


11-1-2005

AB/BA Crossover Trials - Binary Outcome

James F. Reed III

Lehigh Valley Hospital and Health Network, James_F.Reed@lvh.com

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Reed, James F. III (2005) "AB/BA Crossover Trials - Binary Outcome," *Journal of Modern Applied Statistical Methods*: Vol. 5: Iss. 2, Article 19.

Available at: <http://digitalcommons.wayne.edu/jmasm/vol5/iss2/19>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.

AB/BA Crossover Trials - Binary Outcome

James F. Reed III

Interim Chief of Health Studies and Director of Research
Lehigh Valley Hospital and Health Network

On occasion, the response to treatment in an AB/BA crossover trial is measured on a binary variable - success or failure. It is assumed that response to treatment is measured on an outcome variable with (+) representing a treatment success and a (-) representing a treatment failure. Traditionally, three tests for comparing treatment effect have been used (McNemar's, Mainland-Gart, and Prescott's). An issue arises concerning treatment comparisons when there may be a residual effect (carryover effect) of a previous treatment affecting the current treatment. A general consensus as to which procedure is preferable is debatable. However, if both group and carry-over effects are absent, Prescott's test is the best one to use. Under a model with residual effects, Prescott's test is biased. Therefore, a conservative approach includes testing for residual effects. When there is no period effect, McNemar's test is optimal, while McNemar's test is biased.

Key words: Crossover trial, AB/B McNemar's Test

Introduction

In a classic binary crossover trial, patients are randomly assigned to receive either treatment A (τ_A) in the first period followed by treatment B (τ_B) in a second period, or treatment B in the first period followed by treatment A in the second period. On occasion, the response to treatment is measured on a binary variable - success or failure. In a crossover design each subject acts as their own control, which is valuable in the presence of substantial between subject variability. The purpose of the crossover trial is to estimate the treatment contrast $\tau_A - \tau_B$. Period effects, π_1 and π_2 , as well as carryover effects, λ_A and λ_B , are regarded as nuisance

parameters. Traditionally, three tests for comparing treatment effect have been used (McNemar's, Mainland-Gart, and Prescott's). An issue arises concerning treatment comparisons when there may be a residual effect (carryover effect) of a previous treatment affecting the current treatment. Any residual effect, positive or negative, can bias the estimate of treatment effect. The best way of dealing with a potential residual effect is to design the study so that a suitable washout period is included so that the effect of the first period treatment does not affect the treatment applied to the second treatment period. If the researchers cannot safely insure this, then the crossover design is not an appropriate design.

Numerical Methods

For a binary crossover trial, assume that the response data may be arranged as shown in Table 1. Here the response to treatment is measured on an outcome variable with (+) representing a treatment success and a (-) representing a treatment failure.

James Reed III is the Interim Chief of Health Studies and Director of Research at Lehigh Valley Hospital and Health Network. He has published over 100 journal articles and book chapters. His interests include applied statistical analyses, medical education, and statistical methods in simulation studies. Email: James_F.Reed@lvh.com

Mainland - Gart $H_0: \{(\tau_B - \tau_A) = 0 \text{ and } (\lambda_A - \lambda_B) = 0\}$

Gart (1963) gave a rigorous description of two procedures which had been proposed earlier by Mainland (1963) that simultaneously considers the hypothesis $H_0: \{(\tau_B - \tau_A) = 0 \text{ and } (\lambda_A - \lambda_B) = 0\}$. The first, an exact test, is the product of two symmetric binomial variables, while the second is a goodness-of-fit χ^2 with 2 degrees of freedom for large n's (Table 2a).

$$X^2_{H_0} = [(n_{12} - m_{12})^2 / (n_{12} + m_{12})] + [(m_{21} - n_{21})^2 / (m_{21} + n_{21})]$$

Rejection of $X^2_{H_0}$ indicates that there is some deviation from the null hypothesis in either or both of the components of the two-sided hypothesis.

