TRANSACTIONS OF SOCIETY OF ACTUARIES
1986 VOL. 38

CROSSOVER IN MORTALITY RATES BY SEX

ESTHER PORTNOY

ABSTRACT

Mortality patterns among the aged have attracted considerable actuarial attention in recent years, with good cause. Accurate projections are essential in planning for the needs of the increasing numbers of aged persons in our population. Moreover, a careful study of mortality patterns will aid in the effort to understand the aging process.

A second matter of great interest is the difference between mortality rates among males and females. It is well established that female rates are significantly lower than male rates during youth and middle age and that the difference gradually lessens in advanced age. Some recent studies have indicated a "crossover," with female rates exceeding male rates beyond some age around 100. This is qualitatively different from a convergence of rates and requires careful attention.

The purpose of this paper is to develop some statistical tests for the significance of observed crossovers. Two tests based on nonparametric comparisons of mortality experience and one method involving graduation of the observed rates are applied to data presented by Bayo and Faber [1]. For the first data set, charter beneficiaries of Social Security, the tests indicate that the observed crossover is not statistically significant. The second data set, Medicare recipients, appears to offer significant evidence of a crossover at about age 102.5.

1. A MODIFIED WILCONXON TEST

The Wilcoxon test is a standard test for equality of distribution that can be modified slightly to obtain a useful tool in the analysis of the crossover question. It is appropriate for relatively small data sets, like the "charter beneficiaries" group in Bayo and Faber, where not too many individuals are recorded with the same age at death.

The basic one-sided Wilcoxon test (described, for example, in Elandt-Johnson and Johnson, [5], 231–40) begins with random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ from two unknown distributions. The question is whether the distributions are consistently ordered, $F_X(x) \leq F_Y(x)$ for all x. The Wilcoxon statistic W is the sum of ranks of the X_i within the set $\{X_i, Y_j\}$, where rank 1 is assigned to the smallest, rank 2 is assigned to the next smallest,

and so on until rank $n+m$ is assigned to the largest. The null hypothesis $F_X(x) \leq F_Y(x)$ is rejected if W is significantly smaller than would be expected if the two distributions were equal.

The charter beneficiaries data consist of sex, year and month of birth, and year and month of death for each individual. Let X_i, Y_j be the "calendar age" at death (measured in years and months) of females and males, respectively. There are a number of "ties," but not enough to invalidate the procedure. In fact, errors of ranking are likely. For example, a person born in late January 1873 who died in early January 1973 would be recorded as a death at age 100, but actually would have been younger at death than someone born in early January 1873 who died in late December 1972 and consequently was recorded as a death at age 99 years 11 months. It seems reasonable to hope that there is no consistent bias in such errors.

It will be convenient here to reverse the usual order and assign rank 1 to the oldest death (male, aged 107) and so on, with higher ranks indicating death at younger ages. This reverse ranking permits examination of deaths above age x for different values of x without changing anything important. The hypothesis that female mortality rates are consistently lower than male rates will be rejected if W, the sum of female ranks, is too large.

How large is too large? If we had exact ages at death of n females and m males with no ties and if mortality rates were identical for both sexes at all ages, then the ranks of female deaths would be one of the $\binom{n+m}{n}$ subsets of size n of the integers $1,2,\ldots,n+m$—any such subset with equal probability. Since the sum of all these integers is $\frac{1}{2}(n+m)(n+m+1)$, the average rank is $\frac{1}{2}(n+m+1)$. Under the hypothesis of equal distribution, the expected value of the sum W of female ranks is $\frac{1}{2}n(n+m+1)$, and its variance is $(1/12)nm(n+m+1)$. Critical values for n and m no greater than 25 have been tabulated by Pearson and Hartley [9]. For larger values, the normal approximation is commonly used.

Table 1 gives the results of applying this Wilcoxon test with various initial ages. The ages x listed are calendar ages of female deaths; n is the number of females, and m the number of males, recorded as dying at age x or older. The oldest female death was recorded at age 105 (and 0 months), and assigned rank 3 because two males were recorded as dying at higher ages. The two values given for W are determined by different methods of treating "ties." The larger value, W_0, results from assigning an average rank to each of the individuals in a group with the same calendar age at death, and the smaller value, W', from assuming that, within such groups, the females all died older than any of the males. The difference between these two values is an indication of the extent to which ties affect the outcome.

