Statistica Sinica **15**(2005), 857-870

LAST OBSERVATION ANALYSIS IN ANOVA AND ANCOVA

Bin Cheng, Jun Shao and Bob Zhong

Columbia University, University of Wisconsin and Centocor, Inc.

Abstract: In clinical trials with multiple visits, dropouts often occur and the population of patients who dropped out may be different from the population of patients who completed the study. To assess treatment effects over the population of all randomized patients, which is called the intention-to-treat analysis and is required by regulatory agencies, last observation analysis (LOAN) focuses on the last observation of each patient prior to dropout. As a type of LOAN, the last observation carry-forward (LOCF) method treats the last observation prior to dropout as the missing observation at the end of the trial and applies standard tests designed for the case of no dropout. Regulatory agencies such as the U.S. Food and Drug Administration (FDA) have expressed concerns about the validity of the LOCF methods. In this paper we study the validity of LOAN and LOCF tests under an analysis of covariance model, which includes the analysis of variance model as a special case. In situations where LOAN is relevant, we provide explicit conditions under which LOCF tests are asymptotically valid and we derive asymptotically valid tests when LOCF tests are invalid.

Key words and phrases: Dropout, intention-to-treat analysis, last observation carry-forward, balanced design, balanced covariates.

1. Introduction

In clinical trials, data are often collected over multiple visits of participating patients, and statistical analyses focus on observations at the end of the study or change-of-efficacy measurements from baseline to the end of the trial. Despite a thoughtful and well-designed study protocol, it is frequently the case in a clinical trial that patients drop out prior to the end of the study. In the presence of dropout, many regulatory agencies require the intention-to-treat analysis that focuses on all randomized patients with at least one post-treatment evaluation.

An approach focusing on the last observed visit has received some attention recently, the so-called last observation analysis (LOAN). Let μ_{it} be the ith treatment population mean of the last response y (the primary variable of interest) of a patient who dropped out after visit t, where $i = 1, \ldots, I$, $t = 1, \ldots, T$, and visit T is the end of the study so that μ_{iT} is the mean of completers. (Note that

μ_{it} is typically different from the ith treatment population mean of y at visit t in the case of no dropout.) The LOAN evaluates treatment effects by comparing the μ_{it}, $t = 1, \ldots, T$, $i = 1, \ldots, I$. The oldest LOAN is called last observation carry forward (LOCF), an imputation method that imputes the missing response at the end of the study by using the response at the visit prior to dropout. Once this carry forward imputation is done, one applies standard statistical tests that treat all observations (imputed or not) as responses at the end of the study (Ting (2000)). Although the LOCF has a long history of application, there has been concern that treating carried-forward data as observed data creates biases in statistical tests for treatment effects (Heyting, Tolboom and Essers (1992), Lavori (1992), Dawson (1994a) and Ting (2000)). A second type of LOAN defines the overall treatment effect as a weighted average of $D_1, \ldots, D_T$, where, for each t, D_t is a measure assessing the difference among $\mu_{it}, \ldots, \mu_{It}$ (see, for example, Dawson and Lagakos (1993), Dawson (1994) and Shih and Quan (1998)). A key assumption here is that the missing patterns among different treatments are almost the same (i.e., $p_{1t} = \cdots = p_{It}$ for each t, where p_{it} is the population proportion of patients dropping out after visit t under treatment i), but this is not realistic in many applications. The most recent LOAN proposed in Shao and Zhong (2003) assess treatment effects by comparing the weighted averages $\mu_i = \sum_t p_{it}\mu_{it}$, $i = 1, \ldots, I$, which are unbiasedly estimated by the sample means based on LOCF data.

Note that the interpretation of treatment effects in the LOAN is very different from that in the approach of comparing treatment effects at the end of the study (in the presence of dropout). In some practical applications, using μ_i's in the comparison of treatment effects makes sense (e.g., dropout is caused by death), whereas in some situations comparing effects at the end of the study is more reasonable. When dropout is present and depends on y (observed and unobserved), however, treatment effects at the end of the study (even if they are not hypothetical) may not be estimable unless a strong (typically nonverifiable) assumption is imposed on the dropout mechanism and/or the y-population. For example, the treatment effects at the end of the study may be confounded with other effects when patients switch to other medications after dropout. Hence, analysis of the μ_i's may be used if there is no other more reasonable approach.

