Published in final edited form as:
Clin Trials. 2008 ; 5(6): 631–634. doi:10.1177/1740774508098327.

Applying Hodges-Lehmann scale parameter estimates to hospital discharge times

Jonathan J Shuster[a,b], **Douglas W Theriaque**[b], and **Brian M Ilfeld**[c]

[a]Department of Epidemiology and Health Policy Research, University of Florida, Gainesville, Florida, USA

[b]General Clinical Research Center, University of Florida, Gainesville, Florida, USA

[c]Department of Anesthesiology, University of California San Diego, La Jolla, California, USA

Abstract

Background—Clinical trials aimed at shortening the time-to-discharge need to have rational and easily understood effect size estimates for health care management organizations. A natural choice is a scale model, where the distribution of time to discharge on one treatment is assumed to be the same as ρ times that on the other. If, for example $\rho = 0.6$, then one treatment is associated with a 40% reduction in discharge time. In designing and analyzing these studies, we need to have the capability to accommodate right censored data, as it is plausible that some patients may never meet the discharge criteria, even though they do get discharged.

Purpose—Utilizing the ideas of Hodges and Lehmann, to provide methods for analysis of trials aimed at shortening hospital discharge times, using point and interval estimates of scale parameters based on the Gehan generalization of the Mann-Whitney-Wilcoxon test, which accommodates right censoring for situations where patients never meet discharge criteria ($+\infty$).

Methods—For every value of $\rho > 0$, we shall test the null hypothesis that the distribution of discharge times on one treatment is the same ρ times that on the other. The values of ρ that we fail to reject form the confidence interval for the true ρ.

Results—The methods were developed and applied to a real clinical trial for times to meet the three objective discharge criteria in knee replacement surgery for two post-operative pain control strategies (usual care plus a perineural infusion of either placebo or 0.2% ropivocaine, until the morning following surgery). Based on 48 randomized patients, the point estimate (95% confidence limits) for ρ was 0.47 (0.32-0.67), favoring ropivocaine.

Limitations—The methods cannot as yet be applied to group sequential designs or studies with more than two treatments.

Conclusion—This methodology is an effective way to analyze two-arm trials involving continuous hospital discharge time data.

Introduction

In randomized trials designed to improve the objective time to meeting hospital discharge criteria, investigators need to provide the public with estimates of the effect size. These studies are often small, and outliers are common. In fact, since nonmedical factors such as third-party

Author for correspondence: Jonathan J Shuster, Department of Epidemiology and Health Policy Research, University of Florida, PO Box 100177, Gainesville, Florida, 32610-0177 USA. E-mail: jshuster@biostat.ufl.edu.

payer hospitalization limits exist, patients may never meet the discharge criteria, although they do get discharged. Although it might seem that the proportional hazards model of Cox [1] might be ideal for this situation, the problem is that the hazard ratio is not an effective tool for providing a meaningful effect size definition in this setting. An intuitively appealing nonparametric approach, which allows right-censored data, is to estimate a scale parameter, adapting the Hodges-Lehmann [2] method to the Gehan [3] statistic. If there were no censoring, this would be identical to using the Mann-Whitney-Wilcoxon test [4], as detailed in Lehmann [5].

Let the competing treatments have cumulative distributions $F_1(t)$ and $F_2(t)$, where

$$F_1(t) = F_2(\rho t). \tag{1}$$

Under this scale model, known in engineering as the 'accelerated life model', if $\rho = 0.6$, then the time to discharge under Treatment 1, has the same distribution as 60% of the time to discharge under Treatment 2, for a 'savings' of 40%. This concept of effect size is very understandable to patients, the government, and potential insurers. In the next section, the methods are given. Numerical examples follow the methods section. The final section presents a brief discussion.

Hodges-Lehmann point and interval estimates of scale

If a randomized double-blind trial is conducted with observed times to discharge $X_1, ..., X_n$ for Treatment 1 and $Y_1, ..., Y_m$ for Treatment 2, the Hodges-Lehmann [2] method tests the null hypothesis for every value of $\rho = \rho_0$ in (1):

$$H(\rho_0) : \rho = \rho_0, \tag{2}$$

by the two-sided Gehan test [3].