TABLE 1

WILCOXON ANALYSIS OF CHARTER BENEFICIARIES DATA

INITIAL AGE x	FEMALES SURVIVING TO AGE x n	MALES SURVIVING TO AGE x m	WILCOXON STATISTICS		UPPER CRITICAL VALUE (5%)	NORMALIZED WILCOXON
			W_0	W'		
105.00	1	2	3	3	(none)	1.225
103.92	2	10	15	15	22	.430
103.25	3	13	30.5	30	39	.673
102.83	5	14	67.5	67	69	1.620
102.67	6	15	88	87	88	1.713
102.00	8	23	148	146	166	.903
101.75	9	25	182	180	201	.956
101.50	11	31	260	253	294	.672
100.75	12	41	312	304	402	−.255
100.67	13	41	366	358	439	.172
100.58	14	44	422.5	413	504	.173
100.50	15	46	482.5	472	564	.293
100.33	17	54	619.5	605	734	.101
100.25	18	56	692.5	677	806	.220
100.17	19	58	768.5	752	881	.325
100.08	20	58	846.5	830	934	.647
100.00	21	58	925.5	909	989	.949
99.92	23	60	1,088.5	1,070	1,128	1.246
99.83	24	60	1,172.5	1,154	1,187	1.510
99.75	25	63	1,259	1,239	1,291	1.355
99.67	27	65	1,440	1,418	1,448	1.582
99.58	30	68	1,726.5	1,700	1,699	1.862
99.17	33	78	2,052	2,021	2,103	1.316
99.08	36	80	2,394	2,360	2,382	1.719
99.00	38	83	2,632	2,595	2,613	1.754
98.92	39	87	2,756	2,717	2,788	1.475
98.83	42	88	3,141.5	3,101	3,082	1.944
98.75	46	93	3,681.5	3,631	3,588	2.066
98.67	47	100	3,825	3,771	3,875	1.441
98.50	49	104	4,126	4,068	4,194	1.380
98.33	50	112	4,285.5	4,225	4,529	.763
98.25	52	118	4,618.5	4,552	4,922	.583
97.75	54	143	5,006.5	4,935	5,934	−.951
97.58	56	153	5,419.5	5,344	6,517	−1.189
97.50	61	156	6,487	6,404	7,333	−.390

The upper critical value (at the 5 percent level) was taken from Pearson and Hartley for ages 101.75 and above, and calculated by the normal approximation at younger ages. It is easy to check that the normal approximation would not give very different results for ages at which there are at least 20 survivors. The last column gives the observed W_0 in standard deviational units above the mean. For example, at age 105, with two males

and one female surviving, the mean of W is 2 and its variance is 2/3, so that the observed W_0 is 1.225 standard deviations above the mean ($3 \doteq 2 + 1.225 \times \sqrt{2/3}$).

Note in Table 1 that at ages 99.58, 98.83, and 98.75 even the smaller value W' exceeds the critical value. These ages are interspersed with others at which the result is less significant. Not surprisingly, at ages below 97.75 the observed value of W_0 is less than "expected" under the null hypothesis.

At first glance, the result at age 98.75, more than 2 standard deviations above the mean, seems highly significant. But we must be suspicious of the significance level because of the sensitivity to the initial age x. In fact, we have not used a pure Wilcoxon test but a "worst-case" test, selecting *a posteriori* the initial age that gives the most highly significant result. In order to decide how significant this is, we must examine the distribution not of the usual Wilcoxon, but of this "worst-case" statistic.

Let $W(x)$ be the value of W_0 obtained with initial age x (that is, ranking all deaths at age x and above), and

$$M_W = \max_{95 \le x \le x_m} \frac{W(x) - E[W(x)]}{\text{Var } [W(x)]}.$$

Both expectation and variance are to be calculated under the assumption of equal distributions. The upper limit x_m is the greatest age to which at least one male and one female survive; this keeps Var$[W(x)]$ positive. M_W is the new test statistic, and the null hypothesis will be rejected if M_W is too large. (The data would permit us to take the maximum as x ranges between 91 and x_m. It is virtually certain that the maximum value will occur at an age x larger than 95, since female mortality rates are significantly lower than male rates at the younger ages.)

"Too large" means above the upper critical value (for the selected significance level) of the distribution of M under the equal-distribution hypothesis. In theory, this distribution could be obtained from the known distribution of W, but the analysis could be formidable. A simpler alternative in this case is a Monte Carlo estimate of F_M.