The purpose of this paper is to study methods for inference on $\mu_1, \ldots, \mu_I$, assuming that the μ_i's can be used to interpret treatment effects. We adopt the pattern-mixture approach (Little (1993)), which requires very few assumptions about the dropout mechanism.

Although the sample means based on LOCF data estimate the μ_i's, the LOCF tests may not be correct. There is a belief that the size of a LOCF test may be substantially higher or lower than the nominal size α unless $\mu_{i1}, \ldots, \mu_{iT}$

are the same. However, Shao and Zhong (2003) showed that, in one-way ANOVA with two treatments and a balanced design (i.e., the designed sample sizes of treatment groups are the same), the asymptotic size of the LOCF test is still the nominal size α when the null hypothesis is the equality of μ_i's, regardless of whether $\mu_{i1}, \ldots, \mu_{iT}$ are the same or not. It is also shown in Shao and Zhong (2003) that the asymptotic size of the LOCF test is not α when the design is not balanced or more than two treatments are compared. An explanation for these results is that, based on LOCF data, the mean sum of squares for treatment (MSTR) in the one-way ANOVA table is asymptotically distributed as a weighted average of $I-1$ independent chi-square random variables, where I is the number of treatments; when $I=2$, the MSTR is asymptotically distributed as a scaled chi-square random variable; when the design is balanced, this scale is exactly the same as the limit of the mean sum of squares for error (MSE), which ensures the asymptotic validity of the LOCF test; if either $I \geq 3$ or the design is not balanced, the ratio MSTR/MSE is asymptotically distributed as a weighted average of chi-square random variables (not a chi-square random variable) and, thus, the size of the LOCF test is wrong.

Since a LOCF test is often used in clinical trials, it is important to know when it is (asymptotically) valid and, in the case where the LOCF test is not valid, what is a valid testing procedure. The result in Shao and Zhong (2003) only applies to one-way ANOVA. In clinical trials, an analysis of covariance (ANCOVA) is often used to incorporate covariates such as the baseline observations. Also, many clinical trials are multicenter trials, which leads to a two-way (or K-way; $K \geq 3$) ANOVA or ANCOVA.

For a better understanding of the problem, we start with the one-way ANCOVA in Section 2. Our result shows that in order to have an asymptotically valid LOCF test, not only the designed sample sizes of the two treatment groups need to be the same, but also the covariates in the model need to satisfy a balance condition. More precisely, the covariates in two different treatment groups need to have either the same average or the same variability. Furthermore, we derive a test that is always asymptotically valid and can be used to replace the LOCF test when it is asymptotically invalid. In Section 3, we focus on testing the interaction effect in a two-way ANOVA model (without covariates). Our results show that the LOCF test is valid only in some very special situations. A similar conclusion can be drawn for a two-way ANCOVA model. An asymptotically valid test for interaction under a two-way ANCOVA model is derived. Based on the results in Section 3, we consider tests for the main effect (treatment effect) in a two-way additive ANCOVA model in Section 4. The result in one-way ANCOVA is extended to this model. Extensions of our results to general K-way ANCOVA are straightforward. Finally, some simulation results are presented in Section 5.

2. One-Way ANCOVA

Consider a clinical trial consisting of I treatments, n_i patients randomized to treatment group i, and T scheduled post-baseline visits for each patient. Suppose that under treatment i, n_{it} patients drop out after visit t. Hence, $n_{i1}+\cdots+n_{iT} = n_i$ and $(n_{i1},\ldots,n_{iT})$ has the multinomial$(n_i, p_{i1},\ldots,p_{iT})$ distribution, where p_{it} is the population proportion of patients dropping out after visit t under treatment i. We assume a pattern-mixture one-way ANCOVA model, i.e., for patients who drop out after visit t, their last observed responses y_{itk}'s are independent with means $\mu_{it} + \mathbf{b}'\mathbf{z}_{itk}$, where $k = 1,\ldots,n_{it}$, $t = 1,\ldots,T$, $i = 1,\ldots,I$, the μ_{it}'s are unknown fixed treatment effects, $\mathbf{z}_{itk}$ is a q-vector of covariates observed for each patient, and $\mathbf{b}$ is a q-vector of unknown parameters. Unlike the y-response variable, the $\mathbf{z}$-covariate for a patient does not vary with t, although we use the notation $\mathbf{z}_{itk}$ to specify it. No other condition is imposed on the dropout mechanism (i.e., dropout may be nonignorable).

