MORTALITY RATE AS A FUNCTION OF AGE

Donald B. Gennery

January 19, 2010

This document presents some information on mortality rate as a function of age, derived from several sources. A combined estimate of an underlying probabilistic mortality-rate function and the resulting survival-probability function also are presented.

The first three figures show estimates of the instantaneous mortality rate, which is the expected rate at which people die per unit of time (here the year) at each precise age, relative to the surviving population at that age. If this is called M, and if it were constant, then the probability that a particular individual will die within one time unit (year) would be $1 - \exp(-M)$, the remaining life expectancy (mean life) would be $1/M$, and the half-life would be $\ln(2)/M$. Since M actually varies with age, only the first of these relationships is reasonably accurate in general, and only if M is taken to be the value in the middle of the one-year (or other) span.

Figure 1 shows estimates derived from several different sources, and it covers ages 85-125. Figure 2 shows only ages above 110 in greater detail, and hopefully it is helpful because of the large amount of information crammed into that region.

Each source provided data that gives deaths or population at each age. (Table 1 shows the data used from all of these except Social Security.) From these the estimated fraction F of the surviving population at age A that dies within one year was computed and was converted to mortality rate by using the formula $M = -\ln(1 - F)$. Here M is the mortality rate at age $A + 0.5$, if the data is derived from the number of deaths at each yearly age, as stated above. However, in cases where the data is the number of survivors at each yearly age, there is an additional shift of nearly a half year because the precise ages are spread throughout the year. Therefore, in these cases M is considered to be the value at age $A + 1$ as an approximation, or $A + 2$ in one case where two-year intervals are used for the data. Also, except for Greenwood and Irwin, the plotted points should be considered only approximations anyway, because the data was derived from people born in different years, instead of a cohort that was followed as its members die, and the sets of data are incomplete.

For each set of data plotted as individual points, confidence limits are shown by using a tee-shaped mark for the upper limit and an inverted tee for the lower limit (of the same color as the main plotted symbol). These limits take into account only statistical fluctuations due to the sample sizes of the reported numbers; they neglect any errors due to nonuniformity of the population and incomplete data. For cases in which the raw data is deaths at each age (in one-year buckets), these limits are the 68% confidence limits rigorously derived from the binomial distribution. (68% was chosen because it corresponds closely to the one-standard-deviation limits for the normal distribution.) Thus the probability is 16% that an expected mortality rate equal to the upper limit could have produced the observed number of deaths or fewer at this age, and the probability is 16% that an expected mortality rate equal to the lower limit could have produced the observed number of deaths or more. For cases in which the raw data is the number of those alive at each age, a rigorous approach would involve the ratio of two Poisson distributions. Therefore, for these cases a simpler approach was used, in which standard deviations of the Poisson distributions were propagated into one-standard-deviation limits by using a linear approximation.

For some of the plotted points, the upper or lower limit, and even the nominal value itself, can be off scale (in some cases at infinity or zero) and thus is not shown.

The thick yellow-brown curve is derived from the actuarial life table from the Social Security Administration [1]. The separate data for males and females have been combined according to the number of each surviving at each age. Individual points and confidence limits are not shown, because this is not raw data. Apparently it has been heavily processed and is mostly artificial at the high ages. (It indicates that the probability of dying within a year increases at an approximately constant exponential rate from age 100 until it reaches 1 at age 120, so that the mortality rate goes to infinity there. This is not a realistic situation.) According to Robert Young, the Social Security data probably is unreliable above age 95 or so.

The green open upright squares are derived from Greenwood and Irwin [2]. They used data that followed 290 people who had attained the age of 90 in 1920-1922 and recorded the age at which they died. It can be seen that the departure from smoothness is comparable to the given confidence limits. An interesting thing about their paper is that, even though the last subject died at age 102, they used a mathematical extrapolation technique to show that the data is consistent with mortality rate leveling off to a constant value such that the deaths each year would be 43.9% for women and 54.4% for men. These values correspond to instantaneous mortality rates of 0.578 for women and 0.785 for men. Since there are about 6 times as many women as men at very high ages, the weighted average of the final mortality rate would be 0.61. This is not greatly different from the estimates of 0.79 at age 115 and 0.67 at age 120 produced below, based on more recent data that extends to age 122.

