Mortality Rates at Oldest Ages

R.C.W. “Bob” Howard, FSA, FCIA

Presented at the Living to 100 Symposium
Orlando, Fla.

January 5-7, 2011

Abstract

Because of a lack of data, the highest age mortality rates in most tables are conjectural. This paper presents a method for using death records to infer exposure on nonextinguished cohorts, thereby allowing the development of a credible table for high ages. The method uses Whittaker-Henderson graduation in a number of unusual ways. The paper also validates the method by applying it to stochastically generated sets of death records for which the underlying mortality and improvement tables are known. There are some surprising results.

1 Introduction

1.1 The Problem

Mortality rates at high ages in mortality tables are notoriously conjectural. It is common practice to decide on the terminal age of the table and to fit a cubic through 1.0 at that age and the rates for the three highest ages that can be obtained from data. This approach may be safe for insurance because few policies are issued at very high ages and the present value at issue of mortality over age 100 is usually negligible. It may be quite inappropriate for annuities.

High age mortality has become much more important in recent years for two reasons. First, there are many more people reaching those high ages than was anticipated in the past due to mortality improvement and the aging of the baby boom, which is present in a number of countries. Second, with interest rates falling over the past 20 years, the discount factor for events 30 or more years into the future is much larger than it was 20 years ago.

The problem is really in the data. Insurance data at high ages is too scanty. Very few people at the highest ages are covered by insurance policies or annuities, and certainly some of the high age exposure may be for people who have died without their deaths being reported. Population data are too inaccurate due to self-reporting in censuses.

1.2 A Source of Data

The website www.mortality.org has collected population mortality data for many years from several countries. One of the datasets is raw deaths, by sex, age and year. Several countries distinguish individual ages to 110 or higher. Death records are likely to be much more accurate than census records. The law in the countries included requires reporting of each death with both date of birth and date of death. And, of course, the dates are not self-reported. Admittedly, the date of birth may not always be correct, but we have reason to expect few inaccuracies.

In most cases, the deaths are reported by Lexis triangle. That is, the deaths for a given age and year are divided between the two possible years of birth.

This dataset seems to hold promise as being sufficiently complete and accurate for constructing a mortality table. The problem is that the exposure data, derived from censuses, are not nearly as good. Accordingly, we must find a way to work with death records alone.

1.3 Definitions

q_x^y means the probability that a person alive on anniversary *x* of his or her birth in calendar year *y* will die before reaching the next anniversary. The death may occur in calendar year *y* or *y+1*. Thus both *x* and *y* are defined at the beginning of the one-year period. Note that this person was born in calendar year *y – x.*

I_x^y means the improvement rate in mortality for individuals age *x* from calendar year *y-1* to calendar year *y.* In this case, *x* applies to the entire one-year period, and *y* is defined at the end of the period.

Thus $q_x^y = q_x^{y-1}(1 - I_x^y)$

In what follows, I don't use the actuarial symbols, but I do refer to a mortality rate or improvement rate for a year. The above definitions are to clarify what I intend.

2 Table Construction from Death Records

2.1 Extinguished Cohorts

For some of the older birth cohorts, it is reasonable to assume that none remain alive. Therefore, the death records for those cohorts detail when all deaths occurred. It is a simple matter to determine exposures by adding in the deaths at each age in turn and from the calculated exposure to determine mortality rates.

How do we know if a cohort is extinguished? We cannot know with certainty, but it seems reasonable to consider the cohort extinguished if there have been no deaths reported in the most recent three years.

2.2 Continuing Cohorts

Much of the data is for cohorts that likely have survivors. To make effective use of the data, we need to know how many are still alive in each cohort at the end of the study.

The method I have chosen is to infer the exposure at each year and age by extending the known mortality rates for the extinguished cohorts with two-dimensional graduation. Clearly this method will not give a precise answer. It will be important to have some sense of the range of possible outcomes consistent with the observed data.

2.3 Note on Graduation

In all cases in this work, I graduate data using the Whittaker-Henderson method, either in one dimension or in two as require by the data. The graduation is described by:

1. The order of difference used for smoothness (referred to as "order");
2. The factor adjusting the balance between goodness of fit and smoothness ("smoothness");
3. The weights used in determining goodness of fit ("weights"); and
4. Whether the graduation is of the rates themselves or the logarithm of the rates. You may assume the former unless I expressly mention logarithms ("logs"). Of course, when logs are used, the graduated logarithms are immediately exponentiated.

If the graduation is two dimensional, there will be both order and smoothness for each dimension.

I always normalize the weights so that they sum to the number of rates being graduated. Doing so helps me to relate one graduation to another, even when the datasets are very different. Usually, the weights are exposures when graduating the mortality rates; this ensures the sum of deaths is unchanged by the graduation.

There are more extensive notes on graduation below in Section 4.3 and by following the link on graduation in Section 7.

2.4 Assumptions

In doing the calculations, I assume:

1. Death reporting is complete;
2. Ages are accurately reported;
3. No migration occurs after age 95;
4. Cohorts are extinguished before reaching age 116 ($q_{115} = 1$); and
5. It is sufficient to have no deaths for three years to be able to declare a cohort extinguished.

2.5 Method in Detail

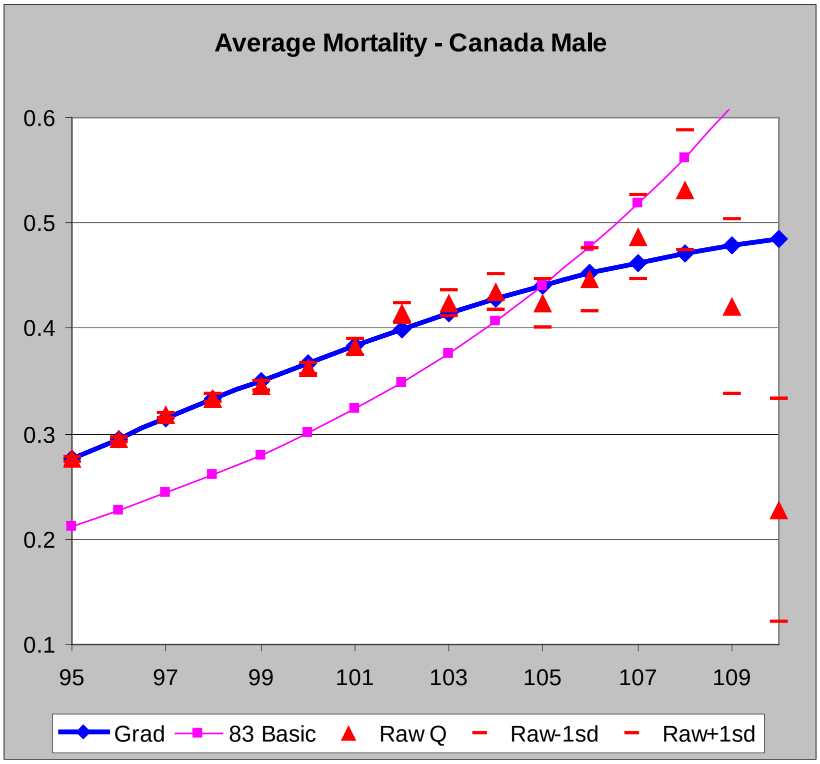
1. Read raw deaths from mortality.org and arrange into Lexis triangles.
2. Add adjacent Lexis triangles to get deaths by age at last birthday for each birth cohort. The result is a matrix very similar to deaths for a "policy year" study period in insurance.
3. Determine which cohorts are extinguished.
4. Smooth the deaths using two-dimensional graduation with order of 3, smoothness of 0.5 for both ages and years, and uniform weights. Then restore the original deaths for all extinguished cohorts. The purpose of this is to reduce the background noise in the data, but the smoothness is small enough that the essential character of the data is unchanged.
5. Calculate the exposure at each age for the extinguished cohorts. Note that what I call "exposure" is in fact the number of lives reaching a specified age during a specified calendar year. My calculations have consistently used 95 as the "youngest" age and 1970 as the earliest year.
6. Calculate an initial set of mortality rates for each year and age. The rates are calculated by attained age using the last 10 cohorts that would have reached age 115 at the end of the period, and graduated by one-dimensional Whittaker-Henderson with order 3 and smoothness 500. The rates at each age are replicated across all years.
7. Estimate the number living for each nonextinguished cohort as of the last birthday within the study period. The estimate minimizes the sum of the squares of the differences between the actual number dying in each year/age and the estimate based of the assumed mortality rates, using the last five years of actual data in each cohort.
8. Graduate the mortality rates over all years and ages using two-dimensional Whittaker-Henderson graduation. I used the logarithm of mortality rates after setting a floor of 0.15. I used the exposure as the weights, but multiplying for all nonextinguished cohorts by 20 percent to give more emphasis on extinguished. I set the smoothness factor fairly high at 1,000 for both dimensions. The order is 2 for years and 3 for ages. The graduated rates are subjected to a floor of 0.15 and a ceiling of 0.75.

9. Repeat the above two steps until the change in the exposure is small in the final year. The measure used is the squared difference in successive estimates of the ending exposure. The iteration continues until the sum of squared error is under 10.

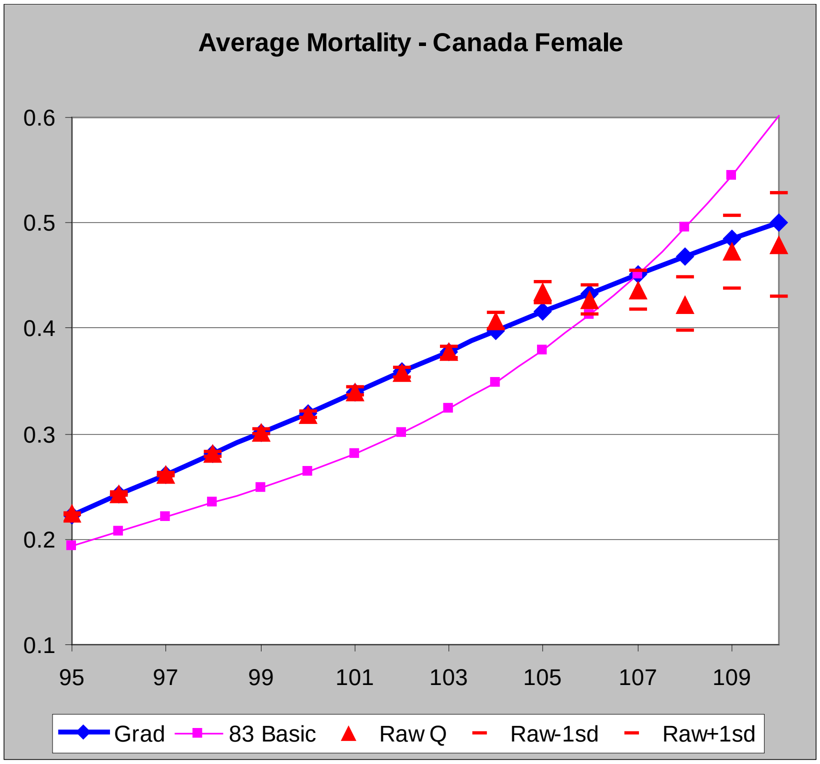
10. The last approximation of exposure is used with actual deaths to calculate the raw average mortality rate for each age. The raw rates are graduated using order 3, smoothness 100 and exposure for weights. The graduation is done over ages 95 to 110. There is too little data at higher ages to justify including them.

