

This Week's Citation Classic

Gehan E A. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 52:203-13, 1965.

[Birkbeck College, London, and National Institutes of Health, Bethesda, MD]

A statistical test was proposed extending the Wilcoxon test to arbitrarily right-censored samples. The test is particularly applicable to comparing two survival distributions in a clinical trial when there is a mixture of individuals who have died and others who are still alive in each group. [The SC[®] indicates that this paper has been cited over 230 times since 1965.]

Edmund A. Gehan
University of Texas System
Cancer Center
M. D. Anderson Hospital
and Tumor Institute
Texas Medical Center
Houston, TX 77030

November 28, 1978

"This paper was written while on a special fellowship from the National Cancer Institute at Birkbeck College, University of London. During the previous four years (1958-1962), I worked as a consulting statistician at the National Cancer Institute and identified the problem of the paper. In col-laboration with Dr. E. J. Freireich et al. on a clinical trial comparing 6-MP to placebo in the maintenance of complete remissions in children with acute leukemia, there were 21 pairs of patients with 21 times to relapse on placebo, and 9 times to relapse and 12 times to censoring on 6-MP.¹ Times to censoring were lengths of remission in individuals who had not relapsed at time of analysis. Clearly, 6-MP was significantly superior to placebo, but there was no statistical test utilizing all the data for demonstrating this.

"The fellowship provided a wonderful opportunity to work on the problem and I was

most fortunate to work in collaboration with Professor David R. Cox who was then head of the small Department of Statistics of Birkbeck. In addition, I had married my wife, Brenda, shortly before leaving for London so it was indeed a 'honeymoon experience.' The test developed was a natural generalization of the Wilcoxon test for comparing two samples. Professor Cox was of great help on approaches to determining the theoretical characteristics of the test. He should have co-authored the paper, but was only willing to be acknowledged. Brenda helped by her encouragement and hand-calculating over 1,000 examples of the test which demonstrated that a normal approximation worked well in small samples.

"The paper has received many references in the statistical literature because it led to many additional theoretical developments for survival studies with censored data, most notably a paper by Cox on regression methods.² There are also many references in the medical literature as justification for the statistical test performed to compare survival times in two groups—some even using the term 'Gehan test,' though Wilcoxon developed the original test. It is still amazing that the problem was not considered much earlier, since it arises so commonly in clinical trials. John Gilbert of Harvard University recognized the problem at about the same time and developed an identical test in an unpublished Ph.D. thesis. I've always been thankful for using the example of acute leukemia (I seriously considered using hypothetical data with censored observations in both samples), since many subsequent authors have used the same data. What may not be known by many statistical authors is that the medical paper by Freireich et al. was a breakthrough—it was the first to demonstrate that maintaining patients on treatment in complete remission was beneficial."

1. Freireich E, Gehan E, Frei E, III, Schroeder L, Wolman I, Anbari R, Burgert E, Mills S, Pinkel D, Selawry O, Moon J, Gendel B, Spurr C, Storrs R, Haurani F, Hoogstraten B, and Lee S. The effect of 6 mercaptopurine on the duration of steroidinduced remissions in acute leukemia. *Blood* 21:699-716, 1963.

2. Cox D. Regression models and life tables. *J. Roy. Statist. Soc. B* 34:187-220, 1972.