The orange open tilted squares are derived from the ages of centenarians living in Europe as of January 1, 2008, provided by N. J. Ruisdael [3]. Apparently Ruisdael estimated total values for Europe from available incomplete data. Therefore, the actual statistical fluctuations are larger than indicated by the confidence limits on the plot, since the latter are computed from the extrapolated total numbers instead of the actual smaller sample sizes.

The magenta diagonal crosses are derived from the ages of people living in England and Wales in 2005 [4].

The red open circles are derived from the number of validated living supercentenarians at each age as of January 7, 2010, as reported by the Gerontology Research Group (GRG) [5]. Because of the small sample size, this data was pooled into two-year intervals in order to make it less noisy and thus to reduce the wide range of the confidence limits. Also, the data is incomplete, especially at the lower ages, and this fact introduces extra uncertainty and tends to bias the results towards lower mortality rates.

The blue circular dots are derived from the number of supercentenarians known to have died at various ages as of January 3, 2010, as reported by Louis Epstein [6], with dubious cases as indicated by Epstein deleted. Also, all individuals born after May 9, 1895, are ignored in order to avoid bias caused by the fact that some of those in that group are still alive, as suggested by Robert Young [7]. (The actual numbers used are shown in Table 1.) The indicated confidence limits are derived from the binomial distribution for the stated numbers. However, it is likely that there is a relative lack of completeness at the lower ages (below about 113), which causes the computed mortality rates to be too low at these ages.

Fits were done to two combinations of Epstein's data. Each function fitted represents a mortality rate that is a straight line on a logarithmic scale such as Figure 1. It represents the fit of one-year differences of a Gompertz function to Epstein's data of deaths at each age (in one-year buckets). Accurate minimum-variance adjustments were done (in the logarithmic space of the figures), assuming that the number of deaths at each age has the Poisson distribution. The results of the fits are plotted in Figures 1 and 2. In the figures, the nominal fit is represented by a dashed line, and the one-standard-deviation error limits are represented by dotted lines of the same color. (The dotted lines are curved because of the correlation between the parameters, but the dashed lines are straight.)

One of the two fits used all of Epstein's data, and it is represented by the relatively short blue dashes. This fit has a fairly small formal uncertainty (represented by the closely spaced blue dots), but it is biased towards an increasing mortality rate with age because of the bias in the data towards lower mortality rate at the lower ages. (It also is affected by the probable curvature in the actual mortality rate function over this span of ages, which the Gompertz function doesn't model.) The other fit used only the data for ages 113 and higher, and it is represented by the long purple dashes. This fit has a larger formal uncertainty (represented by the widely spaced purple dots), but it has less bias, because of the greater reliability of the data at these higher ages. (The actual mortality rate also can be more accurately approximated by a Gompertz function over this narrow range of ages.) Therefore, the latter fit probably is a more statistically reasonable description of what is happening at the highest ages.

In order to obtain a reasonable guess at what the mortality rate curve (without the statistical fluctuations) actually looks like, an approximate manual fit was done (represented by a solid black line in Figures 1 and 2). It coincides at age 90 with the Social Security data and at age 120 (and higher) with the fit to Epstein's data for ages 113 and higher, and it proceeds smoothly between these points by trying to achieve reasonable agreement with the other data.

Figure 3 then shows the resulting curve for all ages. The dotted lines above age 115 represent reasonable error limits for the curve. These are the one-standard-deviation limits from the fit to Esptein's data for ages 113 and higher, so they can be taken to be approximately 68% confidence limits. The combined fitted curve shown in Figure 3 has the value 0.669 at age 120, which if constant would correspond to 48.8% dying in a year. Over the entire range of ages from 105 to 121, the fitted value is between 0.64 and 0.81, which correspond to 47% and 56% dying in a year, respectively, although the confidence limits extend outside of this range.

The sets of data in Figure 1 were obtained from different locations and years. Therefore, perhaps it shouldn't be too surprising that the fitted function disagrees with the Greenwood and Irwin data by about 30% and with the the Ruisdael and UK data by about 10%. Similarly, the exact shape of the curve giving probability of dying per unit time as a function of age may vary from Figure 3. Thus the long fairly straight line extending from age 35 to age 100 representing a relative rate of change of mortality rate of about 0.09 per year may apply only to the United States and a few other countries at the present time. Nevertheless, the leveling off of the mortality rate around age 110 may well be a universal phenomenon for human beings. However, whether the mortality rate keeps on increasing (at a lower rate), becomes constant, or starts to decrease cannot be determined from the available data.