2.6 Applying the Method

The following charts show the results of the method for males and females using raw deaths for Canada between 1970 and 2005 for ages 95 and older.



.



In each case, the chart includes the raw mortality rates +/- one standard deviation (in red), the 1983 IAM Basic table for comparison (in pink) and the graduated rates (in blue).

The most striking features are that the curves are much flatter than the 1983 IAM Basic table and that mortality rates continue to rise, close to linearly throughout the range of ages.

Also note how narrow the range of +/- one standard deviation is at the younger ages and how wide at the older ages. The curves could end in many different ways and still be faithful to the data. Indeed, the main difficulty in graduation is that the method does not cope well with data that is very heavy at one end and very sparse at the other. There is a strong tendency for the graduated rates to follow a smooth curve and have little adherence to the data. This is appropriate to the method but not helpful for us.

3 Mortality Improvement

Because I have, after inferring exposures, mortality rates for each age and year, it should be possible to observe rates of improvement in mortality over those same years. Ideally, graduating the mortality rates in two-dimensions would yield improvement rates directly. Unfortunately, the straightforward approach does not yield good results. However, I was able to make some adjustments to the simple approach that appear to be satisfactory.

3.1 Assumptions

1. The deaths and exposure are sufficiently accurate to be used.
2. The underlying actual improvement in mortality progresses smoothly over ages and years.

3.2 Method

I developed the method described here after considerable experimenting with simulated data. See Section 4.4.

1. Split the data into age groups of four each and add the data for these ages together.
2. For each age group, graduate the logarithm of mortality rates using exposure as weights, order 2, and smoothness 1,000 for ages 95 to 98, 3,000 for 99 to 102, 9,000 for 103 to 106, and 27,000 for 107 to 110. This is quite heavy smoothing. When graduating logs with order 2, perfect smoothness is an exponential. By using heavy smoothing, we get close to a least squares fit to an exponential, whose exponent should indicate the average rate of improvement. The result of each graduation is a vector of mortality rates for that age group for 35 years.
3. Calculate improvement rates for each age group.
4. Arbitrarily set the improvement rate for ages 111 to 114 at 0.
5. In a two-dimensional graduation, use the above improvement rates for each age in the age group. (That is, within each year, the rates are five four-year steps of ages.) Graduate with order and smoothness of 2 and 300 for years and 3 and 300 for ages. The weights are the exposures, but I set the age 114 weight artificially at $1/500^{th}$ of the total exposure. This arbitrary weight at 114 has the effect of forcing the improvement rates to be near the value that I specified.

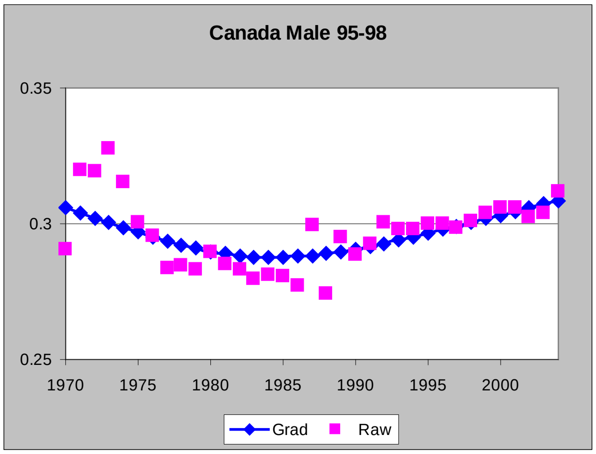
The graduation in the last step is certainly not a traditional application. Its purpose is to smooth out the stepwise rates by age without losing the smoothing by year. Clearly, this is a pragmatic rather than theoretical approach.

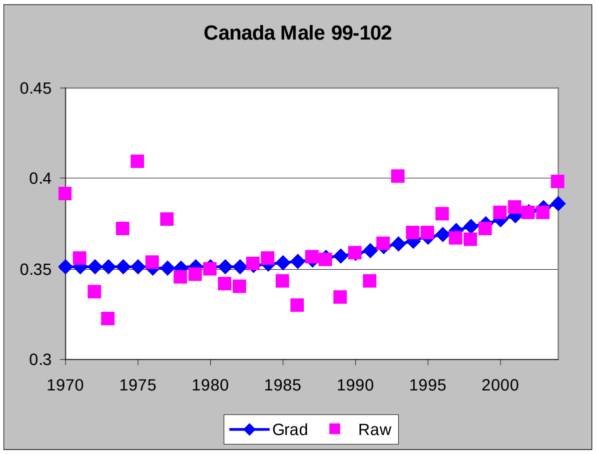
I spent many days searching for a method that would be able to discern the underlying pattern of improvement. Most experiments were disappointing. In the end, I opted for a method that fits the broad trend by year more than by age and that smooths over most of the complexity in the improvement rates.

3.3 Applying the Method

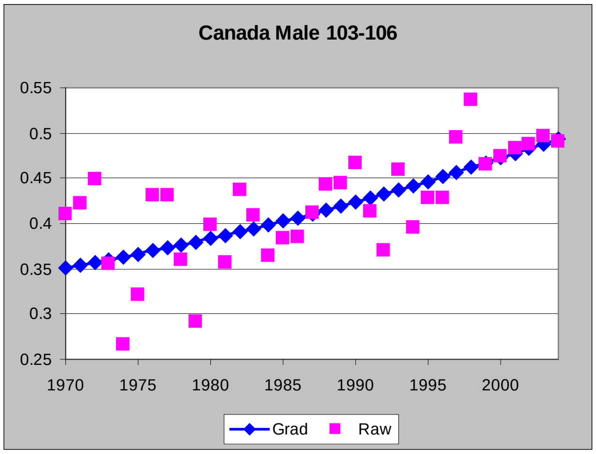
The two charts below show the resulting improvement scales. The numbers are surprising. In both cases, improvement rates were higher in the earlier years and younger ages. In both cases, the improvement rates have been negative in recent years and always at higher ages, in spite of forcing the age 114 rates to be near 0.

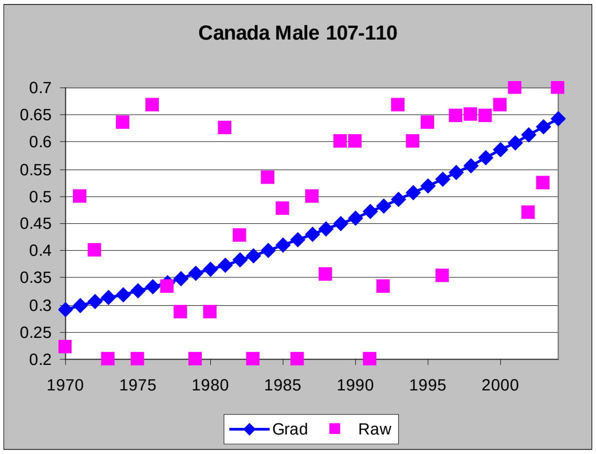
Can these negative improvement rates be believed? Are they an artifact of the method? The following charts compare the raw mortality rates with the graduated for each age band. It seems clear that the deterioration in mortality is evident in each age band and much of the data comes from extinguished cohorts and, thus, not subject to my method of inferring exposures. Bear in mind in looking at these charts that the graduation is very heavily smoothed, particularly at the higher ages.



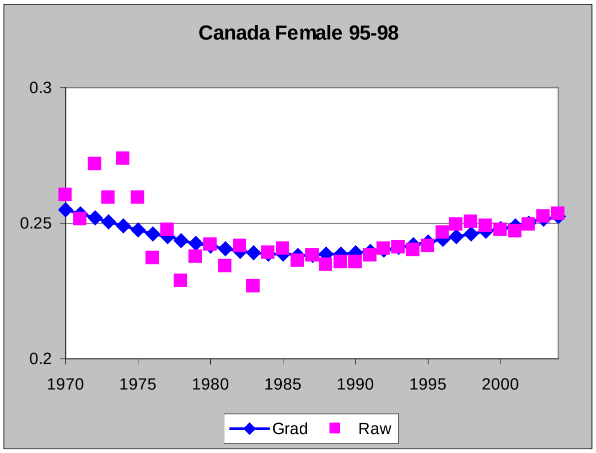


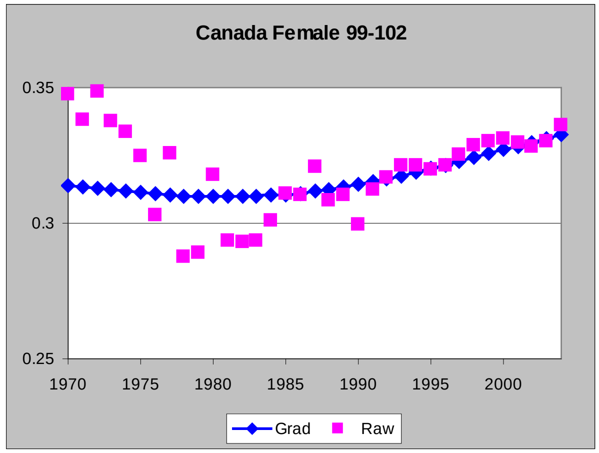
These suggest that there was some mortality improvement until the early 1980s, but mortality has been deteriorating since then. We see a similar picture in the remaining six charts.



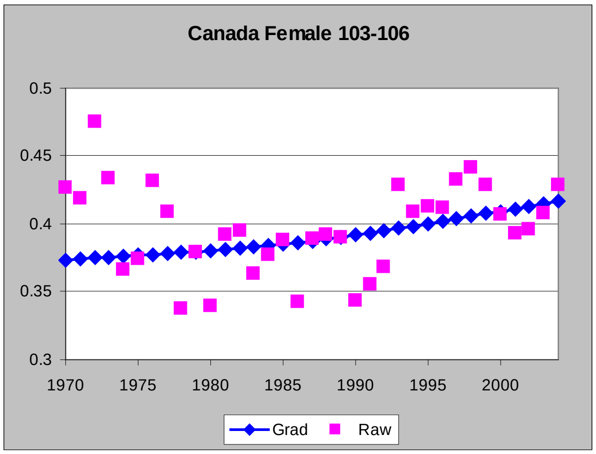


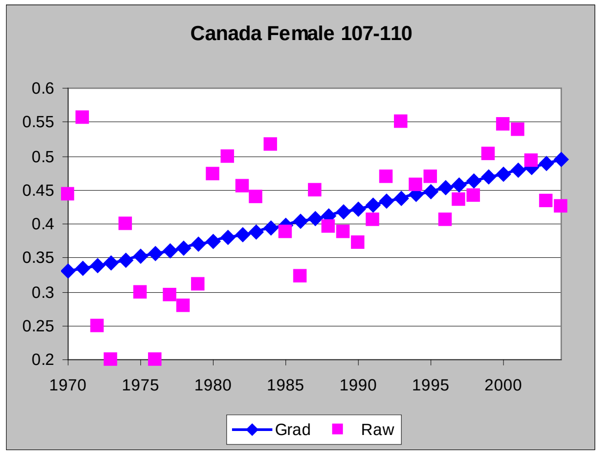
These charts show no indication of mortality improvement in any years. However, the data is much more dispersed. Note that some raw rates show on the graph as 0.2 or as 0.7, which are the floor and ceiling for rates, but the actual rates were outside this range. A non-zero lower limit is essential because the method takes the logarithm of the rates.





