

Chapter VI

STATISTICAL ASPECTS

1. Purpose

The purposes of this chapter are 1) to briefly describe each statistical procedure used in the preceding chapters 2) to state the underlying assumptions of each procedure and 3) discuss the validity of those assumptions in this study. The procedures used in this analysis were survival curve estimates and confidence bands, linear rank tests, relative risk estimation and standardized mortality ratios. Points 1-3 are addressed for each procedure in Sections 2 through 5.

2. Survival Curve Estimation and Confidence Bands

The survival function of a homogeneous population, $S(t)$, is defined as the probability of surviving t years. The problem is to estimate $S(t)$ and make a confidence statement about that estimate based on randomly censored data. Randomly censored data occur in survival studies since analyses are usually carried out before all subjects have failed. In the present application, failure is defined as death and censorship occurs because most subjects are still living at the time of analysis. Other causes for censorship in this kind of epidemiological study are loss to follow-up or death from causes other than those of interest. Thus far in this study, there have been no subjects lost to follow-up, and all causes of death are of interest.

The survival function is estimated here by the product limit estimate $K(t)$, also called the Kaplan-Meier estimate (6). This estimate is derived under the assumption that, in a life testing experiment with n subjects on test, exactly k subjects, with k less than n , are observed to fail; the other $n-k$ remaining are observed only until they are censored. The subjects are assumed drawn randomly from a homogeneous population. Censorship is assumed to be independent of failure. The Kaplan-Meier estimator is asymptotically unbiased and reduces to one minus the empirical distribution function in the absence of censoring.

In the present application, the homogeneous populations are the Ranch Handers, the comparisons and various subgroups of these two groups. Death time is taken as age at death measured to the nearest month; censoring time is age on 31 December 1982, measured to the nearest month. Survival time is age at death or age on 31 December 1982 for those subjects still living.

The process $n[K(t)-S(t)]$ converges weakly to a zero mean Gaussian process, as n tends to infinity, under random censorship when the underlying survival function $S(t)$ and the censoring distribution are continuous on a bounded interval (27). This convergence is the theoretical basis for the confidence band algorithm (7) used in Figures 2 and 3, Chapter III and Appendices VI.

The independence of death and censorship can be assumed to hold here since censorship (survival to December 31, 1982) is not being invoked on individuals because they appear to be at unusually high, or low, risk of death (28). Direct contact has been lost with two Ranch Handers and nine comparisons as described in Chapter II, but these are assumed to be alive, and hence censored at their age on 31 December 1982. The reason for this assumption is that the extensive death ascertainment system is believed to be thorough enough so that, had any of these subjects died, the death would have been detected. Hence, while contact has been lost, loss to follow-up for the purpose of mortality determination has not occurred (29). All other subjects still alive on 31 December 1982 are censored at their age on that date.

The validity of inferences based on the estimate $K(t)$ and its associated confidence band depends on the sample size and the observed number of deaths. The sample sizes and numbers of deaths in every stratum used in these analyses exceed the minimum requirements for these procedures (7).

The survival curve estimates and confidence bands displayed in Figures 2 and 3 and Appendix VI are not adjusted for year of birth. To do so would have required stratification on year of birth, creating many small strata with associated sample size difficulties. Some year-of-birth adjusted plots in the larger occupational strata will be presented in the next report.

3. Linear Rank Procedures

The hypothesis of interest in this analysis is that the actual survival distributions of the Ranch Handers and their matched comparisons are identical. The procedures of choice for testing equality of the two unknown survival distributions based on the matched and censored data in this study are the censored data extensions of the exponential scores and Wilcoxon tests, due to Prentice (8). The first of these is widely known as the logrank test. The test statistics, T , are of the form given by equation 6-23 of (28), where the summands are calculated on matched sets consisting of survival information on one Ranch Hander and his matched mortality comparisons. The statistic T , for either logrank or generalized Wilcoxon summands, is approximately standard normal under the null hypothesis (9).

The large sample normal approximation for T will hold when all distributions are continuous and all censoring times are mutually independent of each other and independent of death. These assumptions are well satisfied in this study since the censorship mechanism, survival to time of analysis, does not favor one group over the other.

In these procedures, the sampling unit is a matched set, so that these tests are adjusted for all matching variables. Prior to calculation, matched sets with Ranch Handers in the same race and job classification having the same year of birth are merged.

The logrank and extended Wilcoxon tests are locally most powerful when the logarithm of the survival times are distributed as extreme value or logistic random variates, respectively. While the efficiency of these procedures peaks at these two underlying distributions, they have been shown to be robust against departures (8). These distributional assumptions, however, are not viewed as strictly valid in this study since there is good evidence in the literature that survival time due to certain cancers and other diseases is log normally distributed (30, 31, 32, 33). A linear rank procedure of the Prentice form, whose efficiency peaks under the lognormal distributional assumption, can be constructed (34), but this algorithm is not available at the present time; it will be included in the next analysis. The effect of this departure from the assumptions is considered mild. It should also be noted that these distributional assumptions cannot be checked since these match sets are small and the observations in the combined samples of all matched sets cannot be assumed to have a common distribution. Therefore, reliance must be placed on historical data to determine which linear rank procedure to use. The logrank and Wilcoxon procedures are used here because they are powerful and widely accepted in epidemiology and statistics.