Figure 4 shows the probability at birth of survival as a function of age, based upon the estimates of current mortality rate shown in Figure 3. (The function in Figure 4 is the exponential function of the negative of the integral of the function in Figure 3.) Because of the uncertainty of fit that lead to Figure 3, the probabilities in Figure 4 are uncertain at the higher ages, especially past age 115. (At age 120, Figure 4 could be off by a factor of 10. Also, remember that the values are derived primarily from current mortality rates in developed countries.)

A rough check on Figure 4 can be made by comparing it to the actual number of people known to have survived to specified ages. For example, for age 115 the number is 22. All of these cases (and 70% of all verified supercentenarians) are people who were born during a

20-year period from 1875 through 1894 and are almost entirely from the United Sates, Canada, Japan, and Europe excluding Russia. (Presumably before then record keeping was too poor to produce many verifiable cases, and anyone born later hasn't had time yet to reach age 115. These places had good record keeping 120 years ago and good enough nutrition and health care to produce an appreciable number of supercentenarians). By combining information from a variety of sources (e.g. [8] and [9]), a rough estimate of the number of people born in these locations during the specified period is 300 million. If Figures 3 and 4 are correct, at birth a person has the probability $3.98\times10^{-7}$of living to age 115. Multiplying these two numbers produces 119 as the number of 115-year olds that would be expected, whereas the observed number is 22. However, the former number is based on current mortality rates, whereas the latter number is the result of mortality rates over the last 130 years or so, which were higher, so the latter number would be expected to be lower. The ratio of the two numbers is 5.41, which perhaps is reasonable. (Remember that there is considerable uncertainty in both the numbers 300 million and 3.98×10^{-7} that led to the value 119 and hence to 5.41.)

The greatest known age to have been achieved by anyone is 122.45 years. According to Figure 4, the probability of achieving this age is 2.11×10^{-9}. By using the above estimate of 300 million births, the expected number is 0.633, and applying the above factor of 5.41 produces 0.117 as the approximate expected number considering the higher mortality rates in the past. With the Poisson distribution and 0.117 for the expected number of events, the probability of having 1 or more events is 0.110, which is a reasonably large probability. Therefore, the agreement is satisfactory.

References

[1] Social Security Administration, 2005 Period Life Table, http://www.ssa.gov/OACT/STATS/table4c6.html.

[2] M. Greenwood and J. O. Irwin, "The Biostatistics of Senility," *Human Biology* Vol. 11, No. 1 (Feb. 1939), available at http://longevity-science.org/Greenwood-Human-Biology-1939.pdf.

[3] N. J. Ruisdael, e-mail forwarded to GRG members by Steve Coles, Feb. 25, 2008.

[4] UK National Statistics, Estimates of the Very Elderly, available at http://www.statistics.gov.uk/statbase/Product.asp?vlnk=15003.

[5] Gerontology Research Group, Validated Living Supercentenarians, http://www.grg.org/Adams/E.HTM.

[6] L. Epstein, The Oldest Human Beings, http://recordholders.org/en/list/oldest.html.

[7] R. Young, e-mail to GRG members, Nov. 23, 2009.

[8] C. Haub, "How Many People Have Ever Lived on Earth?" *Population Today*, November-December 2002, available at http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx.

[9] Wikipedia, World Population, http://en.wikipedia.org/wiki/World_population.

Table 1
Data Used to Compute Mortality Rates in Point-by-Point Plots

Cohort followed in 1920s and 1930s, Greenwood and Irwin [2]

Age	Number alive at this birthday	Deaths before next birthday
90	290	57
91	233	57
92	176	36
93	140	39
94	101	32
95	69	23
96	46	15
97	31	13
98	18	10
99	8	3
100	5	4
101	1	0
102	1	1
103	0	0

Estimated living in Europe, Jan. 1, 2008, N. J. Ruisdael [3]

Age	Number living	Decrease to next age
100	28679	12467
101	16212	6657
102	9555	4158
103	5397	2413
104	2984	1541
105	1443	787
106	656	295
107	361	184
108	177	111
109	66	?