There is strong evidence of improvement followed by deterioration.





The evidence for deterioration is probably weakest for females 103 to 106, but clearly there is no support for improvement. Note that these charts show considerably less dispersion than to males at the same ages. The female population is much larger.

Why are we seeing deterioration at these high ages when there is improvement elsewhere? There is nothing observable in the data to suggest a reason. My initial thought was that because of improvements in medical technology that helps people reach age 95, there are many more entrants to the 95 and older age group, but they are on average less robust than was the case formerly. Furthermore, the medical research for those at these high ages has tended to focus more on palliative issues than on increasing longevity. However, later I came to suspect the problem may lie with inaccuracies in the death data. Regardless of the explanation, deterioration in mortality seems to be very solid in the data as observed.

4 Validation

The method described above seems reasonable to me, but can we be sure it will work in practice? I know of no proof as such. However, we can gain considerable comfort with the method by applying it to simulated data because then we can compare the results from the method with what we know to be the hypothetical underlying “actuals.”

The idea is to choose a mortality table, an improvement scale and a starting population, and then to generate randomly many sets of deaths and exposures. I can then apply my method repeatedly to the simulated sets of data and compare the results to the underlying table. (In the following, I call the underlying table the “actual” table.) This allows me to view a distribution of results for each calculated value and to examine the results for individual scenarios.

4.1 Assumptions for Validation Work

4.1.1 Mortality

I did testing with three different mortality tables for ages 95 to 115. The goodness of fit to the real world is not particularly important, but the rates should be not too far off. In general, the rates should be in the range of 0.15 to 0.6. I used the tables generated for males only in my testing. This is an arbitrary choice; it would not have mattered which sex I used.

1. Gompertz.

   I chose a table defined by a Gompertz function as the underlying mortality table because it is a well-known smooth curve considered appropriate for high age mortality, although we know little about the fit at the oldest ages. I wanted to avoid a polynomial function because Whittaker-Henderson is known to deal well with polynomials. Eventually, I found this table would give the best fit. It may be that the curve was so smooth, it did not present any difficult challenges to the method.

   The equation I used is:

$$q_x = e^{-1.15e^{-0.05(x-98)}}$$

2. World.

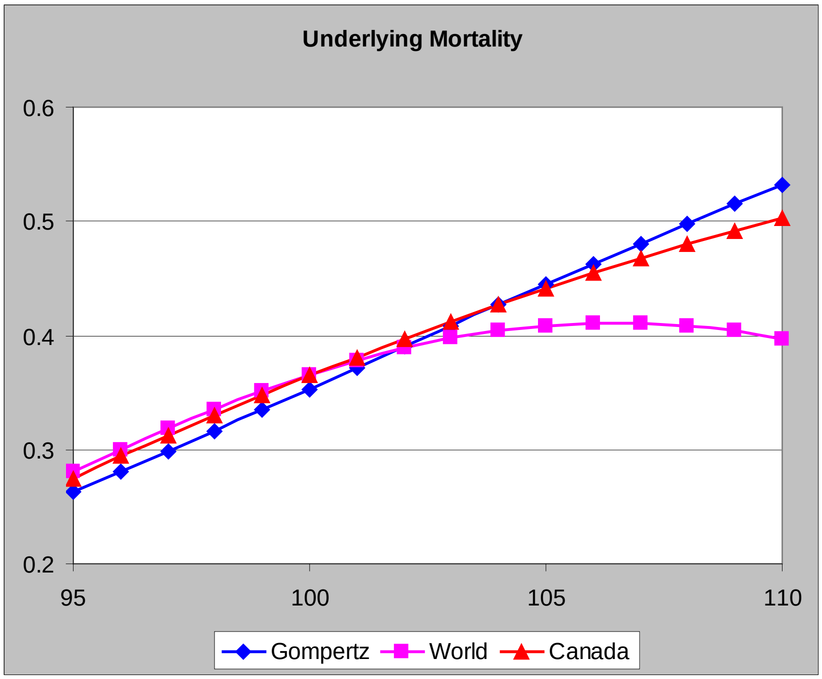
   I developed this table using my method on the aggregate of the seven countries for which I obtained raw deaths—Canada, England (and Wales), Italy, Japan, Sweden, Switzerland and the United States of America. To avoid too much emphasis on one country, I decreased the weight given to Japan and the United States so that those populations would be about the same size as those of England and Italy. I set the mortality rates for the last two ages arbitrarily to 0.5 and 0.8.

   It is an odd feature of the U.S. data, carried into this table, that mortality rates decrease over the last several ages. It seems the negative slope made it more difficult for my method to give a good fit, particularly when combined with a more extreme improvement scale.

3. Canada.

I developed this table using my method on the Canadian data. I used this table because I am most interested in applying my method to Canada. I set the last mortality rate to 0.8.

The three tables are compared in the chart below.



4.1.2 Improvement

I used four improvement scales in testing, focusing especially on the last two. The improvement rates are applied to the mortality rates each year as is done for example with Scale AA.

1. No improvement.

   This is not a likely scenario, but it is a good starting point for testing. If the method does not work with no improvement, it can't be expected to work with a variety of rates of improvement.

2. Flat 1 percent improvement.

   The improvement rate is 1 percent for each year and age, except for age 115, which has no improvement.

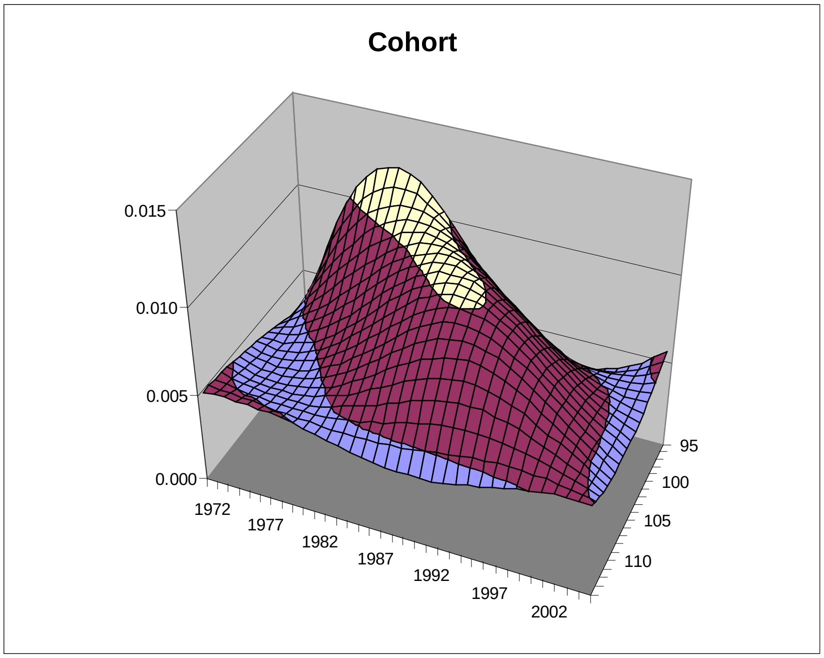
3. Improvement grading by age.

   The improvement rate for age 95 is set at 2 percent grading linearly to 0 at age 115.

4. Cohort effect.

This scale, unlike the others, varies by year as well as age. The point is to test the method with a scale that has sharper variation that the others and has slopes both positive and negative. The maximum improvement rate is 1.2 percent. The minimum is 0.3 percent, except for age 115, which is 0. This scale proved the most difficult to fit well.

The surface chart below shows the improvement rates. Birth cohorts run from back left to front right.



4.1.3 Starting Population

I took as the starting population the numbers calculated on Canadian male data in an early version of my method. Because the simulated population is to represent years of age beginning in 1970 to 2004 and ages 95 to 115, I need those alive on their birthdays in 1970 for ages 95 to 115 and those alive on their 95^{th} birthdays in 1970 to 2004. I intentionally used Canadian males as a smaller population relative to those to which I will apply the method. If the method works well for this population, it should be better for Canadian females and for either sex of countries such as the United Kingdom and the United States.

From these starting values, I generate randomly deaths for each year and age and by subtraction those alive at the next birthday. The deaths are binomial variants based on the number alive at the start of the year of age and the mortality rate for that age-year.

4.2 Inferring Exposure

I generated 100 simulated sets of deaths and exposures for male ages 95 to 115 and for years 1970 to 2004 using each of the above mortality tables and improvement scales. For each set, I inferred a set of exposures from the deaths alone.

To assess the accuracy of the inferred exposure, I calculated the average mortality rate for each age on both the simulated and the inferred exposures to a maximum of 110. There is too little data at higher ages to be useful in this test.

The following tables are based on the Canada table. Each of the improvement scales is shown separately.

The table below compares the actual mortality rates, the raw mortality rates based on simulated deaths and exposures, and the mortality rates based on simulated deaths and exposures inferred by my method, all weighted by the corresponding exposure in each year. (I refer to the three sets of mortality rates as actual, simulated and inferred.) Note that the three mortality rates are very close until age 110.

Some comments are warranted for age 110. The large standard deviation for both simulated and inferred is not comforting. In effect, this says we can be confident the mortality rate lies in the range 0.2 to 0.8, which we could have said without doing any calculations. However, these are raw mortality rates. Graduation is the next step. There are so few lives at this age that statistical fluctuation will be the predominant feature.

I conclude that, although the method will not find the underlying mortality table, the noise level is consistent with what we should expect in the population itself, and the inferred exposures are a satisfactory representation of what we could obtain from an accurate census.

	Average Mortality Rates weighted by exposure, no improvement							
	95		100		105		110	
	μ	σ	μ	σ	μ	σ	μ	σ
Actual	0.275	0.000	0.365	0.000	0.441	0.000	0.503	0.000
Simulated	0.275	0.002	0.365	0.006	0.440	0.027	0.513	0.143
Inferred	0.275	0.002	0.365	0.005	0.441	0.024	0.522	0.134
Inf/Sim	1.000	0.003	1.000	0.010	1.002	0.028	1.030	0.103
Inf/Act	1.000	0.007	1.001	0.014	0.999	0.054	1.038	0.266

A more significant test of the fit takes into account the financial values that we calculate from the mortality rates. The following table shows net single premium immediate annuities and net single premium life insurance at 4 percent based on the mortality tables. The simulated and inferred are very close to the actual. The inferred has slightly more dispersion but not enough to be of any concern.

	Financial Values, no improvement			
	a_{95} at 4%		A_{95} at 4%	
	μ	σ	μ	σ
Actual	1.9931	0.0000	0.9233	0.0000
Simulated	1.9918	0.0085	0.9234	0.0003
Inferred	1.9901	0.0117	0.9235	0.0005
Inf/Sim	0.9992	0.0037	1.0001	0.0003
Inf/Act	0.9985	0.0059	1.0001	0.0005

However, the foregoing has looked only at scenarios with no mortality improvement. I have found little evidence of significant improvement at these very high ages, but there is usually some variation from year to year. A method that works only with no improvement cannot be considered acceptable.

The following two tables are derived from scenarios with 1 percent improvement for each year and age. Note that the standard deviation for actual mortality rates is no longer 0 because the rates vary by year and are weighted by the simulated exposure, which varies from scenario to scenario. The ratio between inferred and simulated has standard deviations a little higher than with no improvement, but the inferred to actual are slightly better.

