Liars, Cheaters and Procrastinators: How They Upset Mortality Studies

R. C. W. “Bob” Howard, FSA, FCIA

Presented at the Living to 100 Symposium

Orlando, Fla.

January 8–10, 2014

Liars, Cheaters and Procrastinators:

How They Upset Mortality Studies

R. C. W. "Bob" Howard, FSA, FCIA

Contents

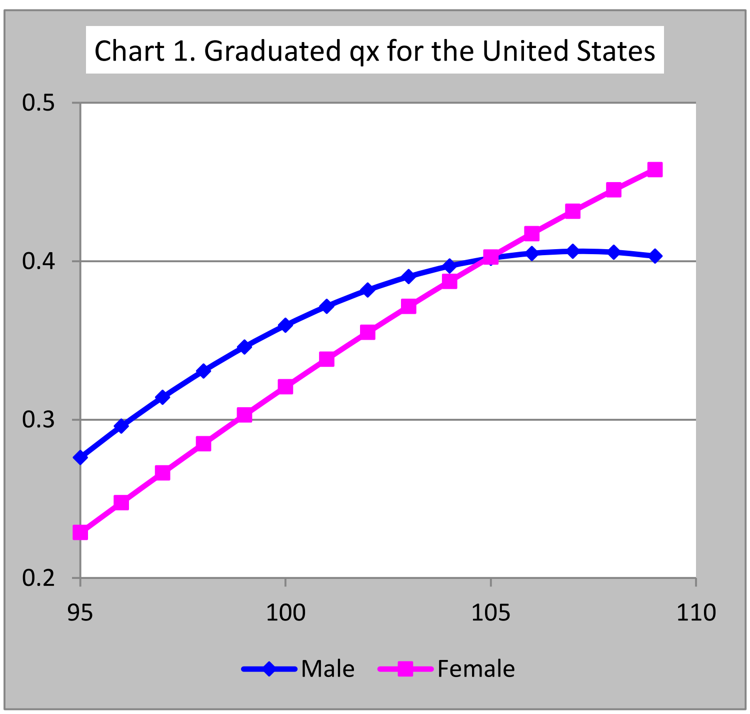
1 Abstract

Odd patterns in mortality curves can be caused by bad data. The author hypothesizes that the negatively sloped mortality rates at the highest ages for U.S. males are caused by overstating age at death, and the low mortality rates for Japanese females by failure to report deaths. The paper upholds the first hypothesis but soundly defeats the second. The paper also demonstrates the impact of mishandling incurred but unreported deaths.

2 Background

At the 2011 Living to 100 Symposium, I presented a paper giving a method for developing high-age mortality rates from death records alone (Howard 2011). My contribution lay in inferring data for cohorts not yet extinguished. I applied the method to the data of five countries all with raw death data distinguished by individual ages at least up to 109. The results were reasonable, and fairly consistent, across most countries for males and females. The one glaring exception was U.S. males. Chart 1 shows mortality rates continue to rise with age for females but plateau and even decline for males.

Finding an explanation for this anomaly led me to the work for this paper.



3 Liars: Overstatement of Age

A comment I remembered from the 2008 Living to 100 Symposium started me on a search for an explanation. I believe it was Bruce Schobel who said that overstatement of age was fairly common among the elderly. Apparently, males, in particular, enjoy a sense of status by being old, and the older they are, the higher the status.

Obviously lying about age, if sufficiently prevalent, will result in inaccurate mortality statistics. But is that a reasonable explanation? To accept the hypothesis of lying, we would need reason to believe that enough people might be willing and able to lie, and we would need to demonstrate that the usual pattern of the observed mortality rates is consistent with overstatement of age at death.

First, we should note there are few, if any, financial incentives to overstate age once attaining 65. Overstating age would have to be done solely for social status. It is unlikely the overstatement would be highly prevalent.

Second, there is even less incentive for surviving relatives or caregivers to overstate age after death, and my method uses only death records, not census records. The overstatement would have to be maintained for long enough and consistently enough that those reporting the age would think it to be true. This also argues for low prevalence.

Third, the presence of a birth certificate would make the fiction more difficult to maintain. The later people were born, the more likely it is they would have documentary evidence of birth and the more frequently such a document would be required. This argues for a pattern of overstatement that correlates to birth cohort.

Intuitively, mortality rates would appear lower than they should because of the overstatement of age at death, but could the mortality curve turn downward? I believe the best way to explore the possibilities is to use a model of a national population.

3.1 Model

My focus is on the data for ages 95+, but to ensure that model conditions have stabilized by that age, I need to start the model much younger.

The starting point for my model is the U.S. male population, as reported by the Human Mortality Database, www.mortality.org, in 1960 for ages 60–119 and at age 60 for years 1960–2009. I assume mortality follows a Gompertz table (derived by fitting to U.S. mortality rates for ages 65–94, but the precise derivation is unimportant) and moderate mortality improvement varying by age. I can then calculate exposure and deaths for all ages and years to 2010. Initially, I thought of generating deaths randomly with a binomial distribution, but the population is large enough that I gained very little "realism" by a stochastic simulation. I then calculate mortality rates for ages 95–110 using data for the last 15 years only. That is my baseline.

Next, I inject systematic errors, such as lying about age and work through the impact on exposure and deaths. Then I recalculate the mortality rates and compare.

3.2 Simulating Liars

Although my model has both the number alive at the start of each year and the number dying during each year for each age-year cell, my focus is on the impact on determining mortality rates from death records alone. Therefore, for each change in death information, there is resulting change in the exposures and continuing backward along the birth cohort.