Gart, Fleiss $H_{0\lambda}: \lambda_A - \lambda_B = 0$

Gart (1963) described an exact test and a normal deviate test statistic for order or carry-over effect (δ) (e.g. $H_{0\lambda}: \lambda_A - \lambda_B = 0$). The exact test corresponds to Fisher's Exact test for a 2 x 2 table with fixed marginals given in Table 2b.

The normal deviate test statistic ($Z_{G\lambda}$) is defined as:

$$Z_{G\lambda} = \{|n_{12} - ((n_{12}+m_{12})(n_{12}+m_{21})/M) | - 1/2\} / \{[(n_{12}+m_{12})(n_{21}+m_{21})(n_{12}+n_{21})(M-n_{12}-m_{21})] / [M^2(M-1)]^{1/2}\}$$

Fleiss (1986) also described a statistic (X^2_λ) for testing the hypothesis of equal residual effects ($H_{0\lambda}: \lambda_A - \lambda_B = 0$). Let $P_1 = (2n_{11} + n_{12} + n_{21})/2n_{..}$, $P_2 = (2m_{11} + m_{12} + m_{21})/2m_{..}$, $V(P_1) = 1/4n^3_{..} (n_{..}[n_{12} + n_{22}] - [n_{11} - n_{22}]^2)$, and $V(P_2) = 1/4m^3_{..} (m_{..}[m_{12} + m_{22}] - [m_{11} - m_{22}]^2)$, then

$$X^2_\lambda \text{ is: } X^2_\lambda = (|P_1 - P_2| - 1/4 (1/n_{..} + 1/m_{..}))^2 / (V(P_1) + V(P_2)). X^2_\lambda$$

is compared to a χ^2 distribution with 1 degree of freedom.

Gart, Prescott, Fleiss $H_{0\tau}: \{\tau_B - \tau_A = 0\}$

The hypothesis of the equality of treatment effects (τ) between the two groups may be tested by one of several procedures. Gart proposed a one-tail exact test (Fisher's exact test applied to a 2 x 2 - Table 2a) and a normal deviate with a continuity correction when the sample size is large ($Z_{G\tau}$) (Gart, 1963).

$$Z_{G\tau} = \{|n_{12} - ((n_{12}+m_{12})(n_{12}+n_{21})/M) | - 1/2\} / \{[(n_{12}+m_{12})(n_{21}+m_{21})(n_{12}+n_{21})(M-n_{12}-n_{21})] / [M^2(M-1)]^{1/2}\}$$

Prescott (1981) derived a normal deviate test statistic (Z_P) for testing $H_{0\tau}: \{\tau_B - \tau_A = 0\}$ (Table 3). $Z_P = (|T - E(T)| - 1/2)/V(T)$, where: $T = n_{12} - m_{12}$, $E(T) = n_{.1} (n_{1.} - n_{3.})/n_{..}$, and $V(T) = n_{.1} n_{.2} \{ (n_{1.} - n_{3.}) - (n_{1.} - n_{3.})^2/n_{..} \} / \{ n_{..} (n_{..} - 1) \}$.

Table 1. Data layout for a two-period crossover with a binary response variable.

AB				BA			
π_2 Response				π_2 Response			
π_1 Response	+	-	Total	π_1 Response	+	-	Total
+	n_{11}	n_{12}	$n_{1.}$	+	m_{11}	m_{12}	$m_{1.}$
-	n_{21}	n_{22}	$n_{2.}$	-	m_{21}	m_{22}	$m_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$	Total	$m_{.1}$	$m_{.2}$	$m_{..}$

Table 2. Classification of binary outcomes by period preference (Mainland-Gart Test).