	Average Mortality Rates weighted by exposure, 1% improvement							
	95		100		105		110	
	μ	σ	μ	σ	μ	σ	μ	σ
Actual	0.255	0.000	0.331	0.000	0.392	0.000	0.440	0.000
Simulated	0.255	0.002	0.332	0.005	0.395	0.022	0.454	0.095
Inferred	0.255	0.002	0.332	0.004	0.396	0.018	0.463	0.083
Inf/Sim	1.002	0.005	1.000	0.010	1.004	0.032	1.037	0.137
Inf/Act	1.001	0.009	1.001	0.013	1.009	0.045	1.053	0.190

	Financial Values, 1% improvement			
	a_{95} at 4%		A_{95} at 4%	
	μ	σ	μ	σ
Actual	2.1922	0.0000	0.9157	0.0000
Simulated	2.1937	0.0094	0.9156	0.0004
Inferred	2.1903	0.0153	0.9158	0.0006
Inf/Sim	0.9985	0.0053	1.0001	0.0005
Inf/Act	0.9991	0.0070	1.0001	0.0006

The following two tables summarize the results for improvement varying linearly from 2 percent at age 95 to 0 at age 115. The results are fairly similar to the previous. It is worth noting that the inferred age 95 rates show essentially the same variability with 2 percent improvement as with 1 percent or none.

	Average Mortality Rates weighted by exposure, age effect 2-0%							
	95		100		105		110	
	μ	σ	μ	σ	μ	σ	μ	σ
Actual	0.239	0.000	0.312	0.000	0.387	0.000	0.469	0.000
Simulated	0.239	0.002	0.312	0.005	0.388	0.020	0.471	0.083
Inferred	0.240	0.002	0.313	0.004	0.388	0.018	0.472	0.078
Inf/Sim	1.001	0.005	1.002	0.011	0.999	0.029	1.009	0.100
Inf/Act	1.002	0.009	1.003	0.014	1.000	0.047	1.006	0.166

	Financial Values, age effect 2-0%			
	a_{95} at 4%		A_{95} at 4%	
	μ	σ	μ	σ
Actual	2.3551	0.0000	0.9094	0.0000
Simulated	2.3539	0.0097	0.9095	0.0004
Inferred	2.3513	0.0162	0.9096	0.0006
Inf/Sim	0.9989	0.0056	1.0001	0.0006
Inf/Act	0.9984	0.0069	1.0002	0.0007

The toughest test is with improvement having a cohort effect because there are both positive and negative slopes and steeper slopes than in the tests mentioned above. The following two tables show the fit is essentially the same with a cohort effect as with the other scales.

	Average Mortality Rates weighted by exposure, cohort effect							
	95		100		105		110	
	μ	σ	μ	σ	μ	σ	μ	σ
Actual	0.265	0.000	0.337	0.000	0.399	0.000	0.459	0.000
Simulated	0.264	0.002	0.338	0.005	0.400	0.018	0.450	0.090
Inferred	0.264	0.002	0.336	0.005	0.397	0.018	0.445	0.084
Inf/Sim	0.998	0.004	0.995	0.010	0.994	0.028	0.996	0.107
Inf/Act	0.997	0.008	0.996	0.014	0.996	0.044	0.970	0.183

	Financial Values, cohort effect			
	a_{95} at 4%		A_{95} at 4%	
	μ	σ	μ	σ
Actual	2.1222	0.0000	0.9184	0.0000
Simulated	2.1220	0.0086	0.9184	0.0003
Inferred	2.1290	0.0138	0.9181	0.0005
Inf/Sim	1.0033	0.0047	0.9997	0.0004
Inf/Act	1.0032	0.0065	0.9997	0.0006

I conclude that the method of inferring exposure has little internal bias and gives acceptable results.

By the way, the method described here is not the first one I tried. I used the tests illustrated here on other methods and rejected those methods because they failed to give good enough results. My first attempt was to have successive approximations of the average mortality rates without using a two-dimensional graduation. That method was fine with no improvement, but it failed with improvement.

I also used these tests to refine the parameters for the process: how much weight to give exposure for nonextinguished cohorts, what smoothness to use, what order and how many years to use when estimating the ending lives for nonextinguished cohorts. The results are not strongly sensitive to most of these parameters; the choices represent small improvements in the fit. One exception is the order of difference to be used in calculating smoothness. I found 2 for years and 3 for ages to be the best and other choices did not yield nearly as good results.

4.3 Graduating Mortality Rates

Graduation attempts to smooth out random fluctuations in the data to discern the underlying function. When there is a very large amount of data, graduation is scarcely needed. When there is very little data, the fluctuations are so significant that it is not possible to identify any relationship to an underlying function. Graduation is for the cases in between.

One of the most difficult graduation exercises is on a set of data that is heavily weighted to one side with very little on the other. This is exactly the case with high age data. There is a lot of data for the youngest of these ages, but little at the oldest ages. A method that uses some weighting from the data (the preferred approach) will put heavy emphasis on the younger ages and largely ignore the older. An unweighted method will give far too much emphasis on the older ages. The challenge is to find a good balance between fidelity at the younger and a reasonably smooth progression into the older.

Probably the ideal way to handle such a set of data would be by fitting a known curve to the data. The problem is that we do not know what curve to use. My investigations have suggested to me that variations by country are wide enough it is unlikely there is a generalized curve that will fit well to mortality within a wide range of countries. It is even less likely there is a curve that could also represent the change in mortality over time.

I have found Whittaker-Henderson graduation to give good results. It has the advantage of allowing direct control over goodness of fit and smoothness. It allows weighting by exposure or other criteria. A software package exists that can readily be integrated into Excel and other tools.

Setting up Whittaker-Henderson involves a number of decisions.

1. What ages to graduate. Because I am interested in mortality at very high ages, I decided to start at age 95. I suspect that using a lower starting age might make it more difficult to have a measurable fit at the oldest ages because too much exposure would be at the lower ages. I trap data to age 115 (for Canada and some other countries, but not all), but I found there is too little data to allow me to go that far. The results look believable to age 113, but the implied standard deviation is very large.
2. What form to graduate. Normally I graduate the rates themselves. Some applications work better with graduating the logarithm of the rates. When there is less data, I have found good results in graduating the ratios of the rates to a standard table. In this case, I will graduate the rates themselves.
3. What weights to use. There is theoretical support for using the reciprocal of the variance as the weights. More commonly exposures are used as weights; the work is simpler, and there is the advantage that the sum of the deaths and the average age at death are kept constant through the graduation, if graduating the rates themselves. If graduating ratios to a standard table, the same advantage is realized by using the expected deaths on the standard table as the weights. I normalize the weights. That is, I multiply them by a factor so the sum of the weights is the number of rates being graduated. Normalizing the weights helps me to find the smoothing factor more readily because two dissimilar graduations will have totals close to each other.
4. What order of difference to use. This depends on the curve we expect to find. If using order 3, then the curve with perfect smoothness is a parabola; if order 4, a cubic. If graduating the logarithms of the rates and using order 2, it is an exponential. Because there is very little data at the oldest ages, the graduated rates at those ages may tend more to the smooth curve than to the raw data.
5. What smoothness factor to use. Here it is a matter more of art than science. I have found it helpful to try many values and examine both the graduation statistics (measures of fit and smoothness) and a graph of the graduated rates compared to the raw. Most of the work is in finding a suitable smoothness factor. The decisions are not as linear at this list suggests. It is often necessary to try different combinations to optimize the whole process.

For this test, it was easy to decide to graduate raw mortality rates using exposures for weights. I graduated ages 95 to 110; there seems too little data at higher ages for the rates at those ages to be meaningful.

The order of difference is more difficult to decide. It turns out my male Canada table, which I use most often in testing, is close to linear. As a result, there appears to be closer fit to the underlying actual table when I use 2nd order difference. Because I should not count on the underlying table being linear, I decided 3rd differences would be better in general.

The following table shows the result of trials for various values of the smoothness factor, traditionally called "h." The numbers shown are means over 100 scenarios, in this case for ages 95 to 110 and for 35 years, for the Canada table with a cohort improvement scale. Over a wide range of values of h, there is little variation in the mean fit to the raw mortality rates. There is little worsening of the fit for h > 300. Smoothness is probably satisfactory for h > 100. We would not normally have access to actual data, but it is interesting to note that with higher values of h, we get a better fit to the actual data.

h	Fit to raw	Smoothness	Fit to actual
1	1.82E-04	1.63E-05	2.07E-04
3	2.01E-04	4.50E-06	1.94E-04
10	2.18E-04	1.09E-06	1.82E-04
30	2.31E-04	3.09E-07	1.73E-04
100	2.44E-04	6.30E-08	1.67E-04
300	2.52E-04	1.03E-08	1.66E-04
1000	2.56E-04	1.09E-09	1.66E-04
3000	2.58E-04	1.27E-10	1.66E-04
10000	2.58E-04	1.16E-11	1.66E-04

A reasonable compromise is to use h=300, but h=100 also looks good. I prefer to err on the side of more smoothness because the underlying actual curve is likely to be smooth.

The following table is very similar to the one earlier for the inferred mortality rates, but this one shows graduated rates.

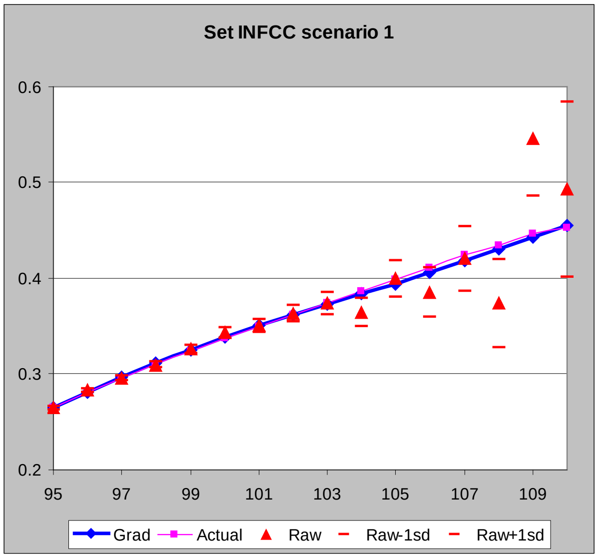
	Average Mortality Rates weighted by exposure, cohort effect							
	95		100		105		110	
	μ	σ	μ	σ	μ	σ	μ	σ
Actual	0.265	0.000	0.337	0.000	0.399	0.000	0.459	0.000
Simulated	0.264	0.002	0.338	0.005	0.400	0.018	0.450	0.090
Graduated	0.264	0.002	0.336	0.003	0.396	0.010	0.447	0.030
Grad/Sim	0.999	0.005	0.996	0.014	0.993	0.041	1.033	0.239
Grad/Act	0.998	0.007	0.997	0.009	0.993	0.025	0.974	0.066

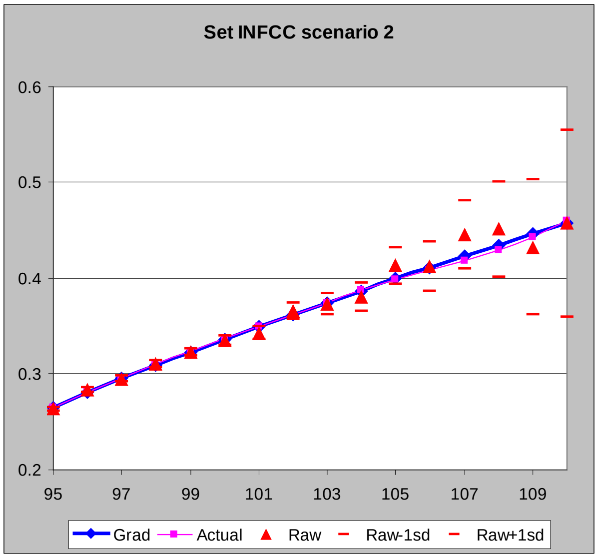
The average graduated mortality rates are almost identical to the simulated mortality rates. What is more significant is that the standard deviations are lower for graduated than for simulated. This means the graduated mortality rates are more likely to be a good representative of the underlying actual rates, not just on average, but also for any particular scenario. The standard deviation is especially improved at age 110; this is important because the standard deviation for simulated is rather high.