For example, if an actual death at age 88 for 1988 is reported at age 98, then the model's

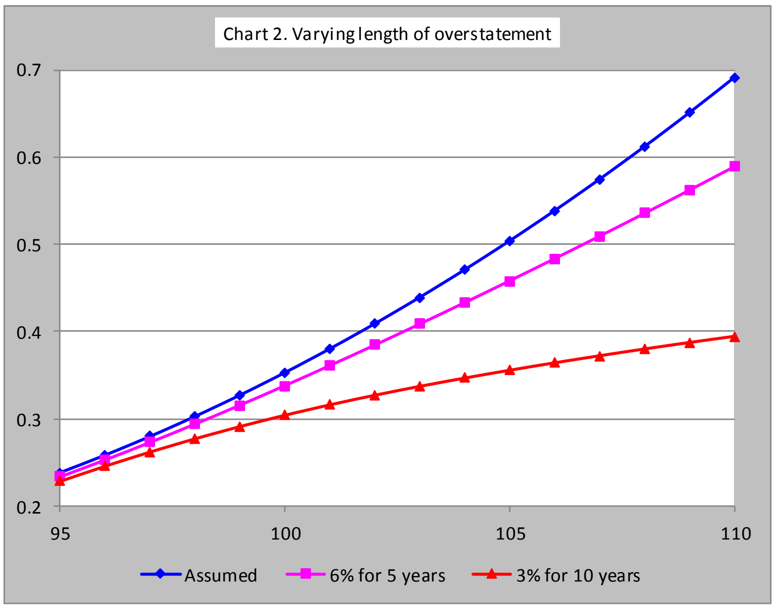
1. deaths are decreased by 1 at age 88 for 1988
2. deaths are increased by 1 at age 98 for 1988
3. census is decreased by 1 for the birth cohort 1900 for all ages up to and including 88
4. census is increased by 1 for the birth cohort 1890 for all ages up to and including 98

Thus, lying about age has the effect of leaving one birth cohort and transferring to another with death in the same year.

Although my model can vary the length of the overstatement, I usually used 10 years. It seems to me that those who want to lie about their age consistently over time will probably shift their birth year by a round number. Of course, some might like to round down to a multiple of 10, such as one born in 1887 saying he was born in 1880, and it might feel somewhat truthful because he got the decade right.

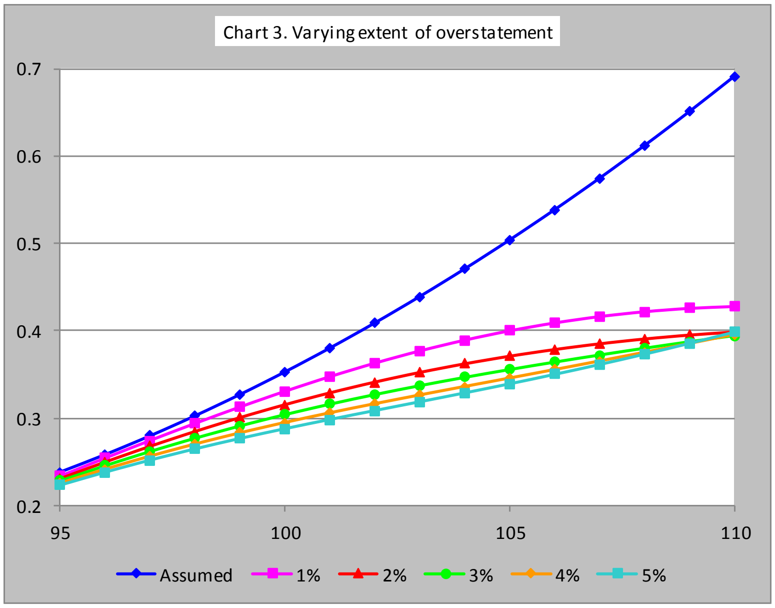
3.3 Observations

I quickly found that the impact on mortality rates increases with the length of the overstatement (see chart 2). An overstatement of 6 percent of deaths by five years each has much less impact than 3 percent overstated by 10 years, even though the number of life-years overstated is the same.

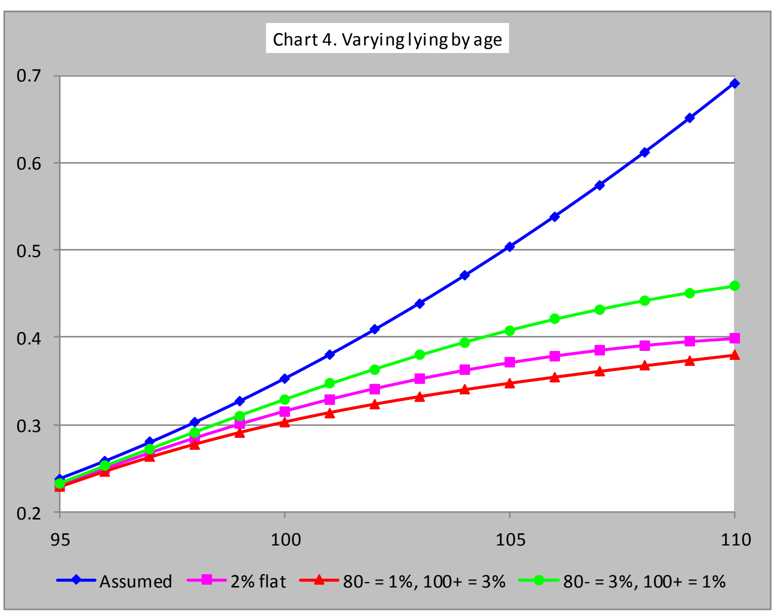


All subsequent charts show overstatements of 10 years only.

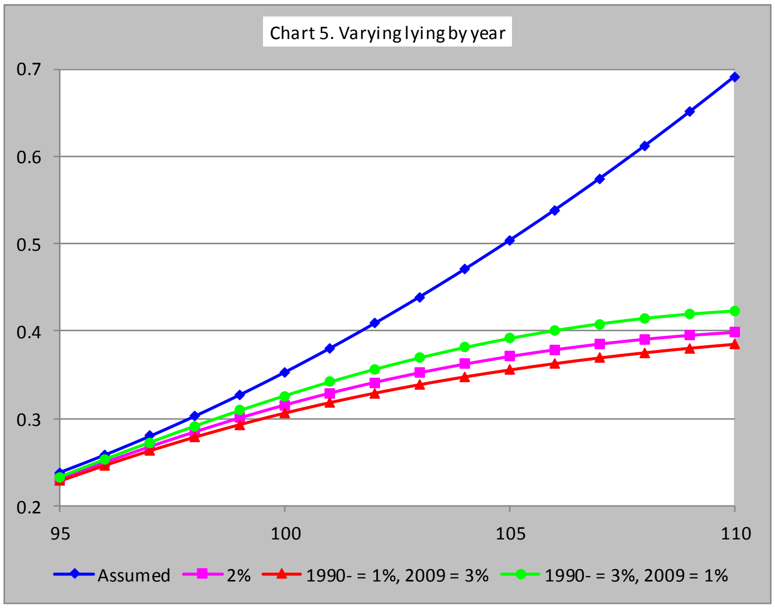
The greater the proportion of deaths with overstated age, the greater the impact on observed mortality rates, but the relationship is certainly not linear (see chart 3). What is particularly surprising is that the impact at age 110, which is very large, seems to depend little on the extent of overstatement.