If we had exact values of X_i and Y_j—and if there were no ties—then the distribution of M would not depend on the assumed common distribution of X and Y. Here there are ties, and more occur at the ages where there are more deaths (the younger ages). Nevertheless, the Monte Carlo method used here ignores the effect of the distribution F_X; this simplifies the computations and probably does not affect the result much.

The Monte Carlo program began with 397 "males" and 127 "females" (the numbers of survivors to age 95 in the charter beneficiaries group) and randomly assigned an integer between 1 and 524 to each as a surrogate for

age at death. This produces some ties but not so many as observed in the actual data. The individuals were then ranked according to these random integers, Wilcoxon statistics calculated for groups of survivors to age x, and the value of the test statistic M_W determined. The program output for each run gave, besides the value M, the number of females and males surviving to the "age" that produced the maximum. The reason for this additional information is that the normalizing transformation is essentially equivalent to using the normal approximation to F_W, which is not appropriate if the number of survivors is less than about 20. (An examination of the critical values given in Table 1 will show that, for small numbers of survivors, the exactly tabulated critical value is consistently higher than that produced by the normal approximation.) Of 200 runs, 69 produced M values greater than the observed 2.066. Even after discarding the 7 of these that were based on consideration of fewer than 30 survivors, it appears that about 30 percent of the time we should expect M_W to be at least as large as observed in the charter beneficiaries data, if mortality rates were actually equal for males and females.

Table 2 lists the largest 100 values of M_W produced in 200 runs, with asterisks indicating runs that found a maximum with fewer than 30 survivors.

TABLE 2

MONTE CARLO VALUES FOR M_W, MAXIMUM NORMALIZED WILCOXON STATISTIC
(Largest 100 in 200 runs)

1.81197	1.97253	2.24045	2.66700
1.82160	1.98649	2.24164	2.67017
1.83541	2.00000*	2.24919	2.68860
1.84258	2.00195	2.24994	2.69974
1.84637*	2.00347*	2.26367*	2.72136
1.85164*	2.06589	2.30134	2.72179
1.85164*	2.07786	2.30274	2.72571
1.85164*	2.08768*	2.32617	2.72646
1.85812	2.09212	2.32906	2.73635
1.86039	2.09265	2.35461	2.75880
1.86306	2.09980	2.35470	2.76487
1.86865	2.10273	2.37129	2.82626
1.87023	2.11688	2.37975	2.87222
1.87228	2.11884	2.38062	2.89200
1.88781	2.12013	2.38210	2.91670
1.88837	2.12132*	2.39600	2.94107
1.89443	2.12258*	2.42831	2.97065
1.89945	2.14682	2.44875	3.02093*
1.90553	2.15575	2.46179	3.09094
1.91881*	2.16242*	2.50000	3.21020
1.92879*	2.18054	2.51966	3.24159
1.93742	2.18650	2.52842	3.28290
1.95979	2.18711	2.59808*	3.53641
1.96383	2.23325	2.60736	3.69444
1.96396*	2.23441	2.65826	4.12350

*Runs having a maximum M_W at an age with fewer than 30 survivors.

The significance of the observed value of 2.066 appears to be far less than it first seemed. A value of M_W greater than 2.9 would be required to reject the null hypothesis at the 5 percent level.

II. A MODIFIED KOLMOGOROV-SMIRNOV TEST

The Medicare data presented by Bayo and Faber are more extensive, still "clean," and almost complete. Since exposures were obtained by summing deaths, and some deaths above age 105 were estimated, one eventually must consider the effect of possible errors in those estimates. But first it is necessary to develop a test suitable for this data set, which shows far too many ties for the Wilcoxon test.

The one-sided Kolmogorov-Smirnov test for equality of distribution begins with two empirical distribution functions, $F_X^0(x)$ and $F_Y^0(x)$, and rejects the hypothesis $F_X(x) \leq F_Y(x)$ if the statistic $D = \max(F_X^0(x) - F_Y^0(x))$ is too large. In the case of complete data based on n samples of X and m samples of Y, critical values of D (depending on n and m) are tabulated, for example, by Kim and Jennrich [6]. If n and m are larger than 100, the normalized statistic $\left(\sqrt{\frac{n \cdot m}{n+m}}\right) D$ gives a simple indication of significance.