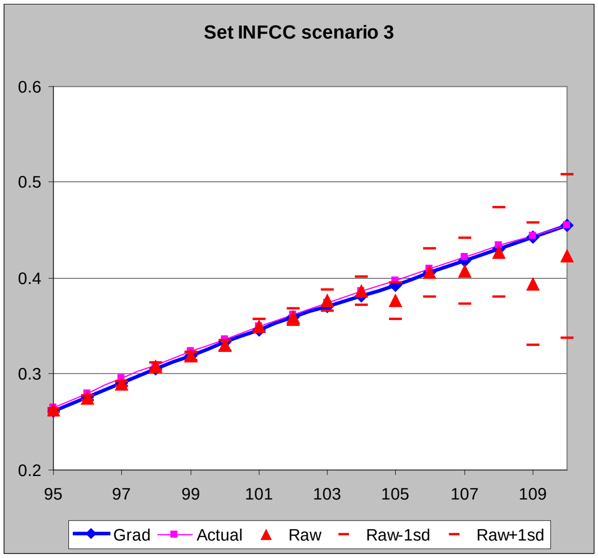
How good is the graduation for a scenario? The following charts show the first four scenarios using the Canada base table and the cohort improvement scale. The red triangle and bars indicate the raw mortality rates +/- one standard deviation.

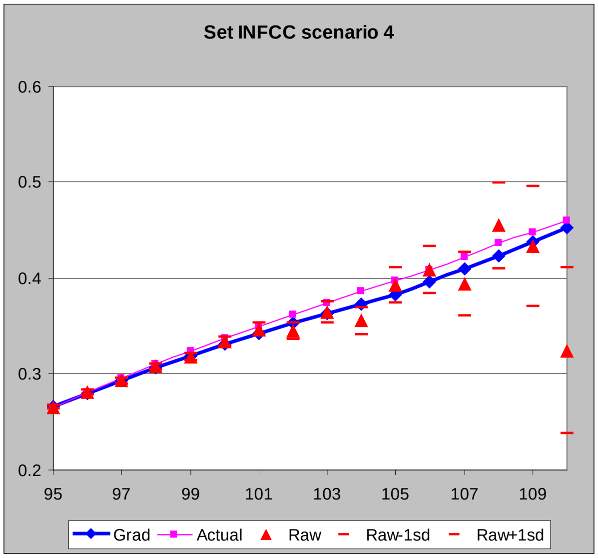
The graduated mortality rates (without arbitrary adjustment) seem fairly close to the actual but tend to wander a bit low at the highest ages. This probably results from the fact that because the order of difference is 3, there is a tendency toward a parabola, although the actual seems closer to linear.

Note that these charts are for the cohort improvement scale. ("INFCC" indicates that exposure is inferred, the base table is "Canada" and the improvement scale is "Cohort".) When the actual improvement scale is smoother than my cohort scale, the fit is better.









4.4 Graduating Improvement Rates

I have generally had good results by applying two-dimensional graduation to data by age and year to get improvement rates. The results are very poor in this case. The problem likely stems from highly skewed exposure and too little data at the oldest ages. I tried a lot of variations in approach, such as using logarithms of mortality rates, varying the weights and using actual/expected ratios rather than mortality rates. None of these gave a fit consistently close enough to be acceptable.

I hypothesized that the difficulty in graduation relates mostly to the rapidly decreasing exposure as age increases. The higher ages then need more smoothing than the lower ages. This suggests a method for which I know no theoretical justification. Nonetheless, it seems to give consistently better results than other methods I have tried. This is an area that warrants further research.

The method is described in Section 3.2.

It is possible this approach could be generalized by replacing the scalar smoothing factors with vectors. My software would have to be revised, but I can see this as a fairly straightforward extension of both one-dimensional and two-dimensional Whittaker-Henderson. This is a subject for further research.

I failed to find a method that seemed to be able to reveal the underlying actual improvement scale in all cases. The cohort scale was particularly difficult. I eventually decided it is best to have a method that would perform well with a scale by age only. Over the long periods of time that I have examined, improvement varying by age seems more common than by year. Cohort effects were unknown until recently. It may be that the rapid decrease in exposure by age makes identify variations other than by age not feasible. A question worth considering is how much data is needed to have adequate discrimination in finding the improvement scale.

The following table shows the mean difference between the actual improvement rates and the calculated ones over 100 scenarios, based on inferred exposures. The scenarios use my Canada base table and the improvement scale varying by age.

	Mean difference from actual			
	1975	1985	1995	2005
95	-0.09%	-0.09%	-0.09%	-0.10%
100	-0.07%	-0.07%	-0.06%	-0.07%
105	-0.07%	-0.05%	-0.05%	-0.05%
110	-0.06%	-0.04%	-0.04%	-0.03%

The mean differences from the actual table are small, although it is a concern that they are all negative. This implies that my method estimates on the low side.

The following table shows the standard deviations for the above differences. They are small enough to indicate the method is very consistent in its estimates. It is remarkable to see the standard deviations so low; there is considerable variation in the simulated data. I take this as an indication the method is smoothing out the variation, as intended.

	Standard deviation in difference			
	1975	1985	1995	2005
95	0.08%	0.05%	0.11%	0.15%
100	0.06%	0.04%	0.08%	0.11%
105	0.03%	0.03%	0.07%	0.09%
110	0.02%	0.04%	0.06%	0.11%

Based on the runs shown above, the method appears to be validated. However, the next set of tables is for my cohort improvement scale. The fit is not nearly as good.

	Mean difference from actual			
	1975	1985	1995	2005
95	0.03%	-0.09%	0.32%	-0.01%
100	0.13%	-0.36%	-0.11%	0.09%
105	0.03%	-0.41%	-0.39%	0.11%
110	-0.21%	-0.36%	-0.43%	0.00%

These mean differences are probably borderline on whether they are good enough to use. The next table shows that the standard deviations remain small. That suggests the method is imposing a pattern on the data rather than discerning what is in the data. Indeed, I concluded it was necessary to do just that because I could not find a way to deal with the varying slopes of the cohort scale.

	Standard deviation in difference			
	1975	1985	1995	2005
95	0.09%	0.06%	0.09%	0.12%
100	0.07%	0.05%	0.08%	0.09%
105	0.04%	0.04%	0.07%	0.09%
110	0.03%	0.04%	0.07%	0.10%

My conclusion is that the method I have outlined is capable of inferring an improvement scales from the data, provided that the underlying actual scale varies primarily by age. A method that could be wrong by 0.5 percent is far from satisfactory. However, given how little understanding of mortality at the oldest ages we have had, this rough method can still be considered an improvement.

4.5 Validating the Calculations on Real Data

There is one last test worth performing, and that is to use the mortality tables and improvement scales developed above for males and females to generate random deaths and then to apply the method to the random data.

The table below, for females, shows the average mortality rates calculated are very close to the starting rates. The rates for ages 95 to 100 should be accurate to almost three significant digits. The higher ages are not quite as good but certainly acceptable.

	Average Mortality Rates weighted by exposure, Canada Female							
	95		100		105		110	
	μ	σ	μ	σ	μ	σ	μ	σ
Actual	0.229	0.000	0.327	0.000	0.423	0.000	0.508	0.000
Simulated	0.230	0.001	0.327	0.003	0.424	0.012	0.510	0.056
Graduated	0.229	0.001	0.326	0.001	0.422	0.005	0.515	0.015
Grad/Sim	0.996	0.003	0.996	0.009	0.996	0.027	1.020	0.109
Grad/Act	0.997	0.004	0.996	0.004	0.998	0.011	1.014	0.030

The male results are not quite as good. The means are about as close to the actuals as for females, but the standard deviations are larger, as should be expected because the male population is smaller than the female. The only concern that might be expressed is for the rate at age 110, but fortunately its financial significance is very low.

	Average Mortality Rates weighted by exposure, Canada Male							
	95		100		105		110	
	μ	σ	μ	σ	μ	σ	μ	σ
Actual	0.284	0.000	0.376	0.000	0.449	0.000	0.491	0.000
Simulated	0.284	0.002	0.376	0.006	0.448	0.023	0.506	0.130
Graduated	0.283	0.002	0.375	0.003	0.449	0.009	0.505	0.028
Grad/Sim	0.998	0.004	0.997	0.015	1.003	0.048	1.079	0.372
Grad/Act	0.997	0.006	0.997	0.007	1.000	0.020	1.028	0.057

The improvement rates are not nearly as good. The female results are shown below. The difference is worse for the early ages and years, compensated for in the later years at the same ages. The method tends to flatten the improvement rates over the years, and that is what is seen here. On balance, these are probably good enough to use.

	Mean difference from actual			
	1975	1985	1995	2005
95	-0.35%	-0.11%	0.15%	0.17%
100	-0.24%	-0.05%	0.16%	0.18%
105	-0.14%	0.01%	0.15%	0.18%
110	-0.04%	0.05%	0.14%	0.17%

The male improvement results are very similar to those of the females, except the standard deviations are generally a basis point or two higher.

	Mean difference from actual			
	1975	1985	1995	2005
95	-0.36%	-0.08%	0.14%	0.12%
100	-0.23%	0.00%	0.17%	0.16%
105	-0.10%	0.06%	0.19%	0.19%
110	0.01%	0.12%	0.19%	0.20%

4.6 Conclusions about the Validity of the Method

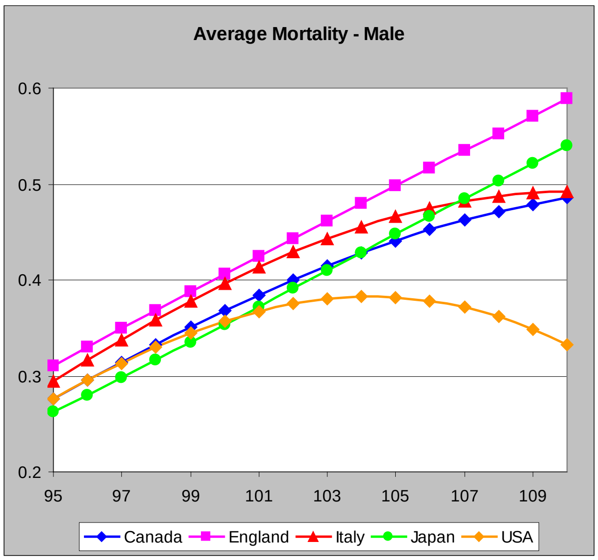
My conclusion is that the method yields mortality rates good enough to be of practical use. The standard deviations are larger than desirable, but given that the rates at these high ages tend to have little financial impact, some variation is tolerable. I believe the method gives a more reliable table at these high ages than we have had before.

