TRANSACTIONS OF SOCIETY OF ACTUARIES
1993 VOL. 45

MORTALITY DIFFERENCES BY HANDEDNESS: SURVIVAL ANALYSIS FOR A RIGHT-TRUNCATED SAMPLE OF BASEBALL PLAYERS

HARRY H. PANJER

ABSTRACT

Previous studies have suggested that left-handers exhibit significant excess mortality; their life expectancy is said to be up to nine years shorter than that of right-handers. The methodologies used in such studies have been criticized for not controlling for variables that can significantly bias the results. We study mortality differences for professional baseball players by examining the times of, and ages of, death of more than 6,000 players who died before 1990. Various cohorts are studied separately to recognize time-varying factors such as changes in the overall level of mortality and the changing handedness mix. The methodology uses the reverse time hazard function and recognizes that the individual observations are drawn from right-truncated populations.

Mortality levels are compared at all ages over 20. No significant mortality differences can be detected for any subgroup in the study, although small (but not highly statistically significant) differences can be observed when all cohorts are combined. This difference appears to be due to changes in the handedness mixture over time.

INTRODUCTION

The question of individual differences associated with natural handedness has been studied by several researchers recently. In particular, Coren and Halpern [4] report on apparent large mortality differences between left- and right-handers in a retrospective study of a community in California. They reported a difference of about nine years in life expectancy. These rather surprising results were also reported in Halpern and Coren [7] in a letter to the editor, which generated extensive press attention. One day after publication of the letter, the American Academy of Actuaries released a statement attacking the methodology used in the study.

In a previous study, Halpern and Coren [6] examined the ages of death of left-handed and right-handed major league baseball players and concluded that although there was little apparent difference between the life expectancies of left- and right-handers, the differences were statistically significant. Six letters criticizing various methodological issues in their analysis appeared in a subsequent issue of the same journal (*New England Journal of Medicine* [11]). The results of research on handedness by many authors is reported in Coren [3], who has himself authored or co-authored many papers on differences between left- and right-handedness.

Various theories have been developed to explain apparent differences based on handedness. For example, it is hypothesized that because left-handers live in an environment largely designed by right-handers, they will have more accidents and hence higher morbidity and mortality rates. It is easy for left-handers to appreciate the awkwardness experienced in using items such as power tools that are designed for right-handers. Common sense suggests that accident risk is increased in such situations, but whether such differences in risk level are significant is not generally known. A variety of other theories based on neurological function and other physiological factors are reported in Coren [3].

The purpose of the present paper is to carry out a more detailed study of the mortality differences between left- and right-handers by using a methodology that avoids the weaknesses of previous methodologies that led to possibly erroneous conclusions. As in Halpern and Coren [6], the study is based on professional baseball players for whom natural (as contrasted to learned) handedness is accurately determined.

The study of Halpern and Coren [6] used data on baseball players listed in the fourth edition of *The Baseball Encyclopedia*, edited by Reichler [10], which provides date of birth and date of death (if known) of major league players. For players who are still alive or who are not known to be dead, only information about birth date is given. *The Baseball Encyclopedia* also indicates the throwing handedness and the batting handedness separately of each player. Halpern and Coren [6] computed average ages at death of left- and right-handed players and constructed survival curves for left- and right-handers. While no significant difference could be observed in the average ages at death, their conclusion of a significant difference in mortality was based on the observation that over a series of successive years the calculated survival curve for right-handers was above that of left-handers. Wood [13], however, pointed out that differences as great as the maximum in the Halpern-Coren study would occur in 98 percent of samples chosen randomly from the same population.

Furthermore, it is easy to see the fallacy in Halpern and Coren's argument. Imagine a group of 1 million newborn left-handers and another group of 1 million right-handers. In each year of life except the first and the last, the numbers of left- and right-handers dying are the same. In the first year one more left-hander dies, and in the last year exactly one less left-hander dies. The survival curve for left-handers is always (except in the first and last years) less than the survival curve for right-handers, although clearly there is no significant difference in mortality. Halpern and Coren's method of analysis would conclude that the difference is statistically significant. This erroneous conclusion results from the misapplication of the Wald-Wolfowitz runs test to data series in which the successive terms are highly correlated. The interested reader is referred to standard texts such as Bradley [2, p. 268], which indicate that the successive sample values or observations must be independent. Successive values of a survival function are highly correlated and hence not independent.