In analyzing the crossover question, we consider survival beyond age x, let x vary, and look for the most significant result—the largest normalized statistic. Then, as before, we must determine how significant that is.

Define $n(x)$ to be the number of females and $m(x)$ the number of males assumed to survive to age x (the "exposure" figures in Bayo and Faber, Table 6), and let

$$D(x) = \max_t \left(\frac{m(x+t)}{m(x)} - \frac{n(x+t)}{n(x)} \right),$$

$$M_{KS} = \max_x \sqrt{\frac{n(x)\, m(x)}{n(x)+m(x)}}\, D(x).$$

The data permits maximization only over positive integral values for t, and $x = 95.5, 96.5, \ldots, 105.5$. (Left-censoring prohibits our taking $x \leq 94.5$.) Table 3 shows the $D(x)$ values and their normalizations.

The maximum value, 1.65814, occurring at age 102.5, appears to be significant at the 1 percent level. Once again, as in the Wilcoxon case, the issue is clouded by the *a posteriori* selection of an extreme value, and a

TABLE 3

KOLMOGOROV-SMIRNOV ANALYSIS OF MEDICARE DATA

AGE x	$m(x)$	$m(x)$	$D(x)$	t	NORMALIZED KOLMOGOROV-SMIRNOV STATISTIC
95.5	17,014	14,494	.00060	10	.05285
96.5	12,515	10,148	.00180	9	.13456
97.5	8,969	6,908	.00402	8	.25113
98.5	6,309	4,767	.00654	7	.34097
99.5	4,289	3,146	.01156	6	.49236
100.5	2,890	2,049	.02054	5	.71136
101.5	1,890	1,319	.03477	3	.96905
102.5	1,250	829	.07427	2	1.65814
103.5	772	554	.07234	2	1.29923
104.5	453	362	.06871	3	.97469
105.5	234	208	.08333	2	.87447
106.5	148	116			
107.5	84	92			

computer simulation is a convenient way of estimating the distribution of M_{KS} under the hypothesis of equal survival functions for males and females. Such a simulation also will compensate for the fact that the data set here violates several of the conditions usually required for applying the Kolmogorov-Smirnov test: too many ties, distribution function available only at discrete times, and questionable normalization at the highest ages.

Because the Medicare data gives age at death by calendar year only, rather than the monthly intervals available for the charter beneficiaries data, and because of the considerably greater volume of data, there are thousands of ties. Consequently, it is not reasonable to ignore the effect of the distribution of deaths. The simulation suggested here has some similarities to the "bootstrap" method (Diaconis and Efron [3], and Efron, [4]) in using the empirical distribution function as a basis for generating random samples for the Monte Carlo experiment. The major steps are as follows:

1. The observed ages at death (males and females together) are taken as defining an empirical survival function past age 95.5.
2. Beginning with 14,494 "males" and 17,014 "females" at age 95.5, the numbers of deaths at each age are generated as binomial random variables, approximated by the normal. Since the exposure is about 100 lives of each sex, even at age 106.5, the normal approximation is acceptable.
3. The maximum normalized Kolmogorov-Smirnov statistic M_{KS} is calculated as shown previously.

4. Steps 2 and 3 are repeated several hundred times to obtain an estimate of the distribution of M_{KS}. From this we can estimate the significance of the observed 1.658.

The output from 600 runs of this simulation is summarized in Table 4. The observed 1.658 is significant at the 1 percent level, and we can reasonably reject the hypothesis that female mortality rates are consistently lower than male rates for this population.

TABLE 4

ESTIMATED CRITICAL VALUES FOR M_{ks} FOR MEDICARE DATA

α	.50	.25	.10	.05	.01
$F^{-1}_{M_{KS}}(1-\alpha)$	.8318	1.0352	1.1919	1.3196	1.6533
The ten largest values of M_{KS} in 600 runs were:					
1.5289	1.5456	1.6245	1.6533	1.7557	
1.8245	1.8354	1.8452	1.9452	2.0857	

As noted earlier, deaths at ages 105 and up, and thus exposures at earlier ages, were partly estimated from the experience of other cohorts. Could errors in the numbers here invalidate the significance of crossover? It is a simple matter to revise Table 3 by subtracting one constant from the exposures for males and adding another to the exposures for females and, then, to recalculate M_{KS}. In order to lower M_{KS} to 1.32, the approximate critical value for the 5 percent level, it would be necessary to subtract 20 from $m(x)$, or add 30 to $n(x)$, or some combination of these. That is, if imputed male deaths are as much as 10 percent too high, or imputed female deaths about 12 percent too low, our result is still significant. Errors of this magnitude are possible but do not seem likely.