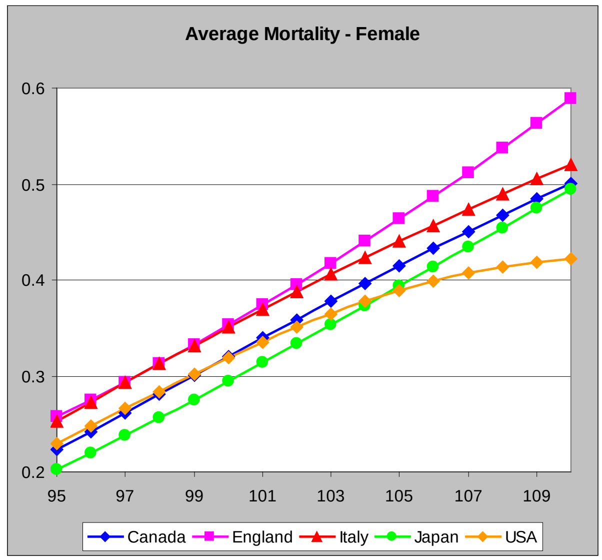
The ability to discern improvement rates is not nearly as good, but I think it is good enough to use with some caution.

5 Other Countries

5.1 Concerns about the Validity of the Method

Because there is comparable data available for England, Italy, Japan and the United States, I applied my method to the data for those countries as well. (There is comparable data for Sweden and Switzerland as well, but I judged the populations too small.) The first year of data is 1970 in all cases. The ending year is the one most recently available and varies from 2005 to 2008. Because all of these countries have more data than Canada, I was concerned the graduation may be smoothing a bit too much, but my testing suggests the parameters for Canada are close enough for the other countries as well. The graphs of the average mortality rates are shown below.

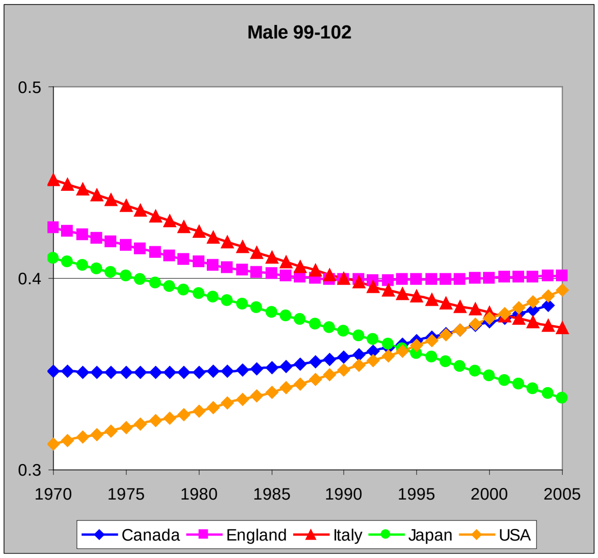


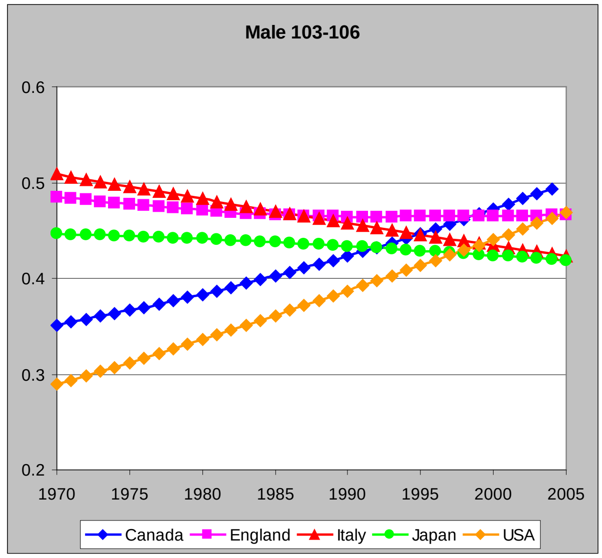


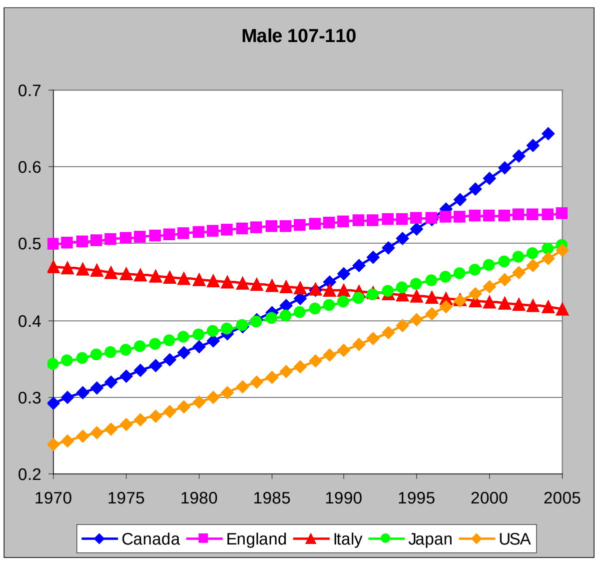
What is happening with the U.S. male? One would think the country with the most data would very likely have the most credible result, but the U.S. curve is such an outlier it really cannot be believed.

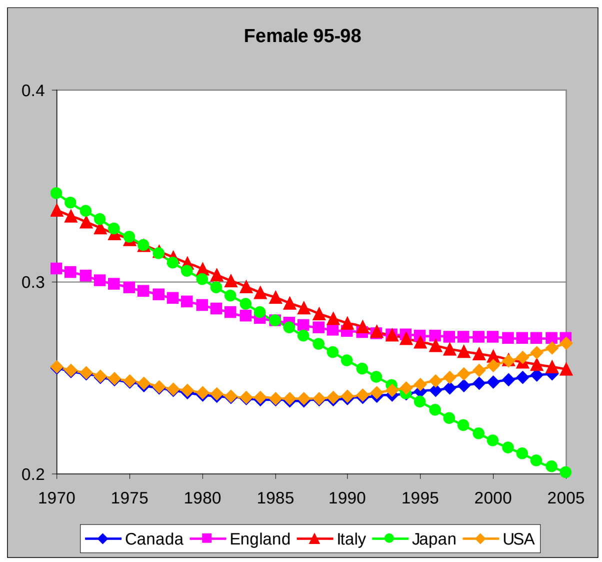
I have heard it theorized that mortality rates level off at high ages rather than increasing steadily toward 1.0. The U.S. data, for both males and females, is consistent with a leveling off, although males seem to go too far. Canadian and Italian males also show some leveling off. Other curves appear to continue, not far off linearly, over the 16-year age range. The differences are so large it seems doubtful they can be explained by national differences.

The following charts show the mortality rates by age group for males. Note how different the curves are by country. What had looked like a nice neat method, when applied to Canadian data, now begins to look out of control. By the way, recall that the smoothing factor is very heavy for these charts.









The variation by country is startling. For Canada and the United States, we see considerable periods of mortality deterioration. For Italy and Japan, there is some improvement. England is close to neutral.

The following table shows what I think to be the most striking feature of the U.S. data.

Male Deaths 1970-2004					
	Canada	England	Italy	Japan	USA
95-99	45240	86132	96795	174989	471319
100-104	6509	8561	9928	19867	69326
105-109	469	375	429	981	6078
110-114	20	2	10	22	613
115+	3	0	0	1	191

The deaths in 95 to 99 are not far off what one would expect given the relative sizes of the populations. The United States shows a far slower decline in deaths at the higher ages than any other country. Indeed, the rate of decline does not look credible.

I recall a comment by Bruce Schobel at the last Living to 100 symposium that there was a lot of exaggeration of age in the Social Security data. If so, that could carry over into the reporting of age at death.

I hypothesized that the reported ages at death is exaggerated in some cases, more so for the older birth cohorts that may have less reliable birth records. I tested the hypothesis by simulating deaths, as described in the validation section, but beginning 10 ages and 10 years earlier. The underlying mortality table is the one I observed for Japanese males, but I used the starting population of U.S. males. I created a second set of deaths with about 10 percent of deaths being exaggerated by one to 12 years. The results were somewhat similar to those shown above for the United States. I did not get the mortality curve to bend down quite as far, but the confirmation was strong enough to lead me to suspect some of the deaths in earlier years.

5.2 Revised Method

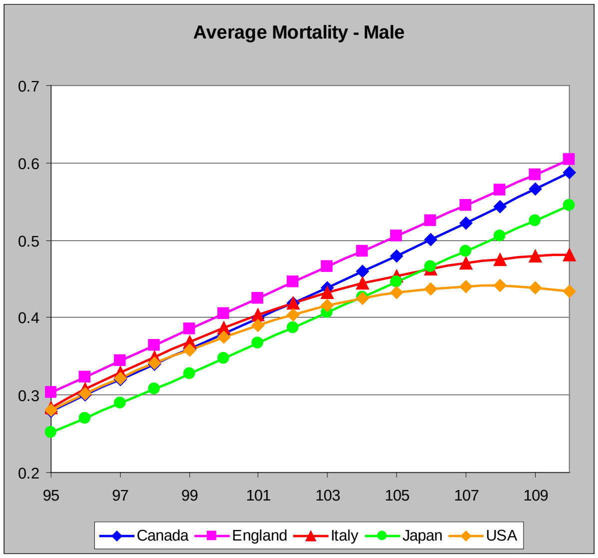
The method I developed still seems reasonable, but it is predicated on accurate data. If the data lacks accuracy, then although my method may be able to infer exposure consistent with the deaths, and develop appropriate mortality tables and improvement scales, the results will be precise but wrong. The challenge is to make the best use of the data available.

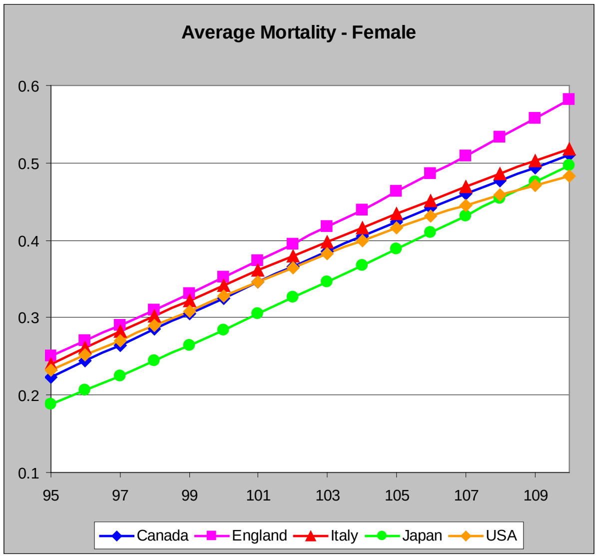
The method I have come to is a pragmatic solution. It lacks precise mathematical justification and formulation. Nonetheless, I believe it is an appropriate response to the data.