In criticizing the Halpern-Coren study of the population of two California counties, Strang [11] points out the implicit assumption that the proportions of left-handers in the population must be the same. If the proportion of left-handers increases over time, relatively fewer old left-handers have the "opportunity" to be observed as deaths at older ages. This biases downwards the average age at death, for known deaths up to a fixed date, of left-handers in the absence of *any* mortality differences. Coren's own work shows that the proportion of left-handers in the population decreases from about 15 percent at age 10 to about 5 percent from ages 50 to 70. One explanation is that older cohorts (persons born in the same time period, such as in the same year) were converted to right-handers. They would then likely be described as right-handed by relatives after death. A similar observation was made by Anderson [1] in connection with Halpern and Coren's 1988 study of baseball players.

A related further criticism of both Halpern and Coren studies is the failure to recognize changes in mortality rates from the late 1800s to the present. The methods used in the studies do not control for the general improving mortality. Differences in the proportion of left- and right-handers over time in the population under study lead to different observed ages at death, even if there are no differences in mortality rates between left- and right-handers.

In addition, Halpern and Coren [7] fail to control for differing levels of mortality and handedness by sex. Females generally exhibit lower rates of mortality than males. Left-handedness appears in a smaller proportion of females than males. These differences potentially bias any results in sampling designs for which explicit control of them is not made.

Implicit in all these criticisms is the recognition that the sample is length-biased due to the study ending at a fixed time, corresponding to the maximum observable age for any person. Persons born at the same time but with different ages at death do not have the same chance of being included in the sample. In statistical terminology, each observed age at death is in the sample *because* it occurred before a fixed age (the age at the end of the study period). This phenomenon is called *right truncation*.

In traditional actuarial studies of mortality, the probability of death in the next year of age for a life aged x is called the mortality rate at age x. It is usually calculated as the ratio of the number of deaths between exact ages x and $x+1$ to the number of exposure units or "potential" deaths, that is, the number of persons alive at age x. Appropriate adjustments are made to the exposures for persons who leave the study, for reasons other than death, before age $x+1$ or who enter after age x.

Such traditional actuarial studies require the knowledge of the exposure basis. Thus a traditional study of mortality by handedness requires the knowledge of both the number of left- and right-handed deaths and the number of left- and right-handed persons in the population under study. In the Halpern and Coren studies [6], [7], no recognition was given to this truncation issue.

In this paper, we show how a mortality study can be conducted without knowledge of the number of survivors beyond the end of the study period. This methodology incorporates the knowledge of the truncation into the underlying model. The methodology is easier to visualize by considering "reverse time," that is, by starting at the highest observed age at death and working backwards through the ordered times of death.

We now show how a quantity closely related to the mortality rate at a specific age can be estimated as the ratio of observed deaths to potential deaths, where the potential deaths are all persons who died at earlier ages but who would have passed through the age in question during the study had death not intervened. The theory is developed in a subsequent section and applied to observed baseball player data.

THE DATA

Previous authors recognized that identification of handedness is probably easiest in sports where preferred hand usage is observable. Baseball players form a large database for whom good records exist. While most players batted and threw with the same handedness, some players batted from opposite sides and some batted both right- and left-handed. However, virtually no players threw with both hands, suggesting that a person can be accurately classified on the basis of a ball-throwing exercise.

Like Halpern and Coren [6], we used baseball player data from *The Baseball Encyclopedia*. They used the fourth edition and used only players who were not exclusively pitchers. In excluding pitchers, they argued that the proportion of left-handers in professional baseball was significantly greater than that in the population. This argument is irrelevant because conclusions based on comparisons of two samples depend on the sample sizes and not on the population sizes. We therefore included all pitchers in the analysis to increase sample sizes. We also used the eighth edition of *The Baseball Encyclopedia* published in 1990 and edited by Wolff [12].

Although *The Baseball Encyclopedia* gives information on all players, we only used the information on the known deaths. This was done because players whose date of death was not given may actually be dead but the death not yet known by the publisher and hence not recorded. Treating all such persons as survivors would bias the results to an unknown extent, rendering the conclusions of questionable value.

Because baseball players in our sample are all male, the potential biases based on sex composition of the sample are avoided. Because the study includes players born from the early 19th century to the mid-20th century and because the handedness composition could change over time, we studied mortality difference for different cohorts of players. For example, we isolated players born before 1880 and studied them separately.

We classified the players (including pitchers) according to handedness as follows:

Classification	Criteria
RR	Batted and threw right-handed
LL	Batted and threw left-handed
RL	Batted right, threw left
LR	Batted left, threw right
Other	Switch hitters and throwers and unknown

The dates of birth and death were recorded for most players. The ages at death were then calculated to the nearest day by differencing the two dates. For a few players, the exact dates of birth or death (usually birth) were not given precisely. If only the month was given, the day nearest to the middle of the month was used as an approximation. Similarly, when only the year was recorded, the same principle was used. Such approximations minimize any error.