Living in England and Wales, mid 2005, UK National Statistics [4]

Age	Number living	Decrease to next age
90	88600	14660
91	73940	15130
92	58810	13610
93	45200	11300
94	33900	8680
95	25220	7040
96	18180	5570
97	12610	4210
98	8400	2980
99	5420	2010
100	3410	1320
101	2090	870
102	1220	540
103	680	300
104	380	?

Validated living supercentenarians, Jan. 7, 2010, Gerontology Research Group [5]

Age	Number living	Decrease to next age (2-year increment)
110 & 111	51	30
112 & 113	21	17
114 & 115	4	4
116	0	0

Deaths of validated supercentenarians born before May 10, 1895, as of Jan 3, 2010, Louis Epstein [6]

Age	Number at this age or higher	Number at this age, by years
110	918	427
111	491	225
112	266	129
113	137	74
114	63	41
115	22	15
116	7	3
117	4	2
118	2	0
119	2	1
120	1	0
121	1	0
122	1	1
123	0	0

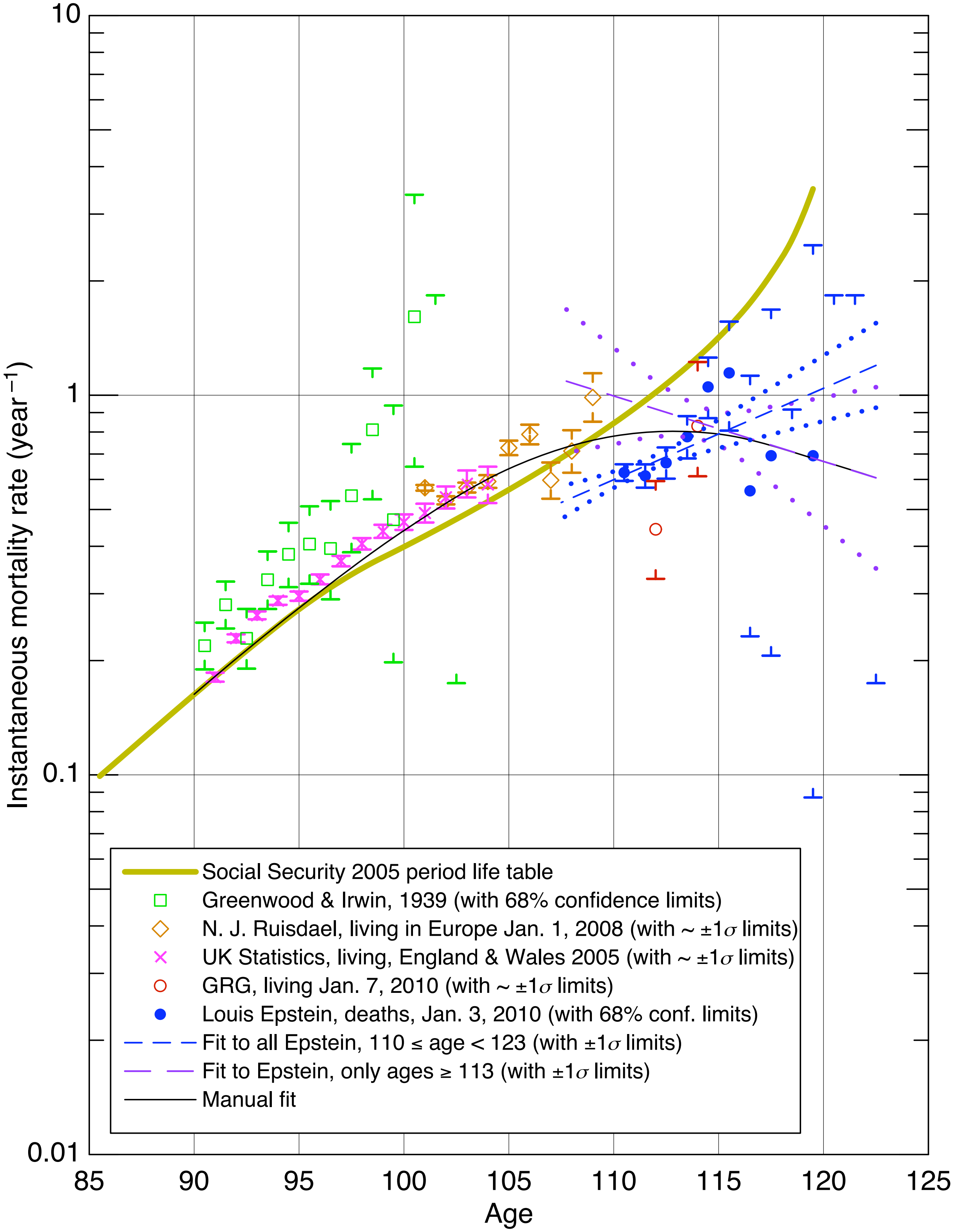

Figure 1. Estimates of mortality rate as a function of age, derived from various data.

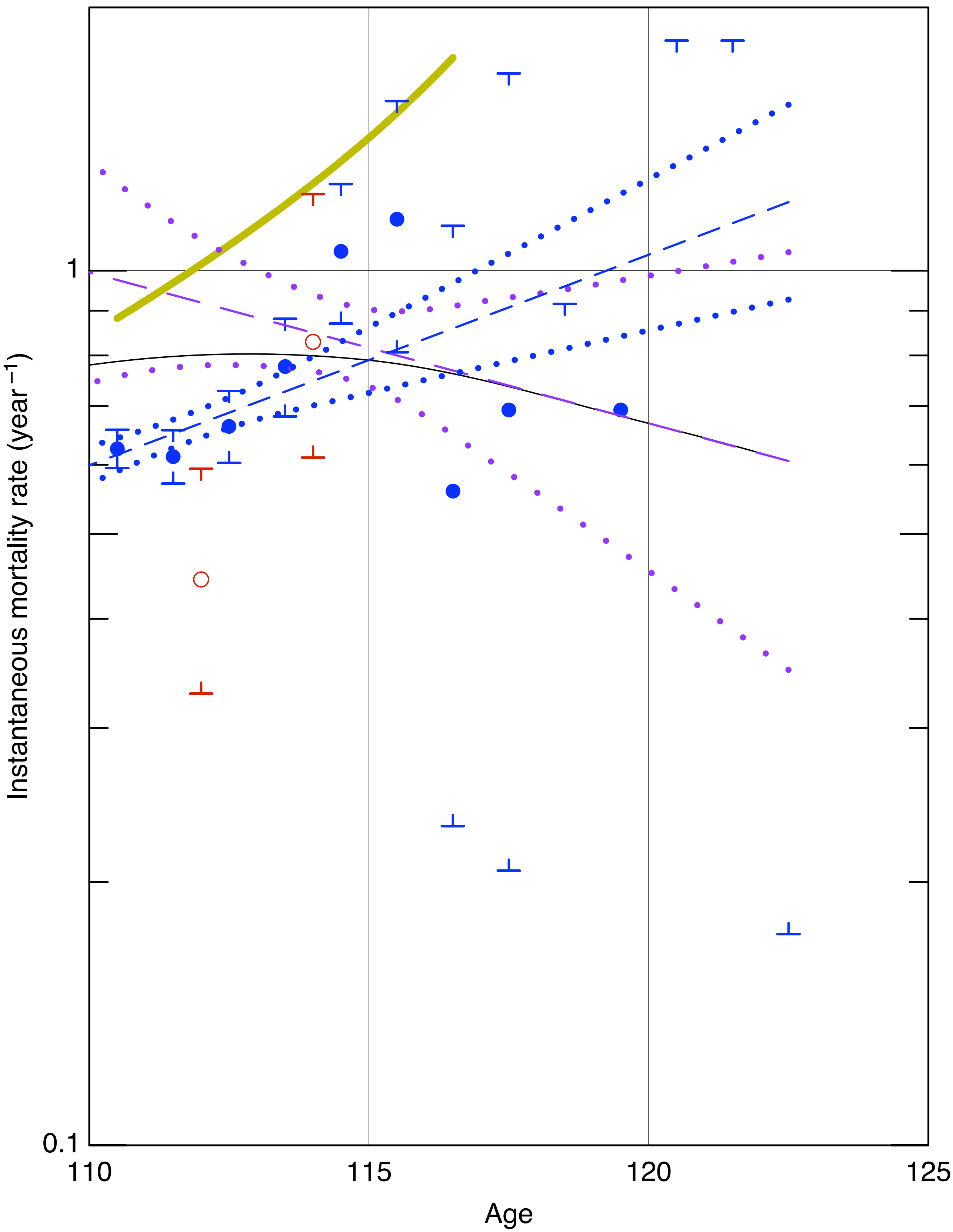

Figure 2. Enlarged portion of Figure 1.