We shall first describe the Gehan test for the special case where $\rho_0 = 1$, that is to test the null hypothesis than the *X*'s and *Y*'s have the same distribution. The Gehan test can be thought of as a permutation test of the 'Gehan ranks', calculated as follows:

For each subject, there are $(m + n - 1)$ contests, one against each of the remaining subjects. The rank score for subject *j* is computed as follows:

Subject *j* wins over subject *k* (contribution = +1) if either (W1) subject *j* is uncensored and subject *k* is censored; or (W2) subjects *j* and *k* are both uncensored, but subject *j* has a shorter time to discharge.

Subject *j* loses to subject *k* (contribution = -1) if either (L1) subject *j* is censored and subject *k* is uncensored; or (L2) subjects *j* and *k* are both uncensored but subject *j* has a longer time to discharge.

Subjects *j* and *k* are tied in all other circumstances (contribution = 0).

The rank score for subject *j* is the net score (wins minus losses), and the groups are compared by a permutation test based on the sum of the Gehan ranks in the first treatment. (This is mathematically equivalent to basing the test on the sum of the ranks in the second treatment, or on the difference in mean ranks between the two treatment groups).

A large sample normal approximation for the test statistic, under the null hypothesis that the X's and Y's have the same distribution is given by:

$$\sum R\left(X_j\right) \mathcal{N}\left\{0, \frac{nmV^2}{(n+m-1)}\right\} \tag{3}$$

where $V^2 = [\Sigma R^2(X_j)+\Sigma R^2(Y_k)]/(n+m)$, with $R()$ the Gehan Ranks. V^2 Represents the population variance of the Gehan ranks in the pooled sample. The p-value is calculated from (3), with large absolute values of $\Sigma R(X_j)$ as significant (e.g., above 1.96 standard errors for a 5% test).

In principal, to construct the confidence interval, we test the null hypothesis that $X \sim \rho_0 Y$, for every value of ρ_0, proceeding by replacing each Y by $\rho_0 Y$, and applying the Gehan test to the new data. Those ρ_0 values in the acceptance region for the two-sided α-level test form a 100 (1-α)% confidence interval for ρ. A point estimate can be obtained as the midpoint of the ρ_0 values, where the p-value is closest to 1.00.

Numerical examples

A double-blind placebo controlled drug study, reported by Ilfeld and colleagues [6] had times (hours to meet discharge criteria) of 23 patients on drug of:

22 25 21 87 28 20 47 23 68 20 31 20

19 22 23 30 27 54 51 18 53 46 23

and of 25 patients on placebo of

71 97 21 46 43 78 46 68 47 70 78 23

95 29 127 101 31 89 79 76 148 94 54 78

30

Despite planning for censoring, quite unexpectedly, none actually occurred on this trial. The point estimate for ρ, based on our macro is 0.47 (95% confidence interval from 0.32 to 0.67).

Purely for the purpose of illustrating the method, we modified the data, one with light censoring (adding three censored patients to the better performing treatment, drug, and one patient to the control) and heavy censoring (adding 10 censored patients to drug and 15 censored patients to placebo). The point estimate (95% confidence interval) changed only slightly for light censoring 0.51(0.34-0.75) and markedly for heavy censoring 0.39(0.22-0.73). Note that as expected, the point estimate moved slightly toward the null value of 1.00 for this light censoring example, and away from the null for the heavy censoring example. It should be noted that situations with substantial censoring risk would be synonymous with trials with very serious adverse events, and trialists would be far more worried about treatment safety and efficacy than duration of hospital stay. In response to a reviewer's question about the adequacy of the large sample normal approximation, we repeated the three numerical examples by using a permutational approach with 10,000 replications, and all three point estimates were the same to two significant digits. The only change in the interval estimation was the lower limit for the heavily censored case, which moved from 0.22 (simulated p = 0.0378) to 0.23 (simulated p = 0.0465).