PRELIMINARY OBSERVATIONS

Table 1 provides a picture of the average ages at death and the numbers of players in each category in the population of baseball players by period of birth.

TABLE 1

AVERAGE AGE AT DEATH AND NUMBER OF OBSERVED DEATHS FOR EACH CLASS BY PERIOD OF BIRTH

Handedness Classifications	Before 1881	1881–1900	1901–1920	1921–1940	1941–1989	Total
			Average Age at Death			
RR	66.61	69.03	63.27	51.66	32.88	66.13
LL	67.53	68.37	65.64	49.92	33.06	66.49
RL	64.34	71.05	63.30	48.77	n/a	66.39
LR	65.34	69.24	63.14	51.39	31.13	66.11
Other	58.31	65.57	65.07	53.26	29.80	60.27
Total	63.57	68.72	63.70	51.29	32.38	65.20
			Number of Observed Deaths			
RR	934	1,630	801	141	25	3,531
LL	165	345	185	36	2	733
RL	40	70	43	7	0	160
LR	153	409	159	35	5	761
Other	722	237	65	8	3	1,035
Total	2,014	2,691	1,253	227	35	6,220

Note that average ages at death decrease in all categories for birth years after 1881; this is due to the effect of right truncation. Also, note that the average age at death for births prior to 1881 is lower than that for the subsequent period in spite of truncation. This is a result of improving general mortality and the fact that truncation is not significant for early birth years. This suggests that, unless the proportions of left- and right-handers are constant over time, the averages for all years are not meaningful.

There are no obvious patterns of differences of the average lifetimes of left- and right-handers either by cohort or by handedness subgroup.

From Table 1, simple proportions of left-handers can be measured in two meaningful ways, either based on pure left- and right-handers (group LL as a percentage of groups RR and LL combined) or on the basis of left- and right-throwers (groups LL and RL as a percentage of groups LL, RR, RL, and LR combined). These proportions are given in Table 2.

TABLE 2

PERCENTAGE LEFT-HANDERS BY COHORT

Period of Birth	Pure Left- and Right-Handers	Left- and Right-Throwers
Before 1881	15.01%	15.87%
1881–1900	17.47	16.91
1901–1920	18.76	19.19
1921–present	18.63	17.93

The proportion of left-handed deaths increases for the pure group by period of birth and increases for all but the last period for the throwers. In this last period there are few deaths. Note that an additional one death of (pure) left-handers and four deaths of left-throwers would make the two columns monotonic.

As discussed earlier, we made no use of information about those players who were either alive or whose death was not recorded in *The Baseball Encyclopedia*. This might be considered a handicap in developing mortality rates. For reasonably large samples, it is possible to estimate mortality rates or, equivalently, to produce survival curves, at least up to a constant of proportionality. The development of these estimators is given in the next section.

THEORETICAL DEVELOPMENT OF THE ESTIMATORS

Let T denote the time of death for a person known to be alive at age x at time 0. Let $f(t)$ and $F(t)$ denote the probability density function (pdf) and the cumulative distribution function (cdf) of the random variable T. In standard actuarial notation, $f(t) = {}_tp_x\,\mu_{x+t}$ and $F(t) = {}_tq_x$.

The probability that death occurs in the infinitesimal interval $(t, t+dt]$ given that death occurs before time s is then given by

$$\frac{f(t)}{F(s)} = \frac{{}_tp_x\,\mu_{x+t}}{{}_sq_x}, \; 0 \le t \le s.$$

It is convenient to define a function similar to the force of mortality, or the hazard rate μ_{x+t}. Let

$$\lambda_{x+t} = \frac{f(t)}{F(t)} = \frac{{}_tp_x\,\mu_{x+t}}{{}_tq_x}$$

denote the "reverse time hazard" function at age $x+t$. The reverse time hazard function (when multiplied by dt) can be thought of as the probability of death in time interval $(t-dt, t]$ given that death occurs *before* time t. This is analogous to the usual force of mortality or hazard function at age $x+t$ which (when multiplied by dt) represents the probability of death in time interval $(t, t+dt]$ given that the person is alive at (that is, that death occurs after) time t.