III. A TEST BASED ON GRADUATION

An important objection can be raised concerning the two methods described previously: they only reject (or fail to reject) the hypothesis $q^f_x \leq q^m_x$ without really addressing the more fundamental question of the pattern of mortality rates. What we would like is a means of comparing the entire sequences of rates. Recognizing that the observed rates involve stochastic variation from "true" underlying rates, the actuary's traditional method has been to graduate, then compare.

Unfortunately, the standard graduation methods do not provide estimates of accuracy. Of course, we recognize that different methods may lead to different graduated values (and maybe to a different answer to the crossover question), so we do not claim that the graduated values are the true values, but we need an estimate of how close they are likely to be.

The problem of obtaining confidence bands for distribution functions and hazard functions has been the subject of research by Liu and Van Ryzin [7], [8], Wahba [10], and others in the past few years. The most promising method for the crossover question is that of Cheng and Iles [2]: select a parametric family for the survival function; obtain a confidence region for the parameter values; and look at the envelope of curves determined by parameter values within the region.

The first two data sets presented by Bayo and Faber are particularly well suited to this method because their statistical structure is clear, both being essentially complete studies of narrow cohorts of lives. The third set, using census populations and vital statistics deaths, would require different analysis.

The method used here assumes a force of mortality of the form *(polynomial)* X $e^{\alpha x}$ + *(polynomial)*, with a confidence region for the parameters determined by a likelihood ratio. This involves constructing a likelihood function, estimating its maximum value, and finally taking as a confidence region that region on which the likelihood is sufficiently close to the maximum.

For the charter beneficiaries data, the likelihood function is

$$\mathcal{L}_x = \Pi\, {}_{t_i}p_x\, {}_{\frac{1}{12}}q_{x+t_i},$$

where the product is taken over all individuals of specified sex who survived at least to age x (as approximated by calendar year and month of birth and death). The ith individual died at calendar age $x+t_i$. Initial tests suggest that Gompertz's law gives a maximum likelihood near the maximum obtained with first-degree polynomials (5 parameters instead of 2). Following standard statistical procedures, we would reject the Gompertz hypothesis and use the more complex mortality law only if $\ln\mathcal{L}$ could be improved by as much as $\frac{1}{2}(7.81) = 3.905$ (see example in Elandt-Johnson and Johnson, page 75). The maximization routine functioned poorly with more than three parameters, and it is possible that the observed improvement of only .01 is too low. Nevertheless, it was decided to use Gompertz's law throughout. This choice has some precedent and has an additional advantage in this situation: the crossover age for two Gompertz forces of mortality can be calculated directly.

It was helpful to transform the parameters, expressing age in centuries and setting $\mu_x = a\, e^{(\alpha x+b)}$. These parameters are related to the familiar Gompertz parameters via the equation

$$a\, e^{(\alpha x+b)} = 100\, B\, c^{100\, x}$$

whence $a = 100 \ln c$ $\quad c = e^{a/100}$

$b = \ln(B) - \ln \ln c$ $\quad B = a\, e^{b}/100.$

Then the function we seek to maximize is

$$\ln \mathcal{L} = -e^{ax+b} \Sigma (e^{ai} - 1) + \Sigma \ln(1 - \exp\{-e^{ax+ai+b}(e^{a/1200} - 1)\}).$$

Maximum likelihood estimators for a few initial ages x are given in Table 5. The erratic variation from age to age indicates the instability of the estimators; $\ln \mathcal{L}$ varies quite slowly with the parameters so that widely different values of (a,b) produce nearly equal values of $\ln\mathcal{L}$. As a result, minor changes in survival times result in large changes in the maximum likelihood estimators. This is distressing because the survival times here are only approximate. Fortunately, the crossover age also varies slowly with the parameters. The crossover age is that age at which the force of mortality is equal for the two sexes, thus,

$$a_m\, e^{(a_m x + b_m)} = a_f\, e^{(a_f x + b_f)}$$

whence

$$x = \frac{b_m - b_f + \ln a_m - \ln a_f}{a_f - a_m}.$$

All the calculated crossover ages are about 99. Recall that the Wilcoxon test used earlier showed the most significant result at age 98.75. With this in mind, we look more carefully at the situation with initial age 98.

