a mature epidemic with current low smoking prevalence." *BMC Medicine* 13(38).

Boffetta, Paolo, Antonio Agudo, Wolfgang Ahrens, Ellen Benhamou, Simone Benhamou, Sarah C. Darby, Gilles Ferro, Cristina Fortes, Carlos A. Gonzalez, Karl-Heinz Jöckel, Martin Krauss, Lothar Kreienbrock, Michaela Kreuzer, Anabela Mendes, Franco Merletti, Fredrik Nyberg, Göran Pershagen, Hermann Pohlabeln, Elio Riboli, Giovanni Schmid, Lorenzo Simonato, Jean Trédaniel, Elise Whitley, Heinz-Erich Wichmann, Carlos Winck, Paola Zambon, and Rodolfo Saracci. 1998. "Multicenter case-control study of exposure to environmental tobacco smoke and lung cancer in Europe." *Journal of the National Cancer Institute* 90(19):1440–1450.

Carter, Brian D., Christian C. Abnet, Diane Feskanich, Neal D. Freedman, Patricia Hartge, Cora E. Lewis, Judith K. Ockene, Ross L. Prentice, Frank E. Speizer, Michael J. Thun, and Eric J. Jacobs. 2015. "Smoking and mortality: Beyond established causes." *New England Journal of Medicine* 372(7):631–640.

Forey, Barbara, John Hamling Jan Hamling, Alison Thornton, and Peter Lee. 2012. *International smoking statistics: A collection of worldwide historical data*. Online resource: `www.pnlee.co.uk`.

Hammond, E. Cuyler and Daniel Horn. 1954. "The relationship between human smoking habits and death rates: A follow-up study of 187,766 men." *JAMA* 155(15):1316–1328.

Holford, Theodore R., Rafael Meza, Kenneth E. Warner, Clare Meernik, Jihyoun Jeon, Suresh H. Moolgavkar, and David T. Levy. 2014. "Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964–2012." *JAMA* 311(2):164–171.

Horiuchi, Shiro. 1997. "Postmenopausal acceleration of age-related mortality increase." *Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 52A(1):B78–B92.

Horiuchi, Shiro and John R. Wilmoth. 1997. "Age patterns of the life table aging rate for major causes of death in Japan, 1951–1990." *Journal of Gerontology: Biological Sciences* 52A(1):B67–B77.

Human Mortality Database. Accessed 15 January 2015. *U.S.A. Complete Data Series: Exposure-to-risk*. `http://www.mortality.org/`.

Jöckel, Karl-Heinz, Hermann Pohlabeln, Wolfgang Ahrens, and Martin Krauss. 1998. "Environmental tobacco smoke and lung cancer." *Epidemiology* 9(6):672–675.

Manton, Kenneth G., Eric Stallard, and James W. Vaupel. 1986. "Alternative models for the heterogeneity of mortality risks among the aged." *Journal of the American Statistical Association* 81(395):635–644.

Moolgavkar, Suresh H., Theodore R. Holford, David T. Levy, Chung Yin Kong, Millenia Foy, Lauren Clarke, Jihyoun Jeon, William D. Hazelton, Rafael Meza, Frank Schultz, William McCarthy, Robert Boer, Olga Gorlova, G. Scott Gazelle, Marek Kimmel, Pamela M. McMahon, Harry J. de Koning, and Eric J. Feuer. 2012. "Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975–2000." *Journal of the National Cancer Institute* 104(7):541–548.

National Center for Health Statistics. 2014. *Mortality multiple cause-of-death data files,* `http://www.cdc.gov/nchs/nvss/mortality_public_use_data.htm`. National Center for Health Statistics. Accessed 15 January 2015.

Preston, Samuel H. 1976. *Mortality patterns in national populations: With special reference to recorded causes of death*. Academic Press, New York.

Preston, Samuel H. and Haidong Wang. 2006. "Sex mortality differentials in the United States: The role of cohort smoking patterns." *Demography* 43(4):631–646.

Russek, Henry I., Burton L. Zohman, and Virgil J. Dorset. 1955. "Effects of tobacco and whiskey on the cardiovascular system." *Journal of the American Medical Association* 157(7):563–568.

Siegel, Rebecca L., Eric J. Jacobs, Christina C. Newton, Diane Feskanich, Neal D. Freedman, Ross L. Prentice, and Ahmedin Jemal. 2015. "Deaths due to cigarette smoking for 12 smoking-related cancers in the United States." *JAMA Internal Medicine* 175(9):1574–1576.

US Public Health Service. 1964. "Smoking and health: Report of the advisory committee to the Surgeon General of the Public Health Service." Public Health Service Publication 1103, US Department of Health, Education, and Welfare, Washington DC.

Wald, Nicholas, Susannah Howard, P. G. Smith, and Knud Kjeldsen. 1973. "Association between atherosclerotic diseases and carboxyhaemoglobin levels in tobacco smokers." *British Medical Journal* 1(5856):761–765.