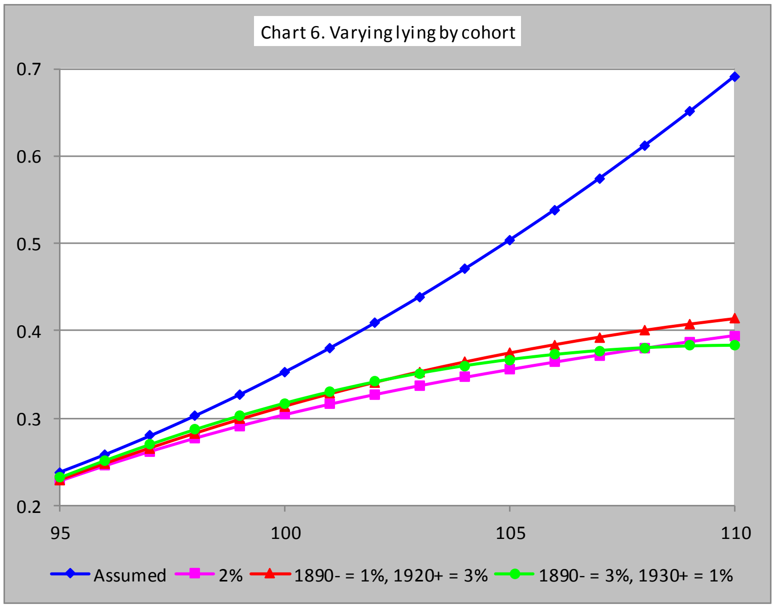
Perhaps the pattern of lying about age has an impact on the mortality curve (see chart 4). The pink line has overstatement on 2 percent of deaths at all ages. The red line shows 1 percent overstating for ages up to 80, 3 percent for ages 100 and up, with linear interpolation in between. The green line is like the red but the percentages are reversed. The impact is stronger when the overstatement is greater at higher ages at death. However, we still do not see the pattern we seek, a mortality curve with a negative slope at the high ages.



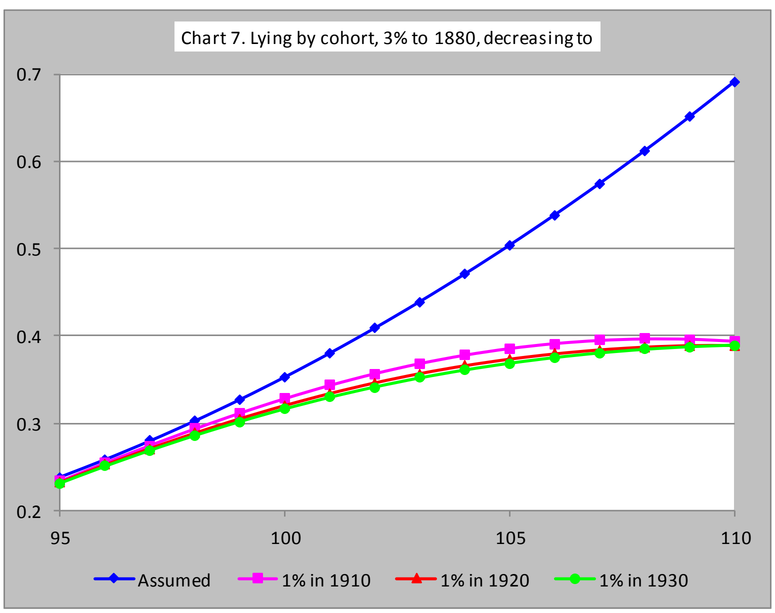
Perhaps we should consider the proportion lying varying by year of death rather than by age. Chart 5 shows three patterns varying by year. The red and green lines show a constant rate of overstatement to 1990, changing linearly to another rate in 2009. None of these look hopeful.



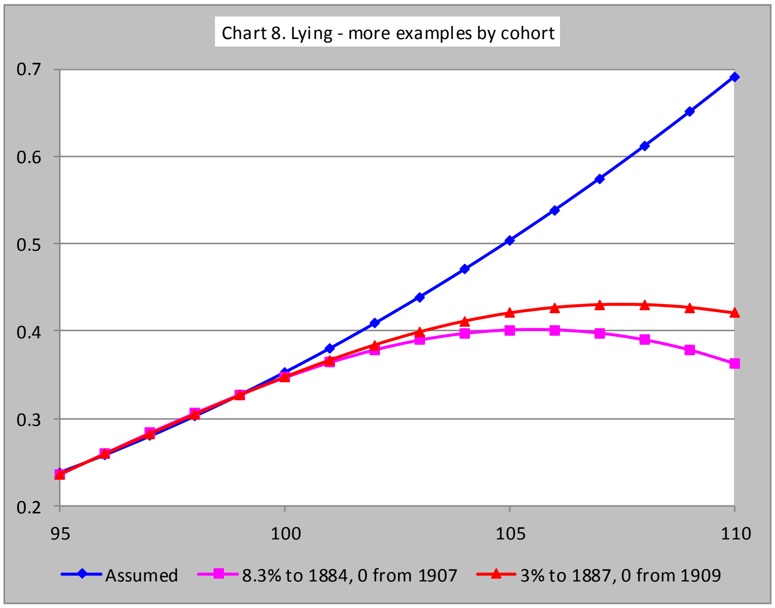
The next pattern to try is varying by birth cohort. This approach seems plausible because the likelihood of having proof of birth date is likely to increase with more recent cohorts. Chart 6 shows three patterns of overstatement by birth cohort. The green line, with 3 percent overstating to 1890, decreasing linearly to 1 percent for 1930 and later comes close to having the desired pattern.