From this it is clear that $f(t) = {}_tp_x\,\mu_{x+t} = {}_tq_x\,\lambda_{x+t}$. Thus the conditional probabilty of death in interval $(t, t+dt]$, given that death occurs before time s, is

$$\frac{f(t)}{F(s)} = \frac{{}_tq_x\,\lambda_{x+t}}{{}_sq_x}, \; 0 \le t \le s.$$

Then the probability that death occurs in time interval $(u, r]$, given that death occurs before time s, is

$$\Pr\{u \le T \le r | T \le s\} = \frac{\int_u^r f(t)dt}{F(s)} = \frac{{}_rq_x - {}_uq_x}{{}_sq_x}, \; 0 \le u \le r \le s$$

In the case of discrete outcomes (for example, if times are measured in integral numbers of years, months or days), the probability associated with any outcome is the probability associated with an interval in continuous time. The subsequent development follows that of Kalbfleisch and Lawless [8].

Suppose that the random variable T takes on only discrete values t_1, t_2, t_3, ... Define f_i, F_i and g_i as follows:

$$f_i = \Pr\{T = t_i\},$$

$$F_i = \sum_{j=1}^{i} f_j,$$

$$g_i = \Pr\{T = t_i | T \leq t_i\} = \frac{f_i}{F_i}.$$

Here, f_i and F_i are the discrete time analogs of the pdf $f(t)$ and the cdf $F(t)$. Then g_i represents the probability that death occurred at time t_i, given that the person is known to be dead by time t_i (that is, dead for the entire period after t_i). This is the reverse time analog of the forward time probability, known by actuaries and others as the "mortality rate" given by

$$q_i = \Pr\{T = t_i | T \geq t_i\} = \frac{f_i}{1 - F_{i-1}}$$

which is the probability that death occurs at time t_i given that the person is known to be alive at time t_{i-1} (that is, alive for the entire period before t_i).

From the above, it can be seen that

$$1 - g_i = \frac{F_{i-1}}{F_i},$$

analogous to the familiar forward time expression

$$1 - q_i = \frac{1 - F_i}{1 - F_{i-1}}.$$

Now, consider a single observation of death occurring at $T = t_i$ arising from a random variable with ordered possible outcomes t_1, t_2, . . . with right truncation time t_m; that is, death must occur at or before t_m to be observed. Persons who die after the study period are not observed at all. This contrasts with the case of right censoring at the end of the study. An observation is censored if it is known that the person died after the censoring point, *but is observed to be alive* at the end of the study period.

This knowledge can be used in the study of survival probabilities. In the case of truncation, the individual who dies after the end of the study is not even observed and does not contribute any information about survival.

Then the contribution to the likelihood of this observation at time t_i with right truncation point t_m is given by

$$\frac{f_i}{F_m} = g_i \prod_{j=i+1}^{m} \frac{F_{j-1}}{F_j} = g_i \prod_{j=i+1}^{m} (1 - g_j)$$

where the product is taken over all possible outcome times greater than the observed one. Thus a single death at time t_i contributes to the likelihood function

(i) a factor of g_i at the i-th

(ii) a factor of $1-g_j$ at each point after the i-th, up to and including the truncation point.

Now consider a sample of size n times-of-deaths arising from the same discrete distribution, where the deaths may have different right truncation points (usually the age at the end of a study period). Now let d_j denote the number of observed deaths at exact time t_j, and let n_j denote the number of "potential" deaths at time t_j, that is, the number of deaths occurring *before or at* time t_j with right truncation times *after* t_j. Then, the likelihood for the entire sample is

$$L = \prod_{j=1}^{m^*} g_j^{d_j}(1 - g_j)^{n_j-d_j}$$

where t_{m^*} is the maximum age at which death is observed to occur. More formally, if a_k and b_k are the ages at death and right truncation, respectively, for the k-th person in the sample, we can write $d_j = \Sigma I(a_k=t_j)$ and $n_j=\Sigma I(a_k\leq t_j \leq b_k)$, where the sums are taken over all persons in the sample and $I(x)=1$ if x is true and $I(x)=0$ if x is false.

From the likelihood, it is trivial to see that the maximum likelihood estimator of g_j is

$$\hat{g}_j = \frac{d_j}{n_j}.$$

This is the reverse time version of the standard actuarial "deaths/exposure" estimator of the mortality rate q_i.

From the above development, we now obtain the estimates of the distribution function F_i. Because we can only estimate ratios of successive values of the distribution function at the possible times of death, F_i is estimable only up to a multiplicative constant. In actuarial terminology, we can calculate the successive mortality rates and only estimate the portion of the survival curve up to the highest age for which exposures exist. We normalize the distribution function so that it becomes one immediately after the last observed death (that is, when exposures disappear).