When there is no dropout, testing for treatment effect may be carried out by using responses from the end of the study and the method of ANCOVA. When there are dropouts, as we discussed in Section 1, the LOAN considers the hypothesis

$$H_0 : \mu_1 = \cdots = \mu_I, \tag{1}$$

where $\mu_i = p_{i1}\mu_{i1} + \cdots + p_{iT}\mu_{iT}$.

The LOCF test is the ANCOVA test that treats y_{itk} as the observation at the end of the trial. Since μ_{it} usually changes with t, one wonders what the LOCF tests for. If (1) is the hypothesis of interest, there is the question of validity of the LOCF test. The following result shows when the LOCF test is asymptotically valid for (1). Since Shao and Zhong (2003) showed that when $I \geq 3$, the LOCF test is asymptotically wrong for testing (1) in a one-way ANOVA without covariates, we only consider the case of $I = 2$ treatments.

In this paper, χ^2_d denotes the chi-square distribution with d degree of freedom, while $\chi^2_{d,\alpha}$ and $F_{l,m,\alpha}$ denote, respectively, the $1-\alpha$ quantiles of χ^2_d and the F-distribution with degrees of freedom l and m, where α is a given nominal level. Let

$$\text{MSTR} = \Big[(\bar{y}_{1..} - \hat{\mathbf{b}}'\bar{\mathbf{z}}_{1..}) - (\bar{y}_{2..} - \hat{\mathbf{b}}'\bar{\mathbf{z}}_{2..})\Big]^2 \Big/ \sum_{i=1}^{2}\sum_{t=1}^{T}\sum_{k=1}^{n_{it}} a_{itk}^2,$$

$$\text{MSE} = \frac{1}{n_1+n_2-q-2}\sum_{i=1}^{2}\sum_{t=1}^{T}\sum_{k=1}^{n_{it}}\Big[(y_{itk} - \bar{y}_{i..}) - \hat{\mathbf{b}}'(\mathbf{z}_{itk} - \bar{\mathbf{z}}_{i..})\Big]^2,$$

where $\mathbf{z}_{itk}$ is the covariate value associated with y_{itk}, $\bar{y}_{i..}$ and $\bar{\mathbf{z}}_{i..}$ are, respectively, the averages of y_{itk} and $\mathbf{z}_{itk}$ over the indexes t and k,

$$\hat{\mathbf{b}} = (\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\sum_{i=1}^{2}\sum_{t=1}^{T}\sum_{k=1}^{n_{it}}(y_{itk} - \bar{y}_{i..})(\mathbf{z}_{itk} - \bar{\mathbf{z}}_{i..}),$$

$\tilde{\mathbf{Z}}$ is the $(n_1+n_2)\times q$ matrix whose first n_1 rows are $\mathbf{z}'_{111}-\bar{\mathbf{z}}'_{1..},\ldots,\mathbf{z}'_{11n_{11}}-\bar{\mathbf{z}}'_{1..},\ldots,\mathbf{z}'_{1Tn_{1T}}-\bar{\mathbf{z}}'_{1..}$ and whose last n_2 rows are $\mathbf{z}'_{211}-\bar{\mathbf{z}}'_{2..},\ldots,\mathbf{z}'_{21n_{21}}-\bar{\mathbf{z}}'_{2..},\ldots,$ $\mathbf{z}'_{2Tn_{2T}}-\bar{\mathbf{z}}'_{2..}$, and $a_{itk}=n_i^{-1}+(-1)^i(\bar{\mathbf{z}}_{1..}-\bar{\mathbf{z}}_{2..})'(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}(\mathbf{z}_{itk}-\bar{\mathbf{z}}_{i..})$.

Theorem 1. *Assume $I=2$ and, for patients dropping out after visit t, the y_{itk}'s are independent with means $\mu_{it}+\mathbf{b}'\mathbf{z}_{itk}$ and variance $\sigma^2>0$.*

(i) *The ANCOVA test based on LOCF rejects* (1) *when the ratio F=MSTR/MSE is larger than $F_{1,n_1+n_2-q-2,\alpha}$.*

(ii) *As $n_i\to\infty$, $i=1,2$, MSE $\to_p\ \sigma^2+\eta$, where $\to_p$ denotes convergence in probability, with*

$$\eta=\lim\frac{n_1\tau_1^2+n_2\tau_2^2}{n_1+n_2} \tag{2}$$

and $\tau_i^2=\sum_{t=1}^T p_{it}(\mu_{it}-\mu_i)^2$.