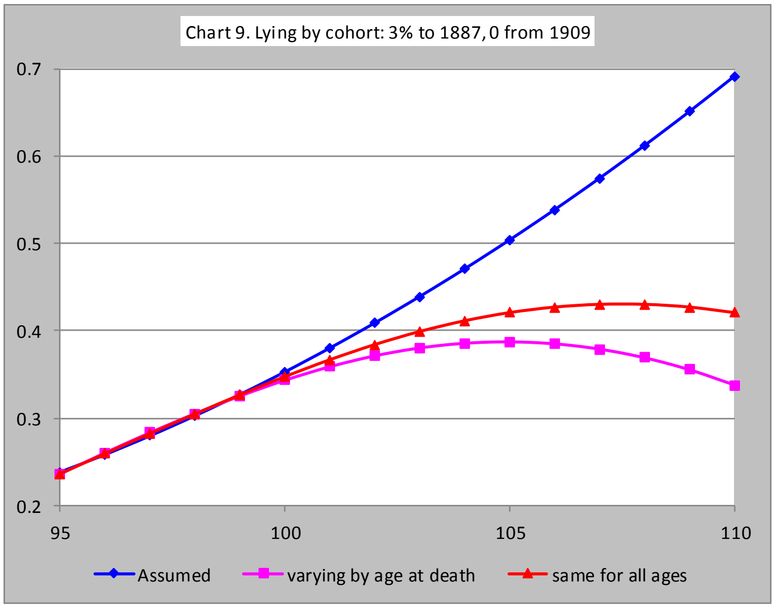
Since a decreasing pattern of lying seems to yield decreasing slope in the mortality curve, let's try varying the length over which the lying decreases. Chart 7 shows the results of the proportion lying being level at 3 percent to 1880 and then decreasing linearly to 1 percent at various years. The clear winner (in getting a negative slope in mortality) is 1910.



It now remains just to do some fine-tuning. The pink line on chart 8 has the strongest downward bend (measured by second central difference at the maximum) that I found with two horizontal line segments and linear interpolation in between. The scenario seems implausible. Could as many as 8.3 percent of deaths for those who were really in birth cohorts up to and including 1884 be overstated by 10 years? The red line is almost as extreme a bend, and it seems much more plausible at only 3 percent of deaths overstated.



There is still one more game to be played. It may be that those who are older at death are more likely to have overstated their age for long enough that it became believed by the ones who report age at death. In chart 9, the red line is the same as the red line in chart 8. The pink in chart 9 has the same base rate for each year of birth, but for those dying at real age 85 or less, the overstatement is only 20 percent of the base rate; for those dying at real age 105 or higher, the overstatement is twice the base rate. The intervening ages are interpolated.

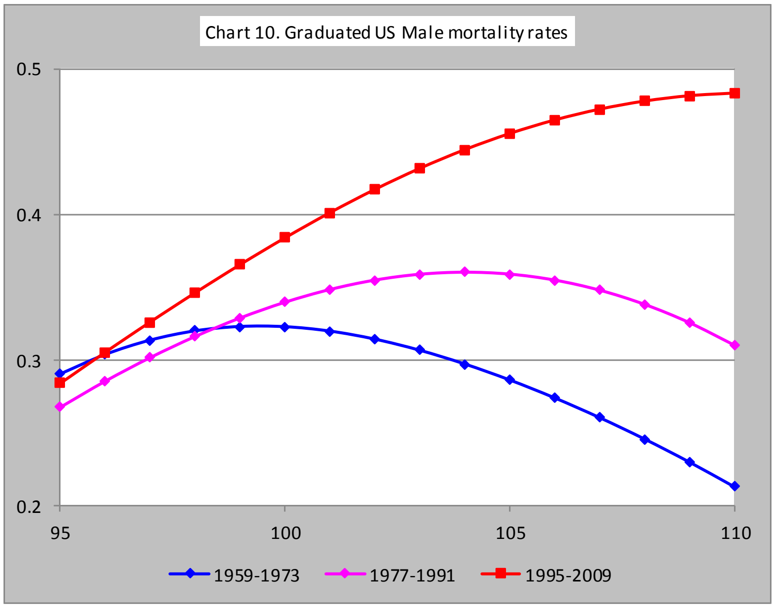


This last scenario seems reasonably plausible. Those born earlier are more likely to overstate their age, and those living longer, within the same year of birth, are also more likely to overstate their age. The observed mortality curve would then depart significantly from the true underlying curve, potentially even to having a negative slope at very high ages.

Of course, this does not prove that the death records in fact contain overstatements. It remains possible that the negative slope is a real feature of the data, but it seems more likely that the negative slope is an artifact of elderly men overstating their age to improve their social status.

If my explanation is valid, then we should see in the actual data (not my model) some variation by observation period. In recent years, it will have become much more reasonable to insist on a valid document to prove date of birth than was the case in the past.

Chart 10 supports my conjecture. The blue line uses deaths reported in 1959–73, the pink line 1977–91, and the red line 1995–2009. Each successive curve shows less downward bend than the one from the earlier period.



3.4 Applications

Few mortality studies by insurance companies or pension administrators have enough data at the oldest ages to make a good estimate of mortality rates. Accordingly, they need to rely on other sources of data. Unfortunately, there is reason to be suspicious of data at the oldest ages. In my experience, the better the data, the more likely the mortality curve is Gompertz. If the observed curve differs materially from Gompertz, it may be wise to look further at the data for inaccuracies.

Those applying for an annuity have a substantial financial incentive to be thought older than they really are. As parents of teenagers often learn, a fake ID is not that hard to obtain. Insurance companies would be wise to insist on good quality proof of age. As chart 3 shows, even 1 percent overstatement of age can have a large impact on the observed mortality at oldest ages.

Pension administrators are unlikely to be prone to this risk. In most cases, they can rely on the date of birth in employment records obtained many years earlier. Except for the very young, if there is a financial incentive to misstate age at time of hire, the incentive would be on the side of a lower age rather than higher.

Note that my example is about a mortality table constructed from death records only. However, the result would be exactly analogous if the date of birth were established at the outset, correctly or incorrectly, and there were immediate reporting of deaths. Although my method is very different from that typically used, I believe the conclusions are equally applicable to more traditional mortality studies.

4 Cheaters: Failure to Report Deaths

There is a second type of action that can upset mortality statistics: failure to report deaths. We usually think of deaths being reported late rather than not all, but a death can be reported so late that its effect is little different from no report.

There was a report in the news a few years ago that the failure to report deaths was shockingly high for elderly Japanese females (McCurry 2010). The motive may have been to continue to receive a government pension, but I don't know of any evidence beyond the news article.

Now it is well known that Japanese females have the longest life expectancy, both at birth and at age 65, of any national group of females. Chart 11 compares the central death rates of several countries to an international average. I calculated the simple average of central death rates over 2000–09 using data posted on the Human Mortality Database, www.mortality.org. The international average is a weighted average of the central death rates for those same countries; the weights are the populations in 2009 at age 65, but to avoid too much impact for the largest countries, I limited the weight to the population of the third largest country. Chart 11 shows that the central death rates for Japanese females are well below those of the other countries.