Then, subject to the boundary condition $F_{m^*}=1$, the remaining values of the resulting conditional distribution function are easily calculated as

$$\hat{F}_i = (1 - \hat{g}_{i+1})\hat{F}_{i+1} = \prod_{j=i+1}^{m^*} (1 - \hat{g}_j).$$

This is the reverse time version of the standard Kaplan-Meier product limit estimator of survival to time t_j

$$1 - \hat{F}_j = \prod_{i=1}^{j} (1 - \hat{q}_i).$$

The asymptotic variance of the reverse time estimator of the probability of death up to and including t_j is then

$$\hat{V}ar(\hat{F}_j) = \hat{F}_j^2 \sum_{i=j+1}^{m^*} \frac{d_i \, I(n_i > d_i)}{n_i(n_i - d_i)}$$

This result is identical in form to the Greenwood estimate of the variance of the product-limit estimator of the survival function in the standard forward time situation (see Cox and Oakes [5]), except that only terms where the "potential" deaths strictly exceeds the actual deaths are counted in the sum. This must always occur in the forward time situation, and all lives must enter observation at the same time. Kalbfleisch and Lawless [8] show that a similar restriction is required for the forward time situation with different starting ages (that is, different left truncation ages). Kalbfleisch and Lawless [8] use the reverse time hazard function in connection with nonparametric estimation of the lifetime of automobile brake pads. Lagakos, Barraj and De Gruttolla [9] use similar methods in connection with AIDS.

RESULTS

The methodology described in the previous section is applied to the various subsets of baseball players. In each case, the conditional survival functions, $1-F_i$, of left- and right-handers are compared. In addition, we use the Greenwood variance estimates to construct 99 percent confidence intervals for the "true" survival function at each age for left-handers. The observed survival function is compared with this confidence band for left-handers. If the right-hand survival function lies within the confidence intervals, we conclude that the right-hand survival function adequately describes left-hand mortality.

We initially compared groups RR and LL, the "pure" right- and left-handers for all years of birth combined. We also compared right- and left-throwers by combining groups RR and LR and combining groups LL and RL. We also compared the "pure" left- and right-handers born in various subperiods. The results for the "pure" handedness groups are given in Figures 1–5. Results for the "throwers" are qualitatively the same as those for the "pure" groups.

Figure 3 indicates slightly increased mortality of left-handers at about age 60 (where the curves separate) for the second cohort. Figure 5 indicates slightly increased mortality at ages around 40 for left-handers. However, the survival functions are equal by age 50, indicating slightly increased mortality for right-handers in the high 40s. From a practical point of view, these differences are not important.

In Figure 1, the survival curve for right-handers lies above that for left-handers for most ages, suggesting that left-handers die slightly earlier than right-handers. The right-hander survival curve is close to the boundary of the 99 percent confidence intervals. This again suggests that the mortality difference, however small, is marginally statistically significant. However, when we examine the corresponding Figure 2–5 for each period of birth, the survival curves for left- and right-handers are very similar. The right-hand survival function lies well inside the confidence intervals at all ages. Hence, it appears that there are no observable differences in mortality within any cohort.

To explain the apparent paradox between the cohort-specific results and the results for all years of birth, we examine the composition of the groups by year of birth. The statistics in Table 2 show that the proportion of pure left-handers increased steadily over time. When the cohorts are combined, proportionately more young deaths from later cohorts will be