1. Infer exposure from death records as described above.
2. Calculate the average mortality rates using data of the last 15 years.
3. Extrapolate the calculated mortality rates after a high age of 100 to 105, depending on the data, at a rate of 0.02 per year of age.
4. Close the table with a mortality rate of 1.0 at age 115.
5. Estimate the improvement rate at age 95 for the last 15 years using my method as described above. Grade that rate down to 0 at age 105 to 110 depending on the data. Allow no deterioration in mortality rates.
6. Set improvement rates for extending after the study period at 0.004 for age 95, 0.002 at age 100, and 0 for ages 110 and higher, with quadratic interpolation.
7. Determine mortality rates for each country for 2011.

5.3 Average Mortality Rates

The concern about possible exaggeration of age at death for older cohorts led me to try limiting my attention to more recent data. I found using the most recent 15 years to be a reasonable compromise. I also looked at using only 10 years, but there seemed not to be enough data to establish a clear pattern. The following charts are based on the most recent 15 years of data.

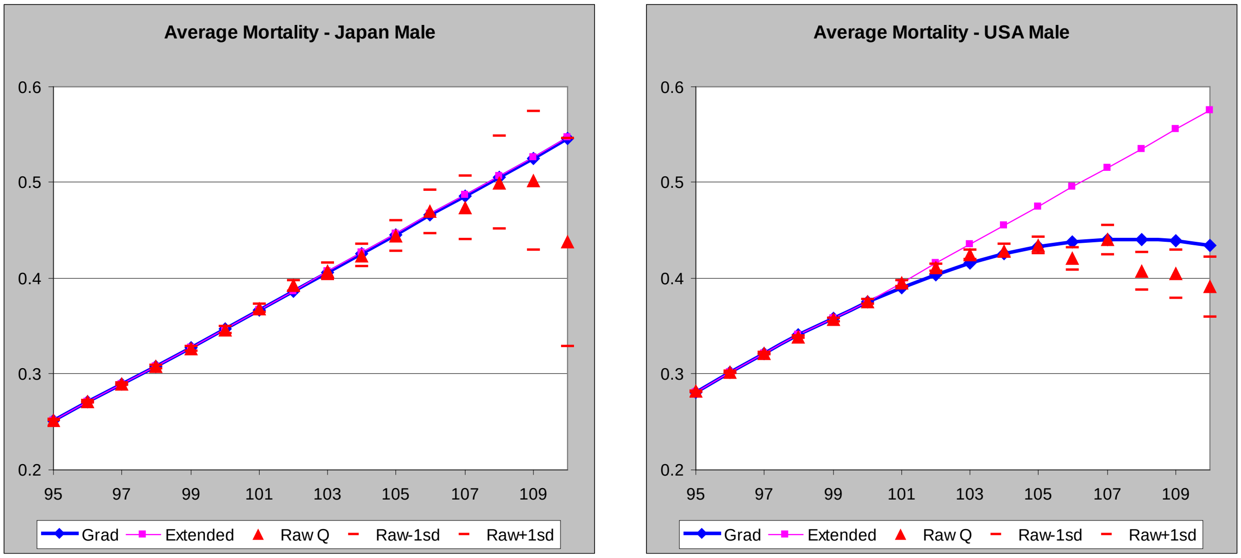




These charts show a more reasonable pattern. The U.S. males are still outliers, now chased by the Italians rather than the Canadians. If my hypothesis is correct, there is still some exaggeration of age at death but less than in earlier years. The female curves are remarkably consistent. Exaggeration of age seems less a problem for females.

The above charts suggest the "normal" curve is very close to linear with an increase in mortality of about 0.02 per year of age. Canada, England and Japan are consistent with that, and so are Italian females. Italian males and the United States are on the same track for at least five years, but then the slope decreases considerably. It is tempting to declare that the lower slope is not real but an artifact of exaggeration of age at death.

These observations lead to a trial set of mortality rates, even with suspect data, of the graduated raw rates for ages 95 to 100, extended at 0.02 per year thereafter. Let's consider, for examples, Japanese and U.S. males.



For Japanese males, the mortality rates extended from age 100 at 0.02 per year fall almost exactly on the graduated mortality rates and clearly well within one standard deviation of the raw mortality rates. However, the extended rates for U.S. males lie farther than one standard deviation above the raw rates also as soon as the extrapolation starts.

This brings us to a serious professional question for actuaries and demographers. Is it appropriate to estimate mortality rates at a level clearly not supported by the observed data? I believe it is in a case such as this. The observed data appears very suspicious, and there is a plausible explanation for the difference between the extrapolated mortality rates and the observed ones. Of course, departing from the data requires prominent disclosure, and I am doing that here.

Fortunately, the financial significance of these ages is low. As a measure of significance, I compared annuities at 5 percent calculated on two bases. The first is the calculated mortality rates by my initial method, extended to 0.6 at age 114 and 1.0 at age 115. The second is mortality rates extended from age 100 at 0.02 per year. The table below compares the present value of annuities for the U.S. male, the dataset with the largest difference between the two methods.

Clearly the difference is unimportant at age 95. Age 99 is the first one with a difference of more than 1 percent; at that age, there are only two unadjusted mortality rates.

Annuity at 5% - USA Male			
	Prior	Revised	Ratio
95	1.886	1.881	99.8%
96	1.754	1.748	99.6%
97	1.639	1.629	99.4%
98	1.538	1.522	99.0%
99	1.449	1.425	98.3%
100	1.373	1.333	97.1%
101	1.309	1.241	94.9%
102	1.255	1.156	92.1%
103	1.212	1.077	88.8%
104	1.179	1.003	85.0%
105	1.155	0.934	80.8%

From a visual inspection of the data, I decided to extend the table at 0.02 per year after age 100 for Italy and the United States, age 103 for Canada, and age 105 for England and Japan.

5.4 Omega

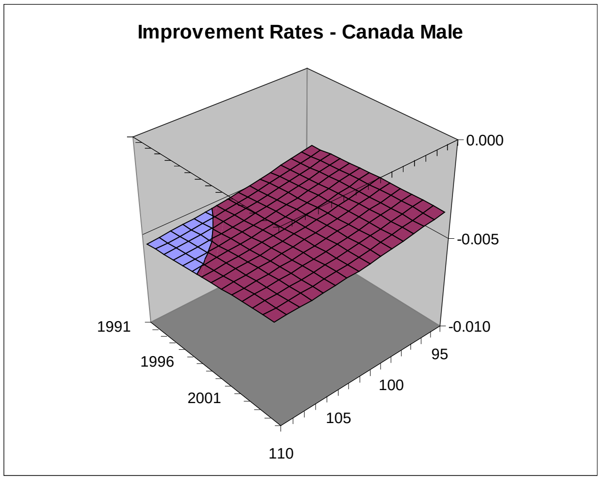
How to end a mortality table is essentially an arbitrary choice. By definition, we cannot calculate mortality rates from experience after there are no lives left. Commonly, tables have used 116 as omega; that is, the mortality rate at age 115 is 1.0. Some more recent tables have chosen 121 as omega.

I compare these two choices by looking at the present value of an immediate annuity of 1 per annum at 5 percent. The table I used was Japanese males extrapolated at 0.02 per year until age 115 or age 120 set to 1.0. I found the ratio between the two annuities to differ from 1.0 by only 1 basis point at age 105, and 0.6 percent at age 110. Because virtually no insurance or annuities are sold over age 100, I conclude mortality rates above 115 are of no financial significance.

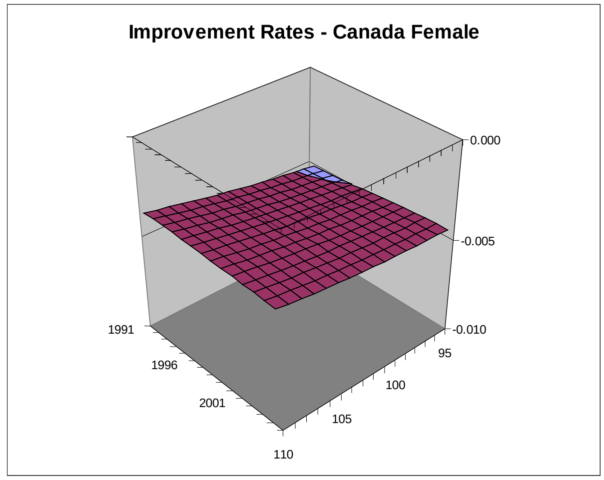
Accordingly, I will use 116 as omega for my tables.

5.5 Improvement Rates

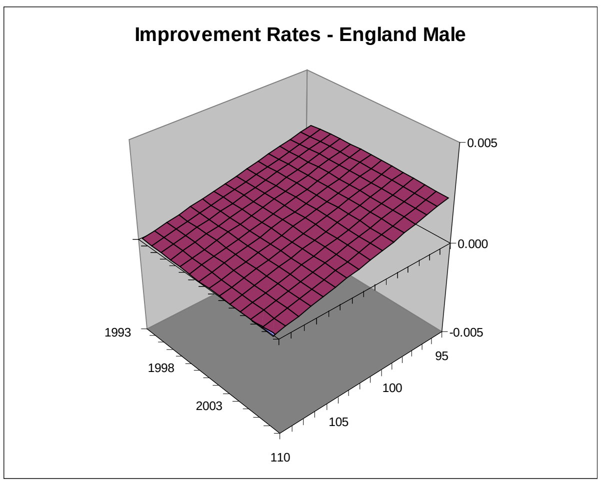
My initial method calculated improvement rates for each age and year. The results are shown in the following surface charts. Because of the heavy smoothing applied, most graphs appear planar.