a) Test $H_{0\tau}: \{\tau_B - \tau_A = 0\}$

Sequence	Preference		Total
	Period 1 (π_1)	Period 2 (π_2)	
AB	n_{12}	n_{21}	$n_{12} + n_{21}$
BA	m_{12}	m_{21}	$m_{12} + m_{21}$
Total	$n_{12} + m_{12}$	$n_{21} + m_{21}$	M

b) Test $H_{0\delta}: \{\lambda_A - \lambda_B = 0\}$

Sequence	Preference		Total
	Period 1 (π_1)	Period 2 (π_2)	
AB	n_{12}	m_{21}	$n_{12} + m_{21}$
BA	m_{12}	n_{21}	$m_{12} + n_{21}$
Total	$n_{12} + m_{12}$	$n_{21} + m_{21}$	M

Fleiss (1986) describes a statistic ($Z_{F\tau}$) that tests the hypothesis of equal treatment effects ($H_{0\lambda}: \lambda_A - \lambda_B = 0$). Let $p'_A = n_{12}/n'$, $n' = n_{12} + n_{21}$, $p'_B = m_{12}/m'$, $m' = m_{12} + m_{21}$, $p' = (n_{12} + m_{12})/(n' + m')$, then $Z_{F\tau} = \{[|p'_A - p'_B| - \frac{1}{2}(1/n' + 1/m')]/\sqrt{p'(1-p')}\}/\sqrt{(n'm')/(n'+m')}$. The hypothesis of equal treatment effects is rejected if $Z_{F\tau} > Z_{\alpha/2}$.

Fleiss $H_{0\tau|\lambda}$

If there were a significant residual effect, the general advice is that the appropriate test for testing $H_{0\tau|\lambda}$ is to use data from the first treatment period. In this case $H_{0\tau|\lambda}$ may be tested using the following test statistic ($Z_{\tau|\lambda}$) described by Fleiss (1986). $Z_{\tau|\lambda} = \{|p_A - p_B| - [1/n_{..} + 1/m_{..}]/2\} / \{p'(1-p)^{1/2}\}(n_{..}m_{..}/(n_{..}+m_{..}))^{1/2}$, with $p_A = n_{1.}/n_{..}$, $p_B = m_{1.}/m_{..}$, and $p' = (n_{1.} + m_{1.})/(n_{..} + m_{..})$. $H_{0\tau|\lambda}$ is rejected if $Z_{\tau|\lambda} > Z_{\alpha/2}$.

Example

Consider the data given by Fleiss (1986, pg 277).

AB				BA			
π_2 Response				π_2 Response			
π_1 Response	+	-	Total	π_1 Response	+	-	Total
+	25	15	40	+	30	0	30
-	5	5	10	-	5	15	20
Total	30	20	50	Total	35	15	50

The test statistics are:

$H_0: \{\tau_A = \tau_B, \lambda_A = \lambda_B\}$
Mainland-Gart chi-square = 6.00, p-value = 0.04841

$H_0: \{\lambda_A = \lambda_B\}$
Mainland-Gart E (1-tail), p-value = 0.29181
Mainland-Gart Z = 0.61, p-value = 0.54029
Fleiss $X^2 = 0.26$, p-value = 0.61914

$H_0: \{\tau_A = \tau_B\}$
Mainland-Gart E (1-tail), p-value = 0.00474
Mainland-Gart Z = 2.50, p-value = 0.00621
Prescott Z = 2.92, p-value = 0.00349
Fleiss Z = 2.55, p-value = 0.01072

$H_0: \{\tau_{A|\lambda} = \tau_{B|\lambda}\}$
Fleiss-T|Residual = 1.96, p-value = 0.04953

The simultaneous test of $H_0: \{\tau_A = \tau_B, \lambda_A = \lambda_B\}$ indicates that there is a difference in treatment effects or that there is a significant residual effect (or both). All of the test statistics and associated p-values associated with $H_0: \{\lambda_A = \lambda_B\}$ are not statistically significant at any reasonable level. Further, the hypothesis $H_0: \{\tau_A = \tau_B\}$ is also significant (0.01072).