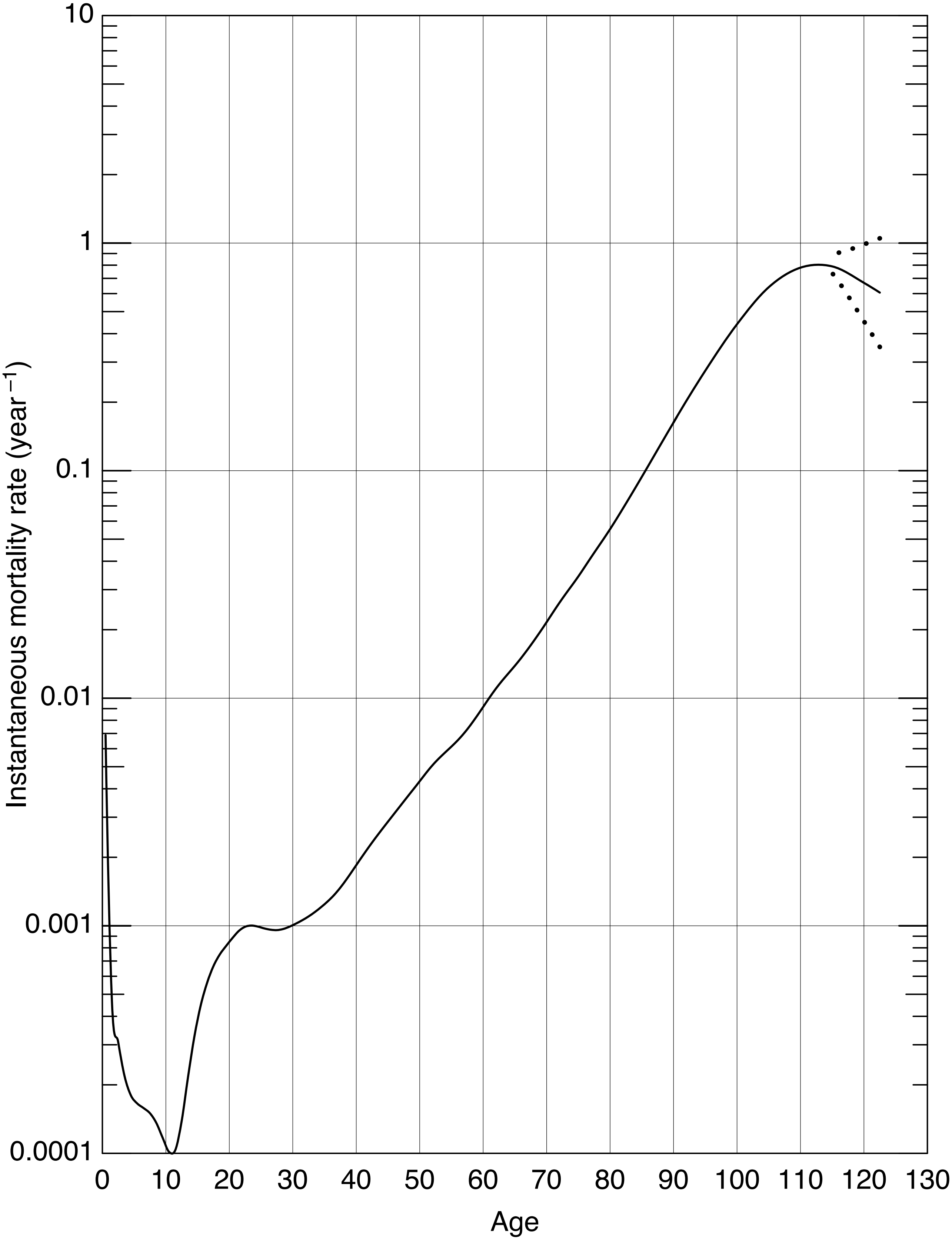

Figure 3. Results from Social Security and manual fit spliced together at age 90 (with approximate $\pm 1\sigma$ limits).