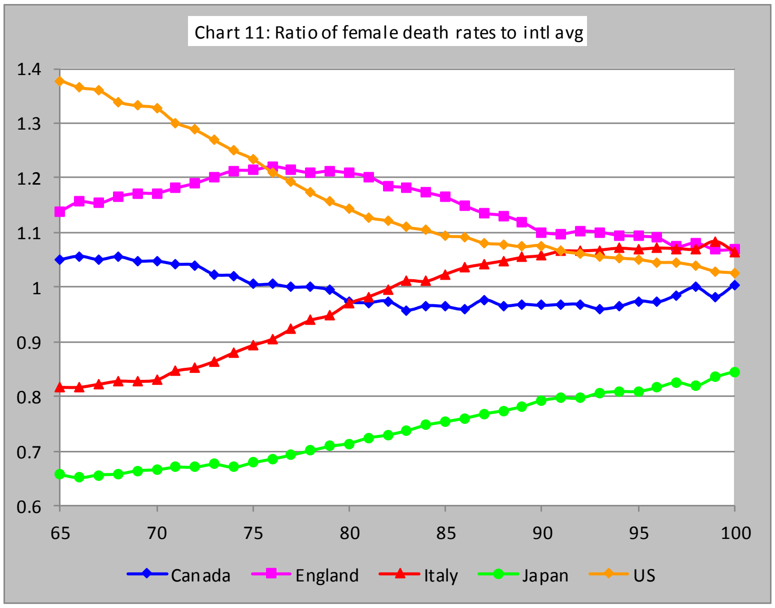
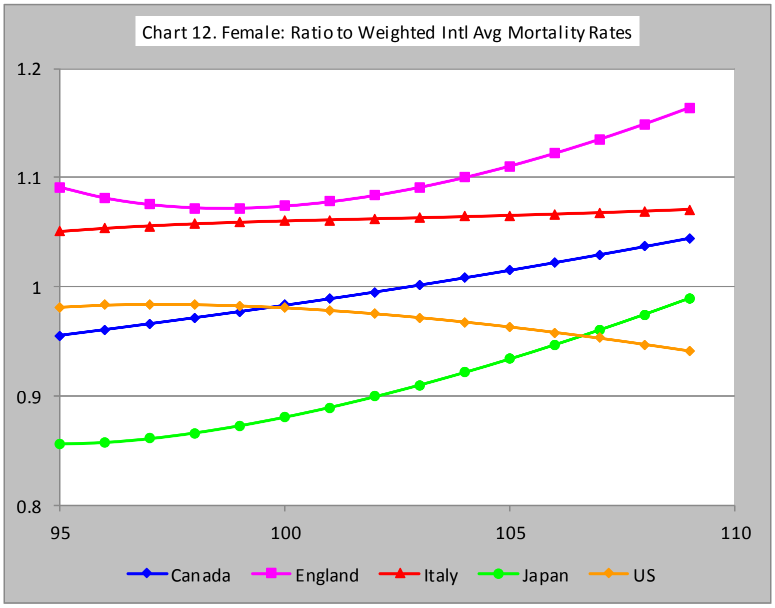


Chart 12 has a similar comparison. The mortality rates are those calculated from death records only using the method in my 2011 paper. Japanese females remain in the lead until a very high age, but the lead is not as large. The two charts overlap for ages 95–100.

If failure to report deaths is prevalent, the line in chart 11 would certainly be lower than it should be because those whose deaths were not reported would continue to be counted as exposure and not as deaths. The line in chart 12 may not be affected. As long as the prevalence of nonreporting was about the same at all ages, the mortality rates calculated from death records alone would not change materially.



The challenge for this section is to determine whether failure to report deaths could explain the very low mortality rates of Japanese females.

4.1 Model

Although the model used to explore overstating age at death was based on U.S. males, there is no reason not to use the same model dataset in this case. There is no need to verify the actual level of mortality. All that matters is whether we see the right shift in the mortality curve. In this case, we are looking for an overall downward shift in the mortality curve. However, I did switch to the female data.

4.2 Simulating Cheaters

I call the failure to report deaths "cheating." Of course, there may not be any desire to cheat the pension plan or insurance company.

I chose to model two approaches. The first, similarly to the approach of the previous section, models the impact on a mortality table developed from death records only. I call this approach D.R. (death records). The second approach models a pension administration system that begins with a list of known lives and records deaths as reported. I call this approach P.A. (pension administration).

Deciding which death not to report is the same for both D.R. and P.A., but adjusting exposure is very different. The exposure for D.R. is adjusted for the year of death and all prior years along the birth cohort. The prior years are adjusted because the reporting of the death is the primary event, which indicates the person had been alive until then. The exposure for P.A. is adjusted after the year of death and all subsequent years along the birth cohort. The later years are adjusted because the life is assumed to be alive until death is reported.

For example, if an actual death at age 88 for 1988 is not reported, then the model's

1. deaths are decreased by 1 at age 88 for 1988
2. D.R. census is decreased by 1 for the birth cohort 1900 for all ages to and including 88
3. P.A. census is increased by 1 for the birth cohort 1900 for all ages 89 and older

4.3 Observations

First, let's consider the impact of failure to report deaths on D.R. Table 1 shows the impact on mortality rates, averaged over the last 15 years for various levels of cheating. The proportion cheating is held constant to age 85 and constant, not necessarily at the same rate, for age 105 and older, with linear interpolation in between.

Table 1. Cheating to age 85, grading to age 105, by DR				
Age	None	2%/2%	2%/0%	0%/2%
95	0.2006	0.1996	0.1998	0.2005
96	0.2209	0.2199	0.2201	0.2207
97	0.2429	0.2418	0.2421	0.2426
98	0.2667	0.2656	0.2660	0.2664
99	0.2924	0.2914	0.2918	0.2921
100	0.3201	0.3191	0.3195	0.3197
101	0.3498	0.3488	0.3493	0.3494
102	0.3815	0.3805	0.3810	0.3810
103	0.4151	0.4141	0.4148	0.4145
104	0.4506	0.4497	0.4504	0.4499
105	0.4879	0.4870	0.4879	0.4870
106	0.5268	0.5259	0.5268	0.5259
107	0.5670	0.5662	0.5670	0.5662
108	0.6085	0.6077	0.6085	0.6077
109	0.6508	0.6501	0.6508	0.6501
110	0.6936	0.6930	0.6936	0.6930

This table shows that the impact of cheating is almost nil for D.R., whether the proportion failing to report deaths is constant, increasing or decreasing with age. Table 1 is not very interesting, but the implications are important.