Discussion

Traditionally, three tests for comparing treatment effect have been used to analyze data from a binary two period crossover trial (McNemar, Mainland-Gart, and Prescott). McNemar’s test in effect ignores information on periods and tests for difference between expected values of $n_{21} + m_{12}$ and $n_{12} + m_{21}$. The Mainland-Gart test is Fisher’s exact test applied to a 2 x 2 table (Table 2). Prescott’s test expands the Mainland-Gart test by using tied responses and is a test for linear trend in a 2 X 3 table (Table 3).

Other models have been proposed to analyze binary data from a two period crossover trial. For instance, Becker and Balagtas (1993) proposed a model in terms of linear models for marginal logits and linear models for log-odds ratios for the analysis of two-period binary crossover trials. Their simulation results demonstrated that their likelihood ratio approach compared favorably with standard procedures such as the Mainland-Gart test for a treatment difference, Prescott’s test for a treatment

difference and the Hills-Armitage test for treatment by period interaction.

Which, if any, of the procedures is best has been debated for nearly 40 years and the debate will probably continue. A general consensus is that if both group and carry-over effects are absent, Prescott’s test is the best one to use. Under the model with residual effects, Prescott’s test is biased, as it can give p-values that are too high or too low (Fidler, 1984). Estimates of treatment effects are biased when residual effects are present. With positive residual effects, the treatment effect is underestimated and the crossover test is conservative; with negative residual effects, the treatment effect is overestimated and the crossover test is liberal (Lehmacher, 1991). Unlike the group effect, the residual effect cannot be eliminated by randomization alone. Therefore, a conservative approach includes testing for residual effects. A non-significant result from the residual effect and external evidence for the absence of a residual effect is essential when using Prescott’s test. When there is no period effect, McNemar’s test is optimal, while McNemar’s test is biased (Fidler, 1984). When a crossover trial has been properly designed, it remains a powerful instrument in the analysts’ toolbox and is a reasonable and sensitive means of determining the efficacy of two treatment regimens (Cleophas & Val Lier, 1996).

Table 3. Classification of binary outcomes by period preference, including tied responses (Prescott’s Test).

Test $H_{0\tau}: \{\tau_B - \tau_A = 0\}$

Sequence	Preference			Total
	Period 1 (π_1)	Tied	Period 2 (π_2)	
AB	n_{12}	$n_{11} + n_{22}$	n_{21}	$n_{1.}$
BA	m_{12}	$m_{11} + m_{22}$	m_{21}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$

Program

An executable program that produces the statistics outlined in this article is available from the author on request. The input file is in free format (sequence AB=1, BA=2; observation 1, 2, n, 1, 2, m; period 1 outcome (1 = success, 0 = failure); period 2 outcome (1 = success, 0 = failure)). Data must be in an integer format.

References

Becker, M. P. & Balagtas, C. C. (1993). Marginal modeling of binary cross-over data. *Biometrics*, 49, 997-1009.

Cleophas, T. J. & van Lier, H. J. (1996). Crossover trials with a binary response: A powerful method despite the carryover effect. *Journal of Clinical Pharmacology*, 36, 198-202.

Fidler, V. (1984). Change-over clinical trial with binary data: Mixed-model-based comparison of tests. *Biometrics*, 40, 1063-1070.

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York, N.Y.: John Wiley & Sons.

Gart, J. J. (1996). An exact test for comparing matched proportions in crossover designs. *Biometrika*, 56, 75-80.

Lehmacher, W. (1991). Analysis of the crossover design in the presence of residual effects. *Statistics in Medicine*, 10, 891-899.

Mainland, D. (1963). *Elementary medical statistics* (2nd ed.). Philadelphia, P.A.: Saunders.

Prescott, R. J. (1981). The comparison of success rates in cross-over trials in the presence of an order effect. *Applied Statistics*, 30, 9-15.