FIGURE 1

MORTALITY OF PURE LEFT- AND RIGHT-HANDERS FOR ALL YEARS OF BIRTH

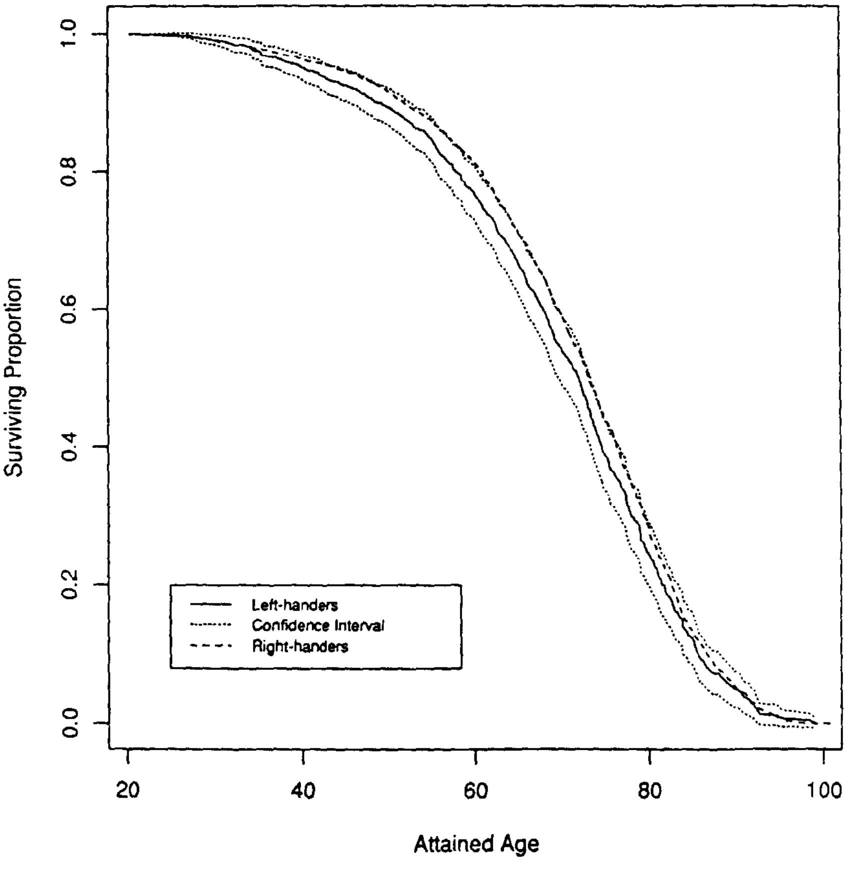


FIGURE 2

MORTALIY OF PURE LEFT- AND RIGHT-HANDERS FOR BIRTHS BEFORE 1881

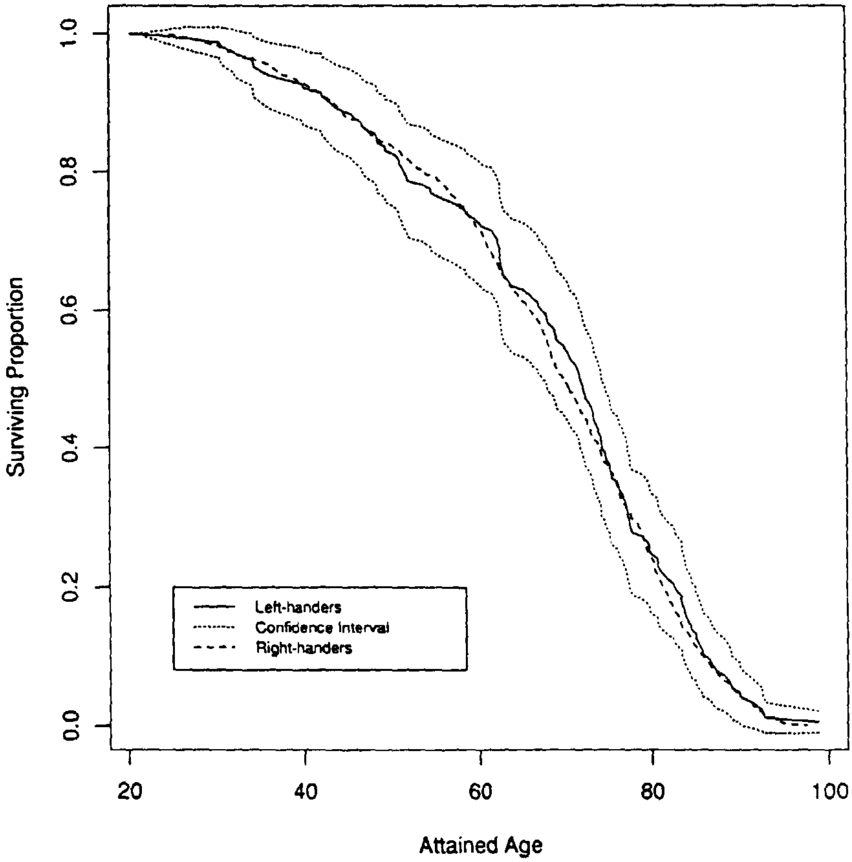


FIGURE 3