Because cheating has so little impact on D.R. mortality rates, let's consider only P.A.

Chart 13 shows the impact of a flat rate of failure to report deaths across all ages and years. Note that the impact at higher ages is very large even though the cheating rate is very low.

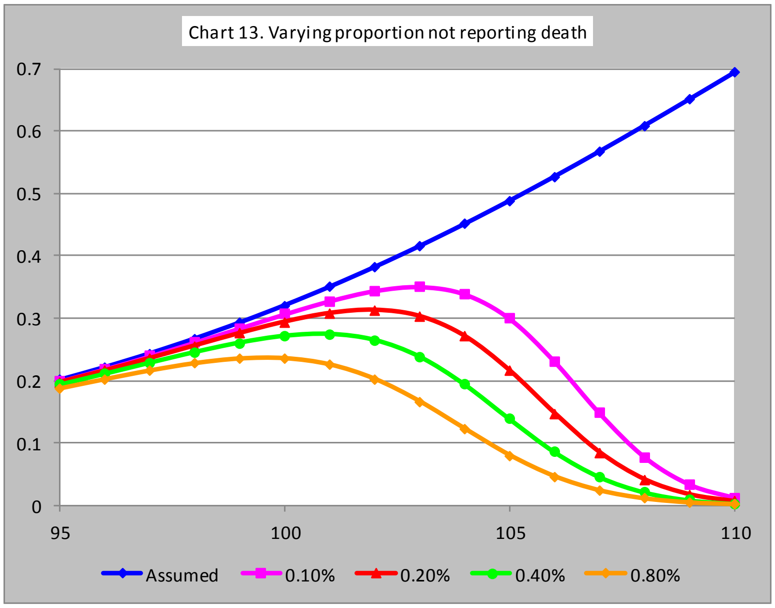
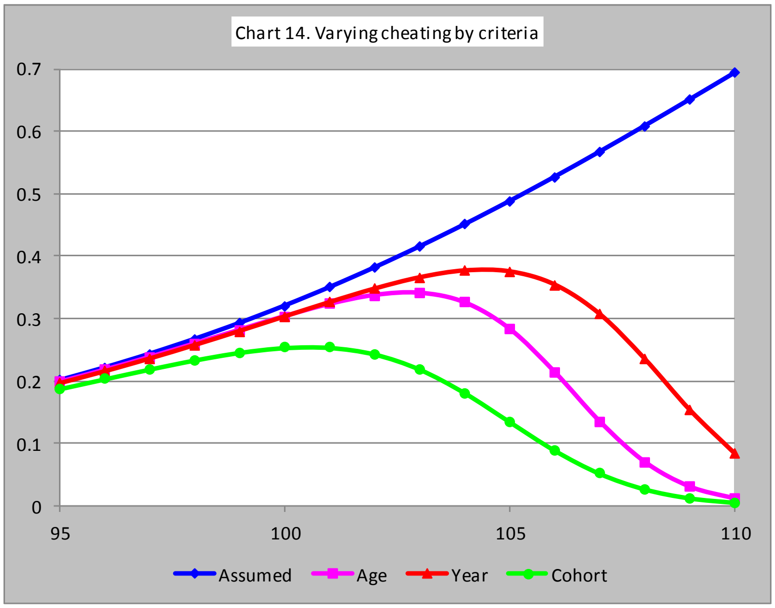


Chart 14 shows the impact of varying the cheating rate by age, year or birth cohort. The line marked "age" uses no cheating up to age 85, then increasing linearly to 1 percent for age 105 and higher. "Year" assumes no cheating to 1990, then increasing linearly to 1 percent for 2005 and later. "Cohort" assumes no cheating for those born on or before 1885, then increasing linearly to 1 percent for 1915 and later. The choices of assumption are arbitrary; the point is to see how the pattern of the cheating rate influences the mortality curve.



Regardless of which criterion changes, the shape of the curve is similar. There is little impact at age 95, and the rates drop precipitously at the highest ages.

My hypothesis was that failure to report deaths of Japanese females explains the overall low mortality rates. Clearly my hypothesis is false. In fact, the pattern of mortality curves is completely different. There may be a lack of reporting, but if so, it has no material impact on the low mortality rates in the 60s to 80s.

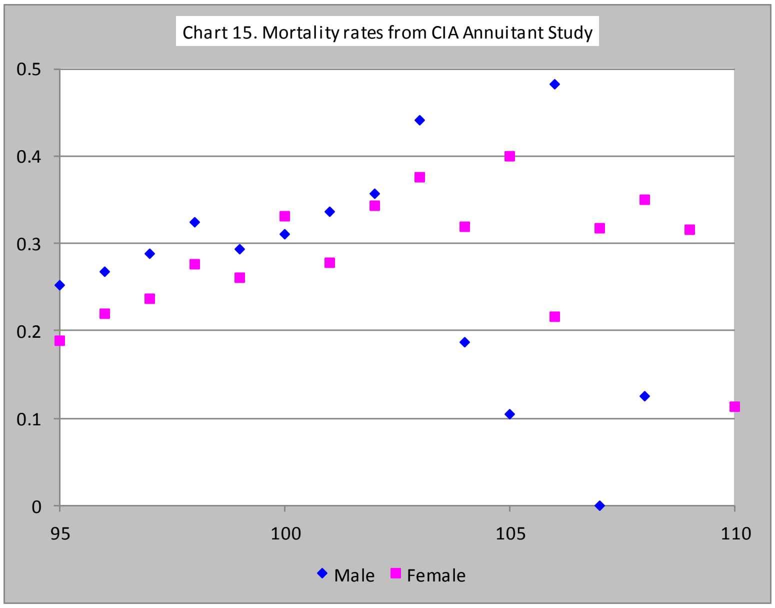
4.4 Application

My hypothesis is false, but there are some important lessons to be drawn.