Discussion

While the Hodges-Lehmann [2] method was designed for location parameters and is easily adopted to scale situations via logs and antilogs, it has not achieved much usage in the past,

perhaps due to a perceived lack of user-friendly software. In the uncensored case, one can employ STATXACT, version 6.0, available from http://www.cytel.com/home/default.asp in this manner, yielding an exact point estimate of 0.465 and exact 95% confidence interval (0.324-0.667), virtually identical to our large sample result. Free user-friendly software that perform the calculations based on the Gehan estimates of scale have been made available at http://ehpr.ufl.edu/sample_size_programs. This site also has an annotated user-friendly sample size program for use with the Mann-Whitney U-Test (twosample.sas), a routine that can perform a sample size computation to design a study for hospital time-to-discharge data. Power analysis can also be done per Chapter 2 of Lehmann [5], for any value of ρ, fully specifying the two cumulative distributions as $F_1(t) = F_2(t)$ and $F_2(t)$. For censored situations, it is recommended that the planned sample size derived from twosample.sas macro be inflated by a factor of 1/(1-C), where C is the fraction anticipated to be censored. This will produce a conservative sample size, with slightly greater power than advertised under no censoring.

The associate editor pointed out that Louis [7] addresses the accelerated life model (scale) example for censored data. His work is somewhat more complex, using the logrank test for statistical inference. A motivation for the Gehan test over the logrank test is as follows. In these trials, all censoring occurs after the last uncensored event time, making the Gehan rank a linear function of the pooled Kaplan-Meier estimate (and therefore mathematically equivalent to the permutation test of these Kaplan-Meier estimates) in the following context: (a) for single uncensored event times, the outcome value is the Kaplan-Meier estimate taken just before that time; (b) if ties occur, we temporarily break the ties and proceed as in (a), but then average these Kaplan-Meier values as the outcome value for all of these uncensored event times; (c) for the censored times, we temporarily make them untied late failures and proceed as in (b). For example, if we had times of 13, 14, 14, 18, 20, 23, 25, 30+, 30+, and 30+, the pooled Kaplan-Meier estimates are, respectively, 1.00, 0.85, 0.85, 0.7, 0.6, 0.5, 0.4, 0.2, 0.2, and 0.2. The uncensored observation at 18 ‘loses’ to three (13, 14, 14) and ‘defeats’ six for a Gehan rank of +3. The censored observations lose to seven and defeat none for a Gehan rank of -7. In general, the Gehan Rank = 20*(Kaplan-Meier value) -11. The logrank test is a large sample permutation test of each subject’s observed minus expected value. Even though the observed values for these yes/no events are either zero or one, the logrank expected value for an individual can exceed 1.00. In the above example, it is 1.082 for the four observations 25, 30 +, 30+, and 30+.

In conclusion, we expect trials aimed at reducing hospital stays will become more common in the future. Since quantitative estimates of savings in hospital stay time are imperative to driving public policy, it is not sufficient to merely test for significance between treatment strategies. Given the continuous nature of hospital stays in today’s health economy, the accelerated life (scale) model appears to be an ideal way to express the relative savings of one method compared to another. Further research, to extend this to Group Sequential trials, to obtain Group Sequential point and interval estimates could lead to valuable contributions in this area.

Acknowledgments

Work was funded by National Center for Research Resources General Clinical Research Center grant RR00082 and National Institute of General Medical Sciences grant GM077026. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of these entities.

References

1. Cox DR. Regression models and life-tables (with discussion). Journal of the Royal Statistical Society, Series B 1972;34:187–220.
2. Hodges JL, Lehmann EL. Estimates of location based on rank tests (Ref: V42 p1450-1451). Annals of Mathematical Statistics 1963;34:598–611.

3. Gehan EA. A generalized Wilcoxon test for comparing arbitrarily singly censored samples. Biometrika 1965;52:203–23. [PubMed: 14341275]

4. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics 1947;18:50–60.

5. Lehmann, EL. Nonparametrics: statistical methods based on ranks. Holden-Day; San Francisco: 1975.

6. Ilfeld BM, Le LT, Meyer RS, et al. Ambulatory continuous femoral nerve blocks decrease time to discharge-readiness following tricompartment total knee arthroplasty: a dual-center, randomized, triple-masked, placebo-controlled study. Anesthesiology 2008;108:703–13. [PubMed: 18362603]

7. Louis T. Nonparametric analysis of an accelerated failure time model. Biometrika 1981;68:381–90.