MORTALITY OF PURE LEFT- AND RIGHT-HANDERS FOR BIRTHS BETWEEN 1881 AND 1900

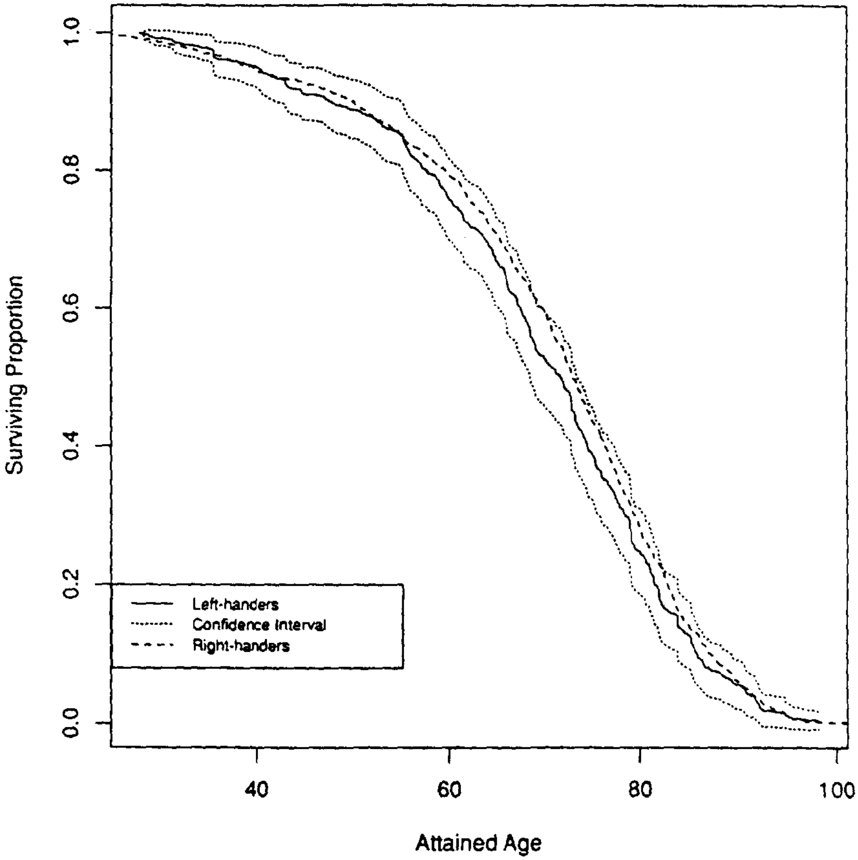


FIGURE 4

MORTALITY OF PURE LEFT- AND RIGHT-HANDERS FOR BIRTHS BETWEEN 1901 AND 1920

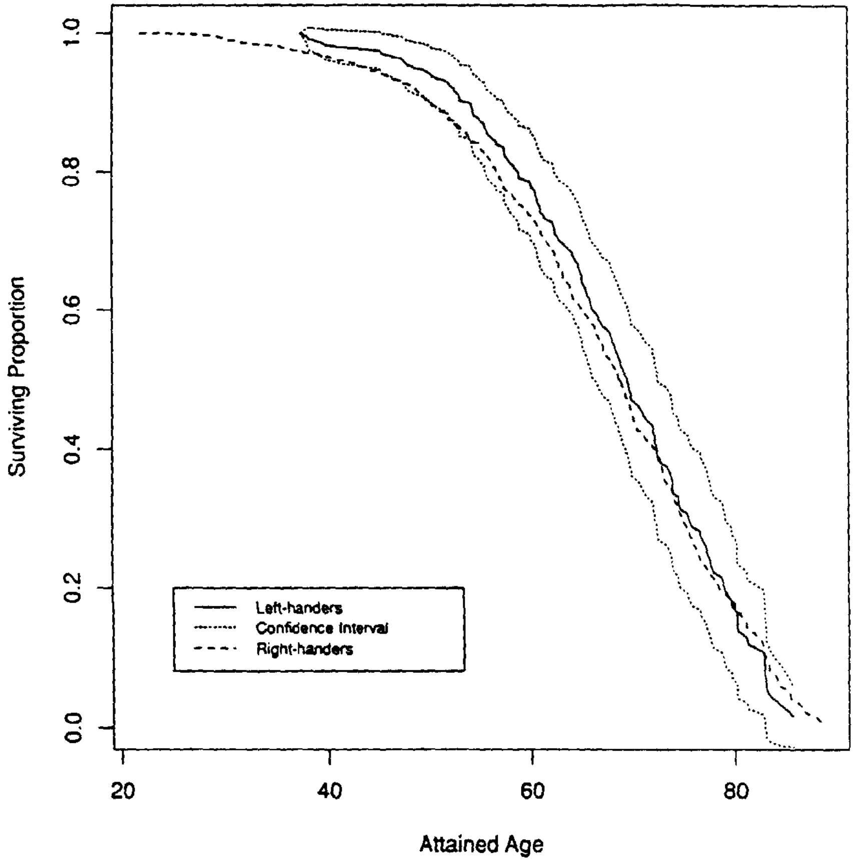
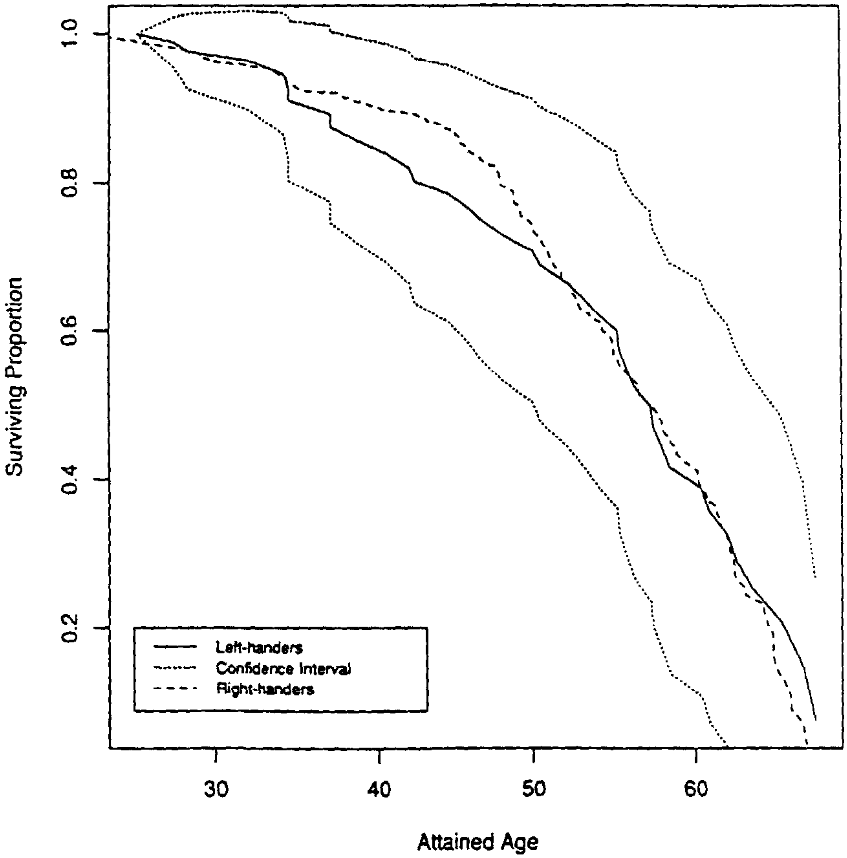