4. Relative Risk Estimation

Two relative risk estimators are used in this analysis, a generalization of the Ejigou-McHugh estimator for one to many matched data (12) and the Mantel-Haenszel estimator for stratified data (14). The Ejigou-McHugh estimate was chosen because it allows full adjustment for the one-to-many year-of-birth matching in this study, it is asymptotically as efficient as the maximum likelihood estimator and it is noniterative. The Mantel-Haenszel estimate was chosen because of its ease of calculation, efficiency (35), and general acceptance. Its variance is estimated according to the advice of Anderson et al. (36). Recent work suggests that the variance of the Mantel-Haenszel statistic might be better estimated by a jack-knife procedure (37); this newer method will be carried out in the next mortality report.

The Ejigou-McHugh estimator in its published form is suitable only for 1 to R matched designs in which the number, R, of controls matched to each case is the same for all cases. Since the number of controls matched to each Ranch Hander is not the same for all Ranch Handers, the Ejigou-McHugh estimate and its variance was extended to a one-to-many matched design in which the number of comparisons is allowed to vary from case to case. Since this extension is unpublished it is stated in Appendix V for reference.

The extended estimate and its variance reduces to the Ejigou-McHugh estimate and variance when all matched sets contain an equal number of comparisons. It is asymptotically efficient and consistent and is noniterative.

The Ejigou-McHugh estimate and the Mantel-Haenszel estimate are based on the assumption that relative risk is constant across levels of the matching variable. Some indication that this assumption holds in this study when the data is grouped, by stratifying on year of birth, is furnished by likelihood ratio testing; there is no evidence in this study to suggest that relative

risk is not constant across levels of the matching variables when the event of interest is death from any cause. Therefore, the Eijou-McHugh and Mantel-Haenszel estimates are appropriate for these data.

5. Indirect Standardization

With either an external or internal standard, the SMR is a good summary mortality index for comparing two or more populations, provided the product model, $P_{ij}=r_i p_j$, holds, where P_{ij} is the probability of death in stratum i of population j , r_i is a set of standard stratum specific rates and p_j characterizes the mortality of population j , $i=1,2, \dots, I$, $j=1,2, \dots, J$, (38, 13). If standard rates are known from some external source and if the product model holds, the best estimate of p_j is proportional to the SMR. If $J=2$, the product model holds, and if one of the two groups is used as the standard, the SMR estimates relative risk. In any case, any SMR summary of mortality data should be preceded by analytical and graphical tests of fit of the product model. Because one of the study groups was always used as the standard in these analyses, the test of fit of the product model was, equivalently, a test of constancy of relative risk across year of birth strata. The fit of the model was verified in each analysis. Further, a likelihood ratio test for equality of population was carried out as described by Gail (13). The results of both tests are summarized by their P-values in each application. The sample sizes in every application are large enough so that chi-square approximations hold; these analyses are, therefore, valid and appropriate.

The expected number of deaths in the SMR used in these analyses was calculated as $\sum n_{ij} r_i$, where n_{ij} is the number of subjects in the i th stratum of the j th population. The person-years SMR was not used here for two reasons. First, its validity as an estimator of relative risk is dependent upon the fit of the proportional hazards model for which an omnibus test is not currently available. Secondly, the person-years calculation is typically carried out from entry into follow-up (5); in this study, follow-up begins at first entry to Vietnam or Southeast Asia and these entry dates are being verified at this writing.

6. Comparing Observed Life Table Data with a Known Survival Curve

The procedure of Gail and Ware (17) is used in these analyses to compare Ranch Hand and comparison group survival data with published period life tables. The basic assumptions of this procedure are that death and censorship are independent competing risks and that the reference curve is a survival distribution for some external population. The test is of the form $\sum (o_j - e_j) / (\sum v_j)^{1/2}$, where o_j and e_j are observed and expected numbers of deaths in age interval j , and v_j is the variance of $o_j - e_j$. The statistic is not an omnibus goodness-of-fit test consistent against all alternatives to the null hypothesis that the observed sample comes from a known survival distribution. Rather, it has good power against proportional hazards alternatives or, more loosely, against alternatives for which the observed survival is better (or worse) in every interval than predicted by the known survival curve.

The independence of death and censorship assumption is well satisfied in these data, as discussed in Section 2 of this chapter. The life tables used in these analyses do not, however, represent the survival distribution of any population since they are period, not cohort, life tables. The appropriateness of this procedure is, therefore, dependent upon the extent to which these period life tables approximate the survival distribution of some relevant reference population. These period tables were used because the more appropriate cohort life tables were not available at the time of analysis.