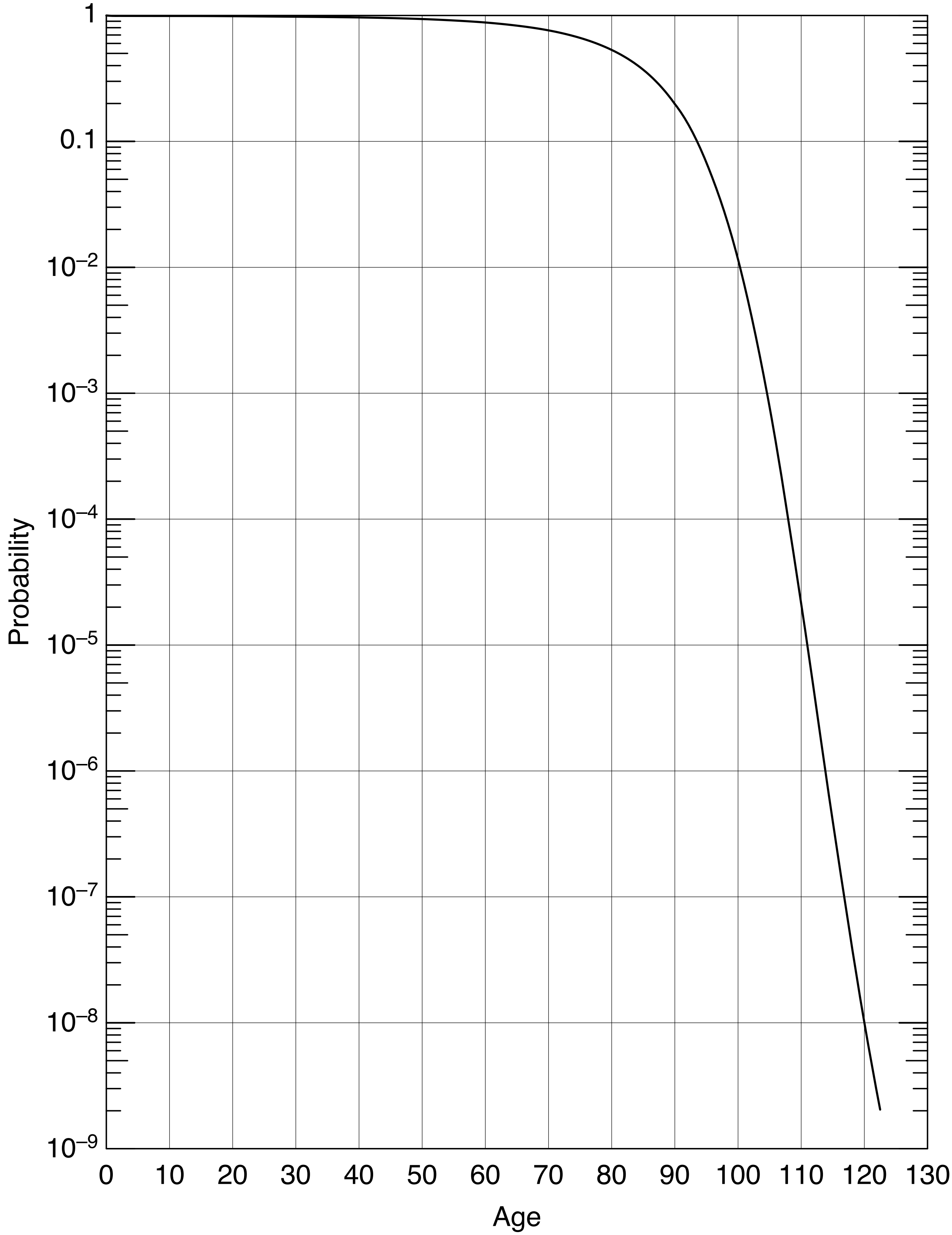

Figure 4. Probability of surviving to each age, derived from the curve in Figure 3.