First, some incompleteness of death reporting will not invalidate a mortality study based on death records only. Because of the need to work with completed cohorts, D.R. will not work well at lower ages, but at the highest ages, it holds much promise of accuracy. It would be worthwhile to examine how many deaths are needed to have credible mortality rates. Because the rates are so very high at these ages, the number of death will not be very large. I explored this a little in my earlier paper, but much more could be done.

Second, more traditional mortality studies using administrative records are prone to mortality rates that are far too low at ages over 100 even with very small incompleteness in death reporting. It is not extremely rare to lose track of an annuitant; losing track of an insured, especially with paid-up insurance, is much more frequent. We have always known that some of those whom we lose are dead and that mortality rates can be understated as a result. It turns out that the secondary result of overstating exposure has a stronger impact on high-age mortality rates. If we see very low mortality rates at extreme ages, we should expect that there has been a lack of reporting of deaths.

For example, consider chart 15. This chart shows raw mortality rates obtained from binary files made available along with "Canadian Individual Annuitant Mortality Experience - Policy Years 2000 to 2009". The rates are measured by policies, for single life annuities only, for years 1989-2009 combined. (The data is available through a link on page 22 of the above document. See references below.) Note that there is clearly a downward trend in the mortality rates at the highest ages. The trend is not as extreme as in chart 13, for example, perhaps because the study specifications call for reporting as a death any policy terminated because the annuitant cannot be located.



5 Procrastinators: Late Reporting of Deaths

Deaths tend to be reported fairly soon for life insurance, most annuities and pensions, but there are always stragglers. A very few may intend to defraud, but most late reports are from those who don't know they need to report a death, don't know to whom to report the death or simply don't get around to it. Keying on that last explanation, I will call those who report deaths late "procrastinators."

In the case of a joint annuity, the report of the first death tends to be delayed far more than the second death or the death of a single life annuity. A delay of 10 years is not uncommon.

While exploring the implications on mortality studies for liars and cheaters, it occurred to me that it might be worth looking at the impact of procrastinators too. Of course, insurance companies and pension funds will often make a bulk adjustment to their liabilities for incurred but not reported (IBNR) deaths, but adjustments in the mortality studies are less common. How material is the impact of not adjusting?

5.1 Model

To use the terms defined in "cheaters," a D.R. model will not be interesting because deaths are eventually reported for the correct age. Therefore, I consider only a P.A. model, and I build on the model used for P.A. with cheaters, except that in this case the death is eventually reported.

There are five ways I have seen late reported deaths handled in a mortality study.

1. As Reported (A.R.): Deaths are included when reported at the age implied at the time of report, not at the actual age at death. The attitude seems to be, "You're not dead until I say you are dead." There is some logic to this approach because it reflects the financial impact if it is not possible to recover payments made after the effective date of death, but it seems inappropriate for the first death of a joint annuity.

2. Corrected Death (C.D.): Deaths are included when reported at the actual age at death. There is no attempt to correct exposure assumed between the actual date of death and the date of report. This approach seems incorrect to me because it will tend to understate mortality rates at higher ages. It may be used when it is thought that the benefit of correcting exposure is less than the cost of doing so.

3. Year of Report (Y.R.): Deaths are included when reported at the actual age at death, and in the year of report negative exposure is included to reverse the impact of exposure previously recorded in error. When composition of the lives under study is experiencing little or no change in its mix of business, as with a stationary population, and there is no material change in the pattern of late reporting, this approach will give correct results.

4. Year of Experience (Y.E.): Deaths are included at the actual age at death and in the actual year of death. Exposure previously recorded in error is reversed for the appropriate year and age. When a death is reported, the study is put back to what it would have been if the death had been reported immediately. This method is appropriate except for one flaw: Deaths still unreported are ignored.

5. Adjustment for IBNR (A.I.): The method is the same as Y.E., but in addition, deaths are increased in the most recent years of experience as an adjustment for those deaths that have not yet been reported. There is typically no adjustment to exposure for IBNR. I would expect this method to be the most accurate, provided that the adjustment for IBNR is reasonably close to emerging experience on late reported deaths.

5.2 Simulating Procrastinators

Table 2. Completeness of reported deaths

Years delayed	Cumulative Reporting
0	85%
1	87%
2	89%
3	91%
4	92%
5	93%
6	94%
7	95%
8	96%
9	100%

I made up a vector of rates for the completeness of reporting, which is shown at left. These rates are artificial. They are not the highest I have seen, but they would be considered fairly high for single life annuities. They are about average for the first death of a joint annuity. To limit the length of the simulation, I assume all deaths have been reported after 10 years.

For each of the five methods of handling procrastinators, I shift the deaths based on the expected year of report, from the rates in table 2. Note that 15 percent of deaths are always removed from the year incurred. Deaths are added back in if the delay does not extend past the last simulated year.

1. A.R.: Exposure is added for each year the report is delayed. Deaths are added at the indicated number of years along the birth cohort.

2. C.D.: Exposure is added for each year the report is delayed. Deaths are added in the year of report at the correct age.

3. Y.R.: Exposure is added for each year the report is delayed. Deaths are added in the year of report at the correct age. Negative exposure is added in the year of report for the ages over which exposure was reported in error.

4. Y.E.: Exposure is added for each year the report is delayed. When the year of report is reached, deaths are added in the correct year at the correct age, and negative

exposure is added along the birth cohort for each year in which exposure was reported in error.

5. A.I.: Exposure and deaths are handled exactly as in Y.E. Deaths are increased in the most recent nine years to reflect the deaths assumed not yet reported. For example, deaths in the last simulated year are multiplied by 1/0.85, in the prior year by 1/0.87, etc. The model uses table 2 for this purpose, but in practice, the assumed rates for IBNR will differ from the actual.

5.3 Observations

Chart 16 shows the resulting mortality rates for assumed and the five methods described above. Mortality rates are obtained by summing deaths and exposure of the last 15 years simulated. The line for A.I. overlaps the line for Y.R. because the two are not materially different. The lines are all close together at age 80, but they deviate noticeably from assumed by age 95.

