Large Sample One-Sample and Two-Sample Tests for Average Incidence Rates

John J.Hsieh

University of Toronto

1. Introduction

Both one-sample and two-sample tests for average incidence rates in censored follow-up studies are frequently used in statistical analysis of epidemiological and other observational studies (see Breslow and Day 1987 and the references cited therein). One-sample tests are usually used to test departure of an observed incidence rate in a study population from that of a general population or of some reference population in order to detect excess morbidity or mortality in the study population. Two-sample tests are used to compare incidence rates of two study groups for measuring effect and association and for comparing event intensities between populations.

In censored follow-up studies, data on event counts and person-time of exposure are available. The basics of the statistical inference procedures based on such cohort data have been described and summarized in Breslow and Day (1987), Sahai and Khurshid (1996) and Rothman and Greenland (1998). The aims of this article are 1) to expand and generalize in a systematic fashion what has been described therein, 2) to bring about properties of various test statistics obtained and, in particular, 3) to point out that analysis of such person-time data arises most naturally from counting processes martingale techniques. We systematically derive many more asymptotic one-sample and two-sample test statistics -- including various types of Wald's statistics, score statistics and likelihood ratio statistics -- using counting processes martingale techniques and likelihood methods applied to the square root and logarithmic transformations of the incidence rate.

These statistical inference procedures yield good statistical properties (such as consistency, asymptotic unbiasedness, efficiency and asymptotic normality). In addition, square root and logarithmic transformations of rates improve convergence to normality. We obtain eight test statistics classified in four groups for the one sample test and five test statistics classified in three groups for the two sample test. Since the asymptotic test statistics so derived give different values with transformed test statistics being more accurate than the untransformed test statistics , we recommend use of the transformed test statistics to make inference and draw conclusion in specific applications, with particular emphasis on specific transformed test statistics when test results give near borderline values.

2. Incidence Rates

It is well to distinguish, at the outset, various types of incidence rates and to identify the appropriate ones for use in statistical analysis. First of all, we shall distinguish between the *net* rates and the *crude* rates. Let X be the waiting time to the occurrence of the event of interest, in the absence of censoring (competing risks included). The incidence rates associated with the distribution of X are the net rates. There are four different types of net incidence rates in use. First, the (net) *instantaneous incidence rate* is the (net) hazard rate or (net) hazard function $h(t)$. It is the instantaneous probability per unit time of failure (occurrence of event) at time t given that the event has not occurred up to t, in the absence of censoring; namely,

$$h(t) = [1/P(X > t)][dP(X \leq t)/dt] = - [1/S(t)][dS(t)/dt], \qquad (1)$$

where $S(t) = P(X > t)$ is the survival function of X. If the event of interest is death, then $h(t)$ is called the *force of mortality*. If the event of interest is the first occurrence of a disease, then $h(t)$ is called the *force of morbidity*.

Second, the *(net) cumulative* or *integrated incidence rate* $H(t)$ is the cumulated sum of the conditional hazard probabilities $h(u)du$ up to time t, defined as $H(t) \equiv \int_0^t h(u)du$.

Third, the *(net) overall* or *average incidence rate* r is the weighted average over the entire follow-up period $[0,\tau]$ of the instantaneous incidence rate $h(t)$ with the expected proportion of person-time of exposure $S(t)dt / \int_0^\tau S(u)du$ at each follow-up duration t as weights:

$$r = \int_0^\tau h(t)S(t)\, dt / \int_0^\tau S(t)\, dt, \qquad (2)$$

where τ is the maximum length of indivivual follow-up periods in the cohort.

Fourth, the *specific incidence rate for the jth follow-up interval* r_j is the weighted average over the jth follow-up interval $(t_j, t_{j+1}]$ of the instantaneous incidence rate $h(t)$ with the expected proportion of person-time of exposure $S(t)dt / \int_{t_j}^{t_{j+1}} S(u)du$ at each follow-up duration t within the jth follow-up interval $(t_j, t_{j+1}]$ as weights:

$$r_j = \int_{t_j}^{t_{j+1}} h(t)S(t)\, dt / \int_{t_j}^{t_{j+1}} S(t)\, dt, \qquad (3)$$

for j =1, 2, ..., k, where the entire follow-up period (o, τ] is partitioned into k follow-up intervals with $t_1 = 0$.

Note that the cumulative incidence rate $H(t)$ is a dimensionless quantity while the other three net rates $h(t)$, r and r_j all have dimension "per unit time". From (2) and (3), it follows that the overall incidence rate r can also be expressed as the weighted average of the specific incidence rate r_j with the expected proportion of person-time of exposure $\int_{t_j}^{t_{j+1}} S(t)dt / \int_0^\tau S(t)\, dt$ for the jth follow-up interval as weights:

$$r = \Sigma_{j=1}^{k} r_j \int_{t_j}^{t_{j+1}} S(t)dt \,/ \int_0^{\tau} S(t)\, dt \,. \qquad (4)$$

It should be noted that the random variable X is not observable and so the net rates defined above are not identifiable without further assumptions. What is observable is the pair $\{T^*, \delta\}$, where T^* is the waiting time to the occurrence of the event of interest, in the presence of censoring and δ is the mode of realization of T^* with $\delta = 1$ if the event of interest (failure) occurs and $\delta = 0$ if censoring occurs. The incidence rates associated with the distribution of $\{T^*, \delta\}$ are called the crude rates. The *crude instantaneous incidence rate* is the crude hazard rate or crude hazard function $\lambda(t)$. It is the instantaneous probability per unit time of failure (occurrence of event of interest) at time *t*, in the presence of censoring, given that neither the event nor censoring has occurred up to *t*; namely,

$$\lambda(t) = [1/P(T^* > t)][dP(T^* \leq t, \delta = 1)/dt] \,. \qquad (5)$$

Other crude rates may be defined in terms of $\lambda(t)$ along similar lines to the definitions of the corresponding net rates.

Crude rates are observable but they do not provide true measures of the intensity of the event of interest, as they are disturbed by the presence of censoring. Our real interest lie in the net rates, as they are not disturbed by censoring and so provide the true picture of the event. For reason of simplicity, we assume that the censoring is noninformative (see, e.g., Kalbfleisch, 1980) so as to be able to identify the net rates. This implies that $\lambda(t)$ is identical to $h(t)$, as the right side of (5) is equal to that of (1) under such assumption. The other net rates H(t), r_j and r defined above can now be identified and calculated by replacing the unobservable $h(t)$ by the observable $\lambda(t)$ in their defining formulas.

To compare the overall levels of event intensity of one or more follow-up populations, analysis may be made either on the (net) cumulative incidence rate or on the (net) average incidence rate. Modern counting process martingale approach has been used to make inference on the cumulative rate (Andersen et al 1993, Fleming and Harrington 1991, Becker 1989). The martingale approach will be used to make inference on the average incidence rate in this article.

3. One-Sample Tests

Suppose we are interested in comparing the (net) average incidence rate r of a follow-up group with the "expected" rate r_o derived from a reference population's (net) hazard function . In other words, we want to test the hypothesis $r = r_o$. The event of interest for which the incidence rate is defined may be the first occurrence of a particular disease, death or some other nonrecurrent vital event. Because of the presence of censoring (which includes both planned and unplanned censoring as well as competing risks) in follow-up studies, the follow-up durations of individual cohort members will vary in length. We observe the pair of data $\{T_i^*, \delta_i\}$, for fact that

basic assumptions are made: (i) that every individual within a given stratum has independent