Table 6 gives values of $\ln\mathcal{L}$ for initial age 98 and some different values of the parameters a and b. The maximum values are approximately -548.953 for males and -217.477 for females. The statistic

$$-2 \ln \frac{\mathcal{L}(a,b)}{\mathcal{L}(\hat{a},\hat{b})}$$

would be approximately $\chi^2(2)$ if a and b were the true values of the parameters, and $\hat{a}$ and $\hat{b}$ the maximum likelihood estimators. Thus, an approximate 97.5 percent confidence region for the parameters consists of those pairs of parameter values that give $\ln\mathcal{L}$ within 3.689 of the maximum, that is, at least -552.642 for males and at least -221.166 for females.

A further transformation of the parameters to $\frac{1}{2}(a-b)$ and $5(a+b)$ facilitates the graphic representation of these regions in figure 1. The "male" region is smaller because the more extensive data on males implies narrower confidence intervals. (The fact that the "male" region lies essentially within the "female" region is coincidental. The maximum likelihood point for the female parameters lies outside the male region, and thus tighter regions, such as the 90 percent confidence regions, would not have so much overlap.)

TABLE 5

MAXIMUM LIKELIHOOD ESTIMATORS FOR CHARTER BENEFICIARIES DATA

INITIAL AGE, x	MALES		FEMALES		CROSSOVER AGE
	a	b	a	b	
98	8.506	−6.867	20.896	−20.053	99.170
97	8.300	−6.629	20.633	−19.780	99.245
96	3.516	−.961	20.357	−19.487	99.576
95	5.738	−3.670	12.332	−10.968	99.066

TABLE 6

SELECTED VALUES OF LN $\mathcal{L}$ FUNCTION FOR CHARTER BENEFICIARIES DATA

MALES			FEMALES			CROSSOVER AGE
a	b	$\ln\mathcal{L}$	a	b	$\ln\mathcal{L}$	
20	−19.2	−552.5	42	−42	−221.0	100.3
15	−13.8	−551.5	20.9	−20.1	−217.5	101.2
9	−7.2	−552.3	10	−8.6	−219.1	129.5
8.5	−6.9	−549.1	8	−6.5	−220.7	67.9
7.5	−5.51	−552.4	7	−5.35	−220.6	18.2
3	−0.12	−552.1	3	0.0	−221.1	none
0.3	+4.8	−551.1	0.5	+3.85	−221.1	219.6

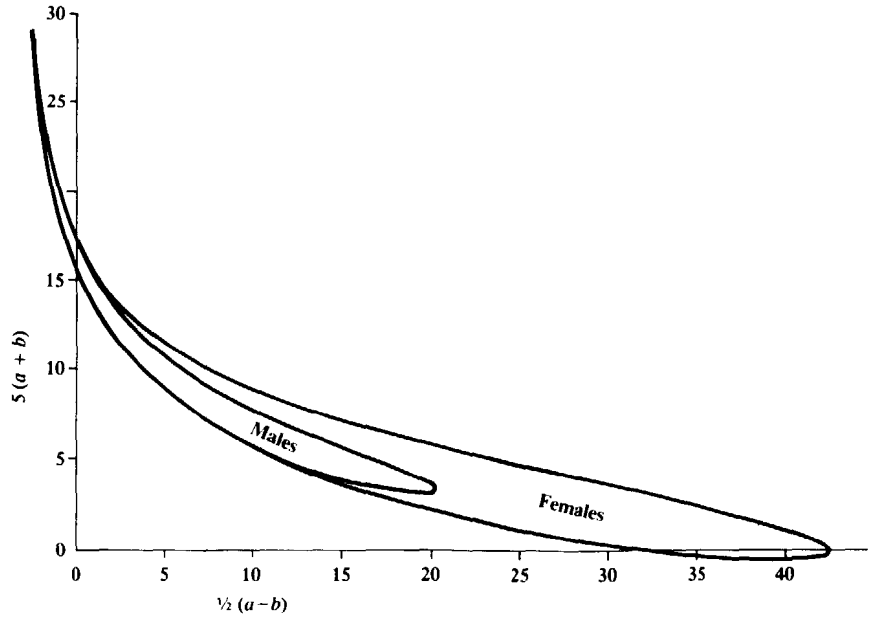

FIG. 1.—95% Confidence regions for transformed Gompertz parameters.