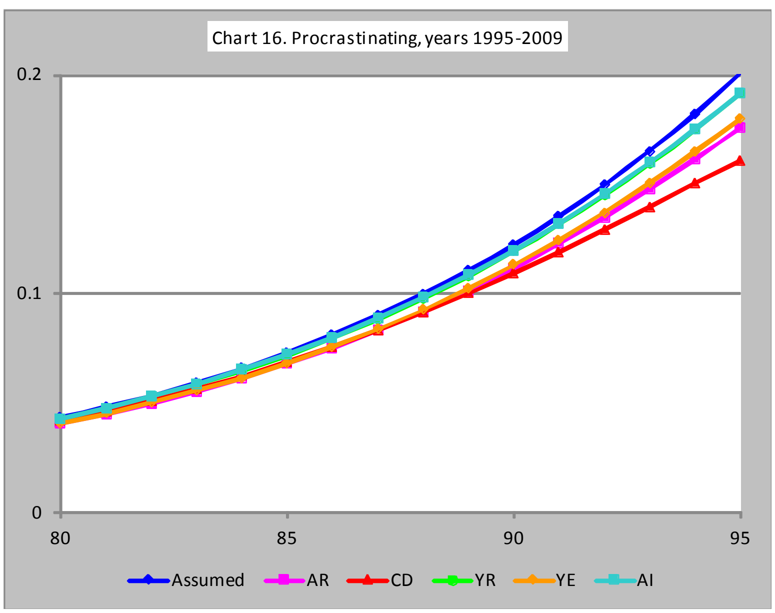


Chart 17 is set 10 years earlier. In this case, all deaths would have been reported. In this case, A.I. and Y.E. are identical because IBNR is zero for years more than 10 before current. They are both also equal to assumed because the adjustments to death and exposure bring the preliminary values back to the actual. A.R., C.D. and Y.R. are all closer to assumed in this chart than in chart 16.

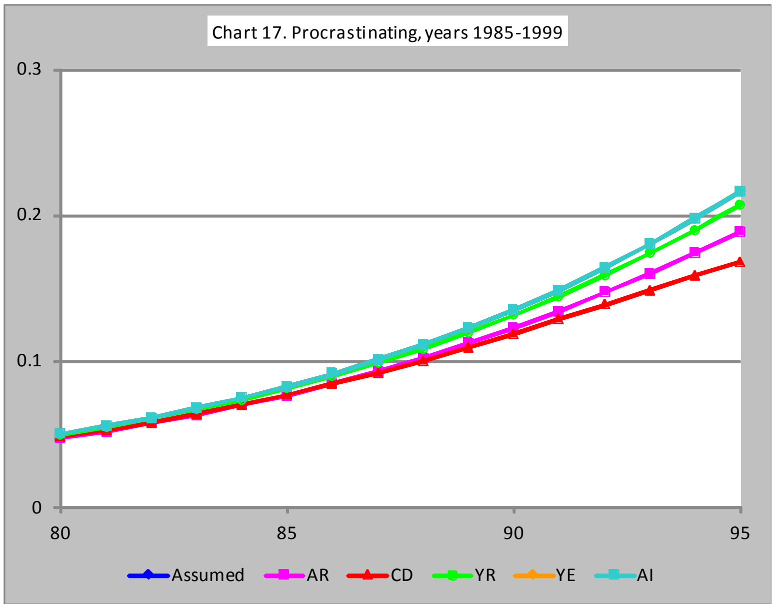
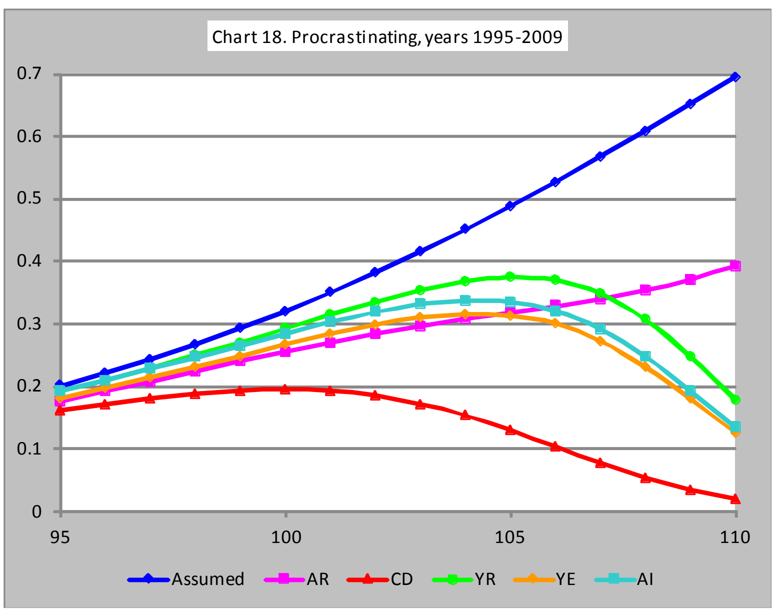


Chart 18 is comparable to chart 16 but for older ages. As we saw with liars and cheaters, the impact of procrastinators is magnified at the highest ages. None is much good over age 100. The reason is likely that there is too much additional exposure passed on to higher ages because of the delay in reporting deaths.



5.4 Application

For ages under 85, it probably does not matter much how late reported deaths are handled, as long as they are not ignored. At higher ages, it becomes increasingly important to make a good adjustment for IBNR. Only Y.R. and A.I. are within 10 percent at age 95, and they fall away by more than 10 percent after age 100 and 99, respectively, given my IBNR factors. (For single life annuities, the factors should be about half those assumed in my model. My testing indicates that these lower rates buys only an additional two to three years of respectability for the method.)

Mortality studies at higher ages should have reliable IBNR rates, and if over age 95, I recommend using a method that adjusts exposures for IBNR as well as deaths.

6 Conclusions

High-age mortality is remarkably easy to upset. Actuaries and demographers need to be wary of misinformation in the data and of mishandling of IBNR. Several studies in the past have shown a flattening of the mortality curve at high ages. The flattening can be explained by errors in the data and is unlikely to be real.

7 References

Howard, Bob. 2011. "Mortality Rates at Oldest Ages." Presented at the Living to 100 Symposium, January, Orlando. http://www.soa.org/library/monographs/life/living-to-100/2011/mono-li11-5b-howard.pdf.

McCurry, Justin. 2010. "Thousands of Japanese Centenarians May Have Died Decades Ago." *The Guardian* (September 10): 23.

Human Mortality Database. University of California, Berkeley. http://www.mortality.org.

Canadian Institute of Actuaries. 2012. "Canadian Individual Annuitant Mortality Experience - Policy Years 2000 to 2009." Canadian Institute of Actuaries, Ottawa, ON. http://www.cia-ica.ca/docs/default-source/2012/212063e.pdf.