(iii) *Under* (1), *as $n_i\to\infty$, $i=1,2$, MSTR $\to_d\ (\sigma^2+\zeta)\chi_1^2$, where $\to_d$ denotes convergence in distribution, and*

$$\zeta=\lim\frac{w_1\tau_1^2+w_2\tau_2^2}{w_1+w_2}, \tag{3}$$

$$w_i=\sum_{t=1}^T\sum_{k=1}^{n_{it}}a_{itk}^2=\frac{1}{n_i}+(\bar{\mathbf{z}}_{1..}-\bar{\mathbf{z}}_{2..})'(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\mathbf{S}_i(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}(\bar{\mathbf{z}}_{1..}-\bar{\mathbf{z}}_{2..}), \tag{4}$$

with $\mathbf{S}_i=\sum_{t=1}^T\sum_{k=1}^{n_{it}}(\mathbf{z}_{itk}-\bar{\mathbf{z}}_{i..})(\mathbf{z}_{itk}-\bar{\mathbf{z}}_{i..})'$.

The proof of Theorem 1(i) is based on the formula (B) in Searle (1987, p.425). It is straightforward and therefore omitted. The proofs for Theorem 1(ii) and (iii) are given in the Appendix.

It follows from Theorem 1 that the LOCF test is asymptotically valid for testing (1) if and only if $\eta=\zeta$. Examining (2) and (3), we find that the LOCF test is asymptotically valid if either τ_1^2 and τ_2^2 are asymptotically the same, or $n_1/(n_1+n_2)$ and $w_1/(w_1+w_2)$ are asymptotically the same. The only practical situation in which τ_1^2 and τ_2^2 are asymptotically the same is when $\tau_1^2=\tau_2^2=0$, which corresponds to the case of $\mu_{it}=\mu_{iT}$ for all t. Hence, if μ_{it}'s for a given i are different, the asymptotic validity of the LOCF test depends on the condition

$$\lim\frac{n_1}{n_1+n_2}=\lim\frac{w_1}{w_1+w_2}. \tag{5}$$

We find that two practical situations in which (5) holds are

$$\lim\frac{n_1}{n_2}=1\ \text{ and }\quad \lim(\bar{\mathbf{z}}_{1..}-\bar{\mathbf{z}}_{2..})=0, \tag{6}$$

$$\lim\frac{n_1}{n_2}=1\ \text{ and }\quad \lim\left(\frac{\mathbf{S}_1}{n_1}-\frac{\mathbf{S}_2}{n_2}\right)=0. \tag{7}$$

That is, the LOCF test is asymptotically valid for testing (1) when (6) or (7) holds. Note that Shao and Zhong (2003) showed that the condition $\lim(n_1/n_2) = 1$ ensures the asymptotic validity of the LOCF test in one-way ANOVA. When there are covariates, our result shows that in addition to the balance condition $\lim n_1/n_2 = 1$, the validity of the LOCF test requires the covariates to be balanced in the sense that either the means of the covariates under two treatments are asymptotically the same (condition (6)) or the covariance matrices of the covariates under two treatments are asymptotically the same (condition (7)).

When there are $I \geq 3$ treatments or the design is not balanced, the LOCF test has the wrong asymptotic size for testing (1). An asymptotically valid test of (1) is derived as follows. For the ith treatment, let $\hat{\mathbf{b}}_i = (\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i)^{-1}\sum_{t=1}^{T}\sum_{k=1}^{n_{it}}\mathbf{z}_{itk}y_{itk}$, where $\tilde{\mathbf{Z}}_i$ is the $n_i \times q$ matrix whose n_i rows are $\mathbf{z}_{i11}' - \bar{\mathbf{z}}_{i..}', \ldots, \mathbf{z}_{i1n_{i1}}' - \bar{\mathbf{z}}_{i..}', \ldots, \mathbf{z}_{iTn_{iT}}' - \bar{\mathbf{z}}_{i..}'$, and let $u_{itk} = y_{itk} - \hat{\mathbf{b}}_i'\mathbf{z}_{itk}$. Then $\bar{u}_{i..} = n_i^{-1}\sum_{t=1}^{T}\sum_{k=1}^{n_{it}} u_{itk}$ is unbiased for μ_i and asymptotically normal, and its variance can be estimated consistently by

$$\hat{V}_i = \frac{1}{n_i(n_i-1)}\sum_{t=1}^{T}\sum_{k=1}^{n_{it}}(u_{itk} - \bar{u}_{i..})^2.$$

Theorem 2. *Suppose that, for patients dropping out after visit t, the y_{itk}'s are independent with means $\mu_{it} + \mathbf{b}'\mathbf{z}_{itk}$ and variances $\sigma_{it}^2 > 0$. Under (1), as $n_i \to \infty$ for all i, $W \to_d \chi_{I-1}^2$, where*

$$W = \sum_{i=1}^{I}\frac{1}{\hat{V}_i}\left(\bar{u}_{i..} - \frac{\sum_{i=1}^{I}\bar{u}_{i..}/\hat{V}_i}{\sum_{i=1}^{I}1/\hat{V}_i}\right)^2.$$

Consequently, an asymptotic size α test rejects (1) if and only if $W > \chi_{I-1,\alpha}^2$. Note that we do not assume the variances of y_{itk}'s are equal in Theorem 2.

When (1) is rejected, we can make inference (such as pairwise or multiple comparison on μ_i's) using the asymptotic results based on $\bar{u}_{i..}$ and $\hat{V}_i$.

3. Tests for Interaction in Two-Way Models

Two-way ANOVA or ANCOVA is often used in clinical trials. In addition to the treatment effect, a common factor in a two way ANOVA or ANCOVA is the center effect in a multicenter trial. Consider a clinical trial carried out in J centers with I treatments, n_{ij} patients randomized to treatment group i at center j, and T scheduled visits for each patient. The total number of patients is $n = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}$. Suppose that under treatment i at center j, n_{ijt} patients drop out after visit t. Then $(n_{ij1}, \ldots, n_{ijT})$ has the multinomial$(n_{ij}, p_{ij1}, ..., p_{ijT})$ distribution, where p_{ijt} is the population proportion of patients dropping out after visit t under treatment i at center j. Let y_{ijtk} be the last observed response

variable of interest from patient k under treatment i at center j who dropped out after the tth visit.

Under the two-way ANOVA model, for patients dropping out after visit t, we assume that y_{ijtk}'s are independent with means μ_{ijt}. Similar to the one-way case, we use the

$$\mu_{ij} = \sum_{t=1}^{T} p_{ijt}\mu_{ijt} \tag{8}$$

as measures for treatment and center effects. Under two-way models, the μ_i and μ_{it} of previous sections should be replaced by μ_{ij} and μ_{ijt}, respectively. Consider the decomposition

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij},$$

where μ is an overall mean, α_i's are fixed treatment effects ($\alpha_1 + \cdots + \alpha_I = 0$), β_j's are fixed center effects ($\beta_1 + \cdots + \beta_J = 0$), and γ_{ij}'s are fixed interaction effects ($\gamma_{i1} + \cdots + \gamma_{iJ} = \gamma_{1j} + \cdots + \gamma_{Ij} = 0$ for any i and j). Although the treatment effects α_i's are of primary interest, the analysis in two-way ANOVA often starts with a test for the treatment-by-center interaction with the null hypothesis

$$H_0 : \gamma_{ij} = 0, \quad \text{for all } i \text{ and } j. \tag{9}$$

To test (9), the LOCF test treats y_{ijtk} as the observation in the end of the trial and rejects H_0 when MSAB/MSE $> F_{(I-1)(J-1),n-IJ,\alpha}$, where

$$\text{MSAB} = \frac{1}{(I-1)(J-1)}\bar{\mathbf{y}}'\mathbf{L}(\mathbf{L}'\mathbf{\Lambda}\mathbf{L})^{-1}\mathbf{L}'\bar{\mathbf{y}},$$

$$\text{MSE} = \frac{1}{n-IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}}(y_{ijtk} - \bar{y}_{ij..})^2,$$

$\bar{y}_{ij..}$ is the average of y_{ijtk}'s over t and k, $\bar{\mathbf{y}} = (\bar{y}_{11..}, \ldots, \bar{y}_{I1..}, \ldots, \bar{y}_{1J..}, \ldots, \bar{y}_{IJ..})'$, $\mathbf{\Lambda} = \text{diag}(n_{11}^{-1}, \ldots, n_{I1}^{-1}, ..., n_{1J}^{-1}, \ldots, n_{IJ}^{-1})$,

$$\mathbf{L} = \begin{pmatrix} \mathbf{1}'_{(J-1)} \\ -\mathbf{I}_{(J-1)} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{1}'_{(I-1)} \\ -\mathbf{I}_{(I-1)} \end{pmatrix},$$

$\mathbf{1}_m$ is the m-vector of ones, $\mathbf{I}_m$ is the identity matrix of order m, and $\otimes$ is the Kronecker product. The following result shows what this tests for, and when it is asymptotically valid. The proof can be found in Cheng (2004).

Theorem 3. *Assume that, for patients dropping out after visit t, y_{ijtk}'s are independent with means μ_{ijt} and variance $\sigma^2 > 0$.*

(i) *As* $n_{ij} \to \infty$ *for all* i, j, $\text{MSE} \to_p \sigma^2 + \eta$, *where*

$$\eta = \lim \frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \tau_{ij}^2 \quad \textit{and} \quad \tau_{ij}^2 = \sum_{t=1}^{T} p_{ijt} (\mu_{ijt} - \mu_{ij})^2.$$

(ii) *Under* (9), *MSAB converges in distribution to a linear combination of* $(I-1)(J-1)$ *independent chi-square random variables with 1 degree of freedom. The LOCF test for interaction is asymptotically valid* (*i.e., MSAB/MSE is asymptotically distributed as* $\chi^2_{(I-1)(J-1)}$) *if and only if*

$$\mathbf{L}'\mathbf{V}\mathbf{L} = \eta \mathbf{L}'\mathbf{\Lambda}\mathbf{L}, \tag{10}$$

where $\mathbf{V} = \text{diag}(n_{11}^{-1}\tau_{11}^2, \ldots, n_{I1}^{-1}\tau_{I1}^2, \ldots, n_{1J}^{-1}\tau_{1J}^2, \ldots, n_{IJ}^{-1}\tau_{IJ}^2)$.

(iii) *When* $I = J = 2$, (10) *becomes*

$$\lim \frac{\sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} \tau_{ij}^2}{\sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij}} = \lim \frac{\sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij}^{-1} \tau_{ij}^2}{\sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij}^{-1}}. \tag{11}$$

When $I = 2$ *and* $J \geq 3$, (10) *becomes*

$$\frac{n_{1j}^{-1}\tau_{1j}^2 + n_{2j}^{-1}\tau_{2j}^2}{n_{1j}^{-1} + n_{2j}^{-1}} = \frac{\sum_{i=1}^{2} \sum_{j=1}^{J} n_{ij} \tau_{ij}^2}{\sum_{i=1}^{2} \sum_{j=1}^{J} n_{ij}}, \quad \textit{for all } j. \tag{12}$$

When $I \geq 3$ *and* $J = 2$, (10) *becomes*

$$\frac{n_{i1}^{-1}\tau_{i1}^2 + n_{i2}^{-1}\tau_{i2}^2}{n_{i1}^{-1} + n_{i2}^{-1}} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{2} n_{ij} \tau_{ij}^2}{\sum_{i=1}^{I} \sum_{j=1}^{2} n_{ij}}, \quad \textit{for all } i. \tag{13}$$

When $I \geq 3$ *and* $J \geq 3$, (10) *becomes*

$$\tau_{ij}^2 = \text{constant}. \tag{14}$$

When $I = J = 2$ and μ_{ijt}'s are different for given i and j (so that the τ_{ij}'s are different), (11) implies that the LOCF test for treatment-by-center interaction is asymptotically valid for testing (9) if the design is balanced in the sense that $\lim n_{ij}/n = 1/4$ for any i and j. Although in many applications the number of treatments $I = 2$, the number of centers J is often more than 2. From (12) through (14), we know that when either I or J is more than 2, the only practical situation that a LOCF procedure is valid is when the τ_{ij}^2 are all the same or, equivalently, the μ_{ijt}'s are the same for fixed i and j. A result similar to Theorem 3 for a two-way ANCOVA model can be derived, but it is omitted since a necessary condition for the validity of the LOCF test is $I = J = 2$, which is a limited special case.

An asymptotically valid test for (9) can be derived based on cell mean estimators $\bar{y}_{ij..}$ and their consistent variance estimators. We now consider the general two-way ANCOVA model (which includes the ANOVA model as a special case) in which the mean of y_{ijtk} (for patients dropping out after visit t) is

$$\mu_{ijt} + \mathbf{b}'\mathbf{z}_{ijtk}, \tag{15}$$

where $\mathbf{z}_{ijtk}$ is a q-vector of covariates observed for each patient, and $\mathbf{b}$ is a q-vector of unknown parameters. For any i and j, let $\hat{\mathbf{b}}_{ij}$ be the least squares estimator of $\mathbf{b}$ based on the data from the patients who received the ith treatment in the jth center and let $u_{ijtk} = y_{ijtk} - \hat{\mathbf{b}}_{ij}'\mathbf{z}_{ijtk}$. Then $\bar{u}_{ij..} = n_{ij}^{-1}\sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}} u_{ijtk}$ is an unbiased and asymptotically normal estimator of μ_{ij} with a consistent variance estimator

$$\hat{V}_{ij} = \frac{1}{n_{ij}(n_{ij}-1)}\sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}}(u_{ijtk} - \bar{u}_{ij..})^2.$$

Theorem 4. *Suppose that, for patients dropping out after visit t, the y_{ijtk}'s are independent with means given by* (15) *and variances σ_{ijt}^2. Under* (9), *as $n_{ij} \to \infty$ for all i, j, $W \to_d \chi^2_{(I-1)(J-1)}$, where $W = \bar{\mathbf{u}}'\mathbf{L}(\mathbf{L}'\hat{\mathbf{V}}\mathbf{L})^{-1}\mathbf{L}'\bar{\mathbf{u}}$, $\hat{\mathbf{V}} = \mathrm{diag}(\hat{V}_{11}, \ldots, \hat{V}_{I1}, \ldots, \hat{V}_{1J}, \ldots, \hat{V}_{IJ})$, and $\bar{\mathbf{u}} = (\bar{u}_{11..}, \ldots, \bar{u}_{I1..}, \ldots, \bar{u}_{1J..}, \ldots, \bar{u}_{IJ..})'$.*

Consequently, a test of (9) with asymptotic size α rejects H_0 if $W > \chi^2_{(I-1)(J-1),\alpha}$. It is clear that the result in Theorem 4 can be extended to K-way ANCOVA models.

4. Additive Two-Way ANCOVA

In a multicenter clinical trial, treatment-by-center interaction can sometimes be ignored, especially when covariates related to centers are introduced into the model. Hence, in this section we consider an additive two-way ANCOVA model, which includes the additive two-way ANOVA model as a special case.

Consider the multicenter clinical trial described in Section 3, where the mean of y_{ijtk} for patients dropping out after visit t is $\mu_{ijt} + \mathbf{b}'\mathbf{z}_{ijtk}$, $\mathbf{z}_{ijtk}$ is a q-vector of covariates observed from every patient, and $\mathbf{b}$ is a q-vector of unknown parameters. Let μ_{ij} be given by (8) and assume the additive model

$$\mu_{ij} = \mu + \alpha_i + \beta_j, \tag{16}$$

where μ is an overall mean, α_i's are fixed treatment effects ($\alpha_1 + \cdots + \alpha_I = 0$), and β_j's are fixed center effects ($\beta_1 + \cdots + \beta_J = 0$). The null hypothesis of no treatment effect is

$$H_0 : \alpha_1 = \cdots = \alpha_I = 0. \tag{17}$$

The LOCF procedure for testing (17) treats μ_{ijt} as μ_{ijT} for all t and uses the following model $E(\mathbf{Y}) = \mathbf{X}\theta + \mathbf{Zb}$, where $\theta = (\mu, \alpha_1, \ldots, \alpha_I, \beta_1, \ldots, \beta_J)'$, $\mathbf{Y}$ is the column vector formed by listing the elements y_{ijtk} in the order of i, j, t and k, $\mathbf{Z}$ is the matrix formed by listing the row vectors $\mathbf{z}_{ijtk}$ in the order of i, j, t and k, and $\mathbf{X}$ is the usual design matrix in a two-way additive ANOVA model. Define $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ and

$$\hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{PZ})^{-\mathbf{1}}\mathbf{Z}'\mathbf{PY}. \tag{18}$$

The following theorem shows when the LOCF test is asymptotically valid for testing (17). Its proof is similar to that of Theorem 1 and is omitted. First, let

$$\mathrm{MSTR} = \Big(\sum_{i=1}^{2}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}} a_{ijtk}y_{ijtk}\Big)^2 \Big/ \sum_{i=1}^{2}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}} a_{ijtk}^2,$$

where the a_{ijtk}'s are the components of the vector that is the difference between the second and the third rows of $(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{PZ})^{-\mathbf{1}}\mathbf{Z}'\mathbf{P})$, and

$$\mathrm{MSE} = \frac{1}{n-(2J+q)}\sum_{i=1}^{2}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}}\Big[(y_{ijtk} - \bar{y}_{ij..}) - \hat{\mathbf{b}}'(\mathbf{z}_{ijtk} - \bar{\mathbf{z}}_{ij..})\Big]^2,$$

where $\bar{y}_{ij..}$ and $\bar{\mathbf{z}}_{ij..}$ are averages of y_{ijtk} and $\mathbf{z}_{ijtk}$ over indices t and k, respectively.

Theorem 5. *Assume that $I = 2$ and, for patients dropping out after visit t, the y_{ijtk}'s are independent with means $\mu_{ijt} + \mathbf{b}'\mathbf{z}_{ijtk}$ and variance $\sigma^2 > 0$, and the μ_{ij}'s have form given in* (16).

(i) *The ANCOVA test based on LOCF rejects hypothesis* (17) *when the ratio F =MSTR/MSE is larger than $F_{1,n-2J-q,\alpha}$.*

(ii) *As $n_{ij} \to \infty$, $i = 1, 2$ and $j = 1, \ldots, J$,* $\mathrm{MSE} \to_p \sigma^2 + \eta$, *where*

$$\eta = \lim \frac{\sum_{i=1}^{2}\sum_{j=1}^{J} n_{ij}\tau_{ij}^2}{\sum_{i=1}^{2}\sum_{j=1}^{J} n_{ij}}, \quad \tau_{ij}^2 = \sum_{t=1}^{T} p_{ijt}(\mu_{ijt} - \mu_{ij})^2. \tag{19}$$

(iii) *Under* (17), *as $n_{ij} \to \infty$, $i = 1, 2$ and $j = 1, \ldots, J$,* $\mathrm{MSTR} \to_d (\sigma^2 + \zeta)\chi_1^2$, *where*

$$\zeta = \lim \frac{\sum_{i=1}^{2}\sum_{j=1}^{J} w_{ij}\tau_{ij}^2}{\sum_{i=1}^{2}\sum_{j=1}^{J} w_{ij}}, \quad w_{ij} = \sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}} a_{ijtk}^2. \tag{20}$$

(iv) *For testing hypothesis* (17), *the LOCF procedure described in* (i) *is asymptotically valid if and only if*

$$\lim \frac{\sum_{i=1}^{2}\sum_{j=1}^{J} n_{ij}\tau_{ij}^2}{\sum_{i=1}^{2}\sum_{j=1}^{J} n_{ij}} = \lim \frac{\sum_{i=1}^{2}\sum_{j=1}^{J} w_{ij}\tau_{ij}^2}{\sum_{i=1}^{2}\sum_{j=1}^{J} w_{ij}}. \tag{21}$$

When the design is balanced, i.e., $n_{ij} = n_0$ for all i, j, (21) can be simplified.

Corollary 1. *Assume the conditions in Theorem* 5 *and* $n_{ij} = n_0$ *for all* i, j. *Let* $\tilde{\mathbf{Z}}$ *be the* $(IJn_0) \times q$ *matrix whose rows are* $\mathbf{z}'_{ijtk} - \bar{\mathbf{z}}'_{i\ldots} - \bar{\mathbf{z}}'_{.j.} + \bar{\mathbf{z}}'_{\ldots.}$. *Then*

$$w_{ij} = (Jn_0)^{-1} + (\bar{\mathbf{z}}_{1\ldots} - \bar{\mathbf{z}}_{2\ldots})'(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\mathbf{S}_{ij}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}(\bar{\mathbf{z}}_{1\ldots} - \bar{\mathbf{z}}_{2\ldots}),$$

where $\mathbf{S}_{ij} = \sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}} (\mathbf{z}_{ijtk} - \bar{\mathbf{z}}_{i\ldots} - \bar{\mathbf{z}}_{.j.} + \bar{\mathbf{z}}_{\ldots.})(\mathbf{z}_{ijtk} - \bar{\mathbf{z}}_{i\ldots} - \bar{\mathbf{z}}_{.j.} + \bar{\mathbf{z}}_{\ldots.})'$, *and* (21) *holds if and only if* $\lim \bar{\mathbf{z}}_{1\ldots} = \lim \bar{\mathbf{z}}_{2\ldots}$ *or* $\lim n_0^{-1}\mathbf{S}_{ij}$ *are the same for all* i *and* j.

When (21) does not hold, the LOCF procedure described in Theorem 5(i) is no longer