The substantial overlap of the two regions means that we cannot confidently distinguish between male and female mortality patterns for this data set.

If the two regions did not overlap, we could conclude that the mortality patterns were significantly different, and then turn to the question of whether the difference includes a crossover, and at what age. The method of Cheng and Iles [2] would be to construct envelopes for the two force-of-mortality functions, using parameter values in the confidence region. For the crossover question, an easier alternative is to estimate a 95 percent confidence interval for the crossover age. With probability approximately 95 percent, the cross product of the two-parameter confidence regions includes the "true" parameter values a_m, b_m, a_f, and b_f. The maximum and minimum values of crossover age calculated from parameters within this region gives us an approximate 95 percent confidence interval for crossover age. The last column of Table 6 gives the crossover age corresponding to a few pairs of points. It is clear that the calculated crossover age can be arbitrarily large or unreasonably small. Thus, the data here are not sufficient to reject, at the 5 percent level, a hypothesis such as "the crossover age is above 200."

Parametric methods like this one tend to be more powerful than nonparametric methods. In this case, we have found that the parametric method confirms the result of the Wilcoxon test, namely that the data are insufficient to draw a firm conclusion regarding crossover. It is possible that more extensive data (such as that on Medicare recipients) would give confidence regions like those in figure 1 but smaller so that the possible a-values would not overlap, and then we might have a useful confidence interval for crossover age.

Finally, one should note that any conclusion about crossover drawn from any of the tests described here would properly apply only to the particular population studied. To apply it to the general population would require an assumption that the studied population was "typical," an assumption that would appear unwarranted in the case of the charter beneficiaries. The more extensive Medicare data is more interesting from this viewpoint.

Vaupel and Yashin [10] have given an interesting discussion of crossovers and other surprising effects of heterogeneity.

REFERENCES

1. BAYO, FRANCISCO R., AND FABER, JOSEPH F. "Mortality Experience Around Age 100," *TSA* XXXV (1983), 37–59.
2. CHENG, R.C.H., AND ILES, T.C. "Confidence Bands for Cumulative Distribution Functions of Continuous Random Variables," *Technometrics* 25.1 (February 1983), 77–86.
3. DIACONIS, PERSI, AND EFRON, BRADLEY. "Computer-Intensive Methods in Statistics," *Scientific American* Vol. CCXLVIII, No. 5, May, 1983, 116–30.

4. EFRON, BRADLEY. *The Jacknife, the Bootstrap, and Other Resampling Plans*. Philadelphia: SIAM, 1982.
5. ELANDT-JOHNSON, REGINA C., AND JOHNSON, NORMAN L. *Survival Models and Data Analysis*. New York: John Wiley and Sons, 1979.
6. KIM, P.J., AND JENNRICH, R.I. "Tables of the Exact Sampling Distribution of the Two-Sample Kolmogorov-Smirnov Criterion $D_{m,n}$ (mn)." *Selected Tables in Mathematical Statistics*, Volume I, The Institute of Mathematical Statistics, Chicago: Markham Publishing Co., 1970, 79.
7. LIU, Y.C., AND VAN RYZIN, JOHN. "A Histogram Estimator of the Hazard Rate with Censored Data," *Columbia University Statistical Reports, Technical Report No.B*-30, 1983. *Annals of Statistics* XIII, 1985, 592–605.
8. ———. "The Limiting Distribution of the Maximal Deviation of a Density Estimate and a Hazard Rate Estimate," *Columbia University Statistical Reports, Technical Report No. B*-40, 1984. To appear in *Zeitschrift für Wahnscheinlichkeits Theorie*.
9. PEARSON, E.S., AND HARTLEY, H.O. *Biometrika Tables for Statisticians*, Volume II, Table 22. Cambridge: Cambridge University Press, 1972.
10. VAUPEL, J.W., AND YASHIN, A.I. "Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics," *The American Statistician*, Vol. 39, No. 3, August 1985, 176–85.
11. WAHBA, GRACE. "Bayesian 'Confidence Intervals' for the Cross-Validated Smoothing Spline," *Journal of the Royal Statistical Society* (B) 45(1983), 133–50.