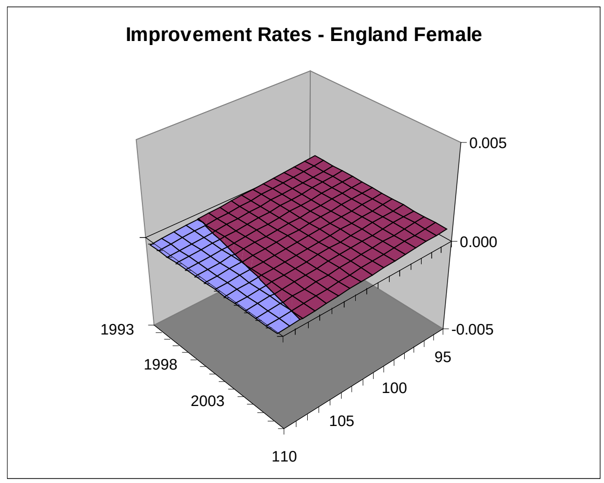
Improvement Rates - Canada Male



Improvement Rates - Canada Female

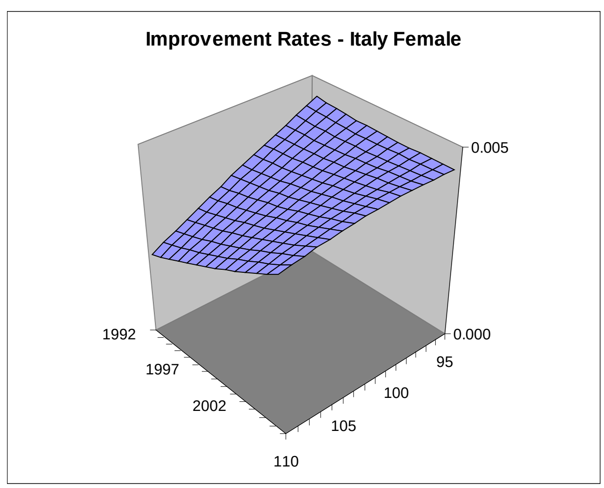


Improvement Rates - England Male

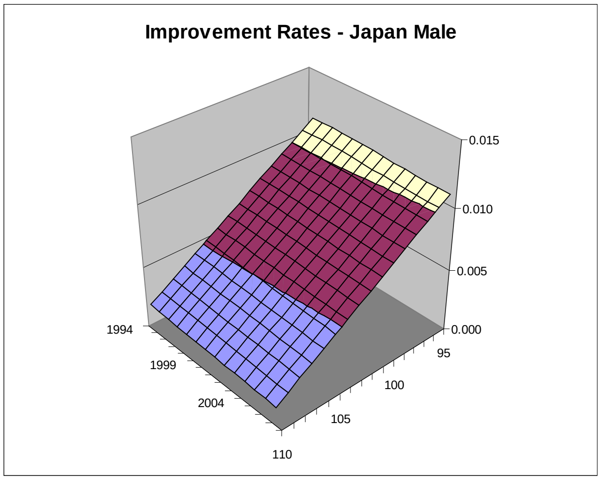


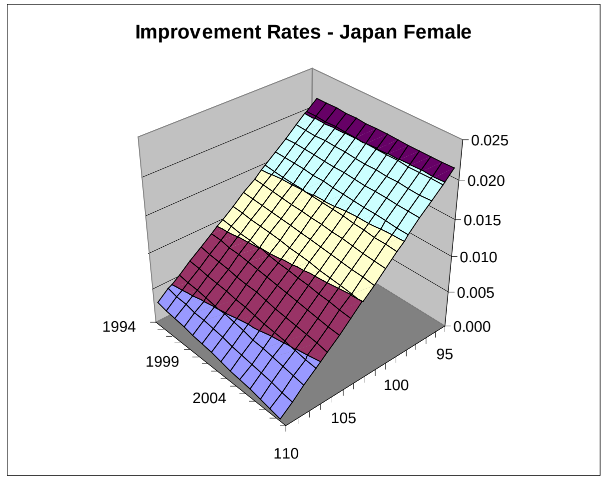
Improvement Rates - England Female

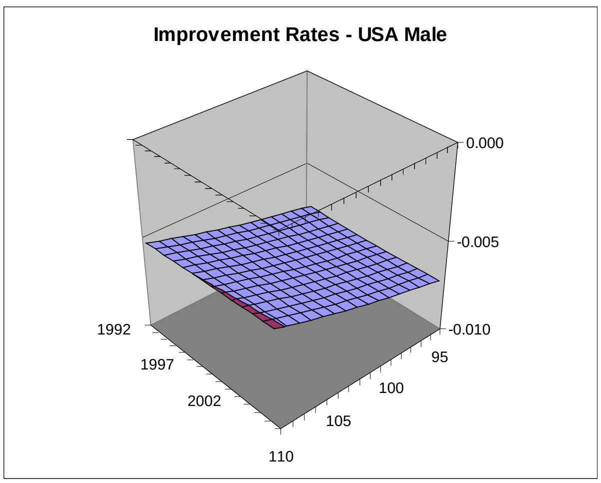
Improvement Rates - Italy Male

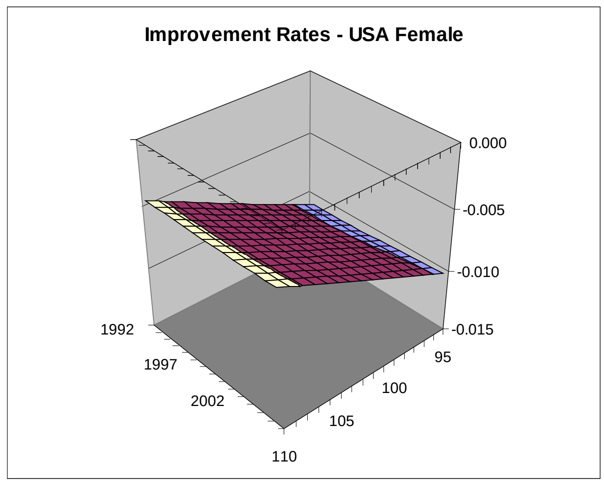


Improvement Rates - Italy Female









Based on my observations, I chose the following as improvement rates during the study period. These rates are used to move the average mortality rates from the middle of the study period to its end.

Canada Male: No improvement.

Canada Female: No improvement.

England Male: 0.0022 at age 95 grading linearly to 0 at age 110.

England Female: 0.0005 at age 95 grading linearly to 0 at age 105.

Italy Male: No improvement.

Italy Female: 0.004 at age 95 grading linearly to 0 at age 110.

Japan Male: 0.011 at age 95 grading linearly to 0 at age 110.

Japan Female: 0.021 at age 95 grading linearly to 0 at age 110.

U.S. Male: No improvement.

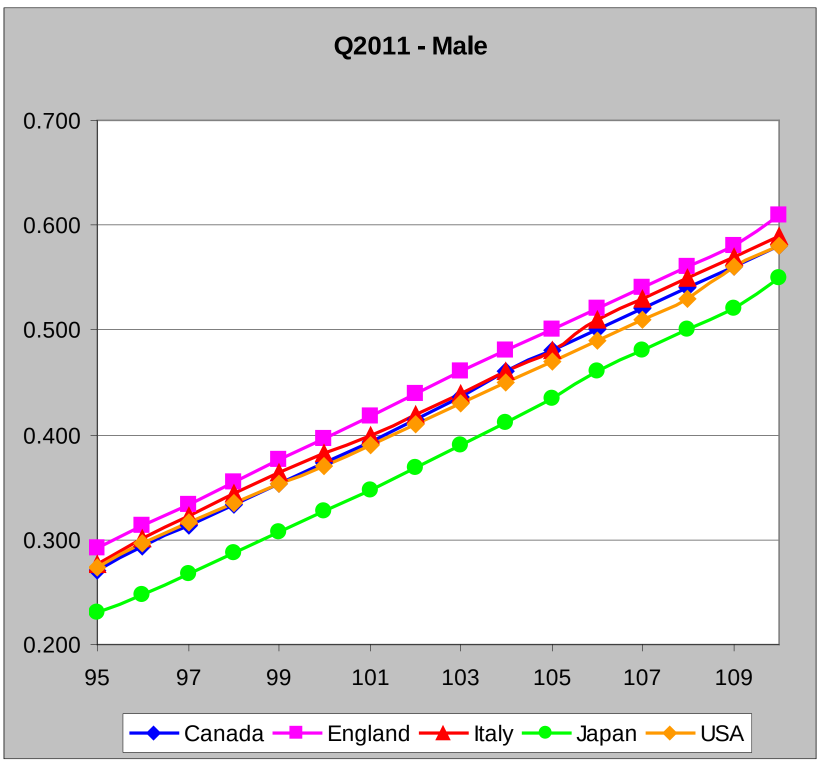
U.S. Female: No improvement.

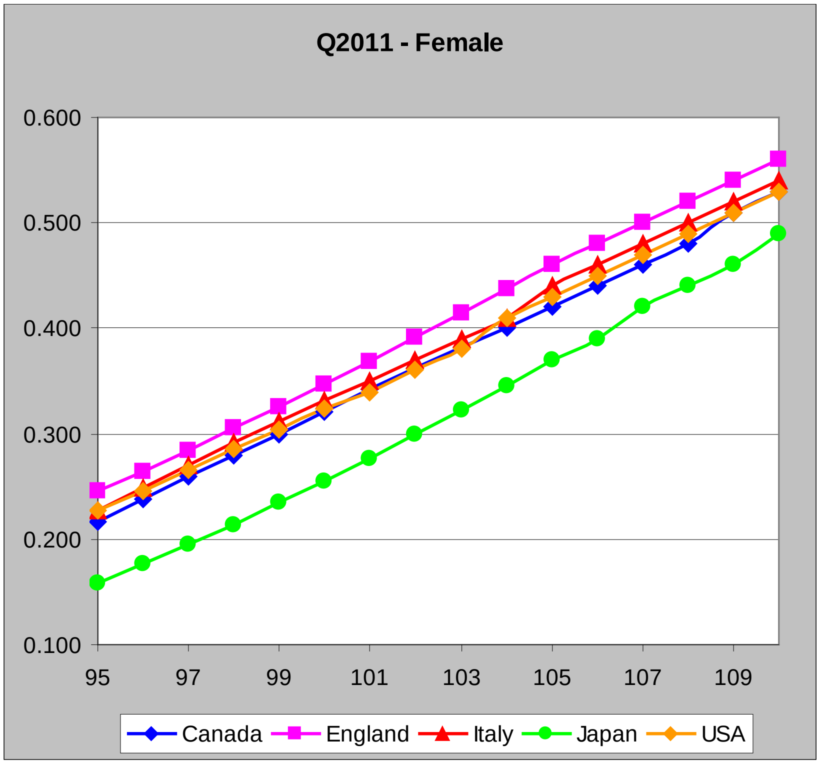
I decided to use the same improvement rates for all countries and both sexes after the study period. Clearly, these rates cannot be founded directly on data because they are about the future. I am inclined to use less than we have recently seen in most countries that exhibit improvement at these ages. I believe there is justification for pessimism about improvement at these high ages for several reasons. Although we have seen many medical advances in recent years, the increase in cost is slowing them down, and they tend to be more applicable to younger ages. The combined effect of pollution, global warming and population growth will tend to have most impact on the frail. The decline in smoking prevalence, which has produced much improvement in mortality, cannot be repeated. The emphasis of geriatrics is much more on the quality than the quantity of life. Society increasingly seems to take a utilitarian view of health care that is less likely to justify significant expenditure of the very elderly.

My improvement scale for after the study period is 0.004 at age 95, 0.002 at age 100, and 0 at ages 110 and up, with quadratic interpolation in between.

5.6 'Oldest 2011' Tables

The following charts display the final tables as of 2011. I have named these tables "Oldest 2011 Canada Male," etc. A workbook with the mortality rates and improvement rates is available at http://www.howardfamily.ca/~bob/oldest.




Q2011 - Female
0.600
0.500
0.400
0.300
0.200
0.100
95
97
99
101
103
105
107
109
Canada
England
Italy
Japan
USA

6 Applications

I commend my tables to anyone constructing a table on lives of the countries covered. It is rarely possible to construct a table reaching past age 90. I think it would often be reasonable to use my table to extend the table being constructed.

For example, consider an ultimate table that has good support to age 90. I would extend that table by using my table for ages 100 and older and linking the two segments by Lagrange interpolation using the rates for ages 89, 90, 100 and 101. The interpolated segment is then a cubic with the same slope at either end as the other segments.

In some applications, one might consider using a multiple of my table, but I am not convinced doing so is warranted. I know of no study that supports differentials in mortality by socio-economic class or underwriting classification at these very high ages. I conjecture mortality at these ages is more a matter of genetics and attitude than of past advantages in lifestyle.

7 References

Howard, Robert C.W. Graduation software and a description of Whittaker-Henderson graduation and variants available at: http://www.howardfamily.ca/~bob/graduation.

Human Mortality Database available at http://www.mortality.org.

London, Dick. 1983. *Graduation: The Revision of Estimates.* Abington, CT: ACTEX Publications.