FIGURE 5
MORTALITY OF PURE LEFT- AND RIGHT-HANDERS FOR BIRTHS AFTER 1921



present. This appears to explain the apparent difference in the survival curves when all cohorts were combined.

From the statistics and the figures, we conclude that there is no significant difference in mortality on the basis of handedness.

Of course, we are not interested only in the results for baseball players. The more interesting question is one of extrapolation to the general population. If there is any meaningful difference in mortality between left- and right-handers in the general population, it should be evident in specific subsets of the population. Our subset of the population, namely, baseball players, shows no such difference.

REFERENCES

1. ANDERSON, M.G. "Lateral Preference and Longevity," *Nature* 341 (1989): 112.
2. BRADLEY, J.V. *Distribution-Free Statistical Tests*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.
3. COREN, S. *The Left-Hander Syndrome: The Causes and Consequences of Left-Handedness*. New York, N.Y.: The Free Press, 1992.
4. COREN, S., AND HALPERN, D.F. "Left-Handedness: A Marker for Decreased Survival Fitness," *Psychological Bulletin* 109 (1991): 90–106.
5. COX, D.R., AND OAKES, D. *Analysis of Survival Data*. London and New York: Chapman and Hall, 1984.
6. HALPERN, D.F., AND COREN, S. "Do Right-Handers Live Longer?" *Nature* 333 (1988): 213.
7. HALPERN, D.F., AND COREN S. "Handedness and Life Span," *New England Journal of Medicine* 324 (1991): 998.
8. KALBFLEISCH, J.D., AND LAWLESS, J.F. "Some Useful Statistical Methods for Truncated Data," *Journal of Quality Technology* 24 (1992): 145–52.
9. LAGAKOS, S., BARRAJ, L.M., AND DE GRUTTOLA, V. "Nonparametric Analysis of Truncated Survival Data, with Application to AIDS," *Biometrika* 75 (1988): 515–23.
10. REICHLER, J., ed. *The Baseball Encyclopedia*. 4th ed. New York, N.Y.: Macmillan, 1979.
11. STRANG, J. "Left-Handedness and Life Expectancy," *New England Journal of Medicine* 325 (1991): 1041–42.
12. WOLFF, R., ed. *The Baseball Encyclopedia*. 8th ed. New York, N.Y.: Macmillan, 1990.
13. WOOD, E.K. "Less Sinister Statistics from Baseball Records," *Nature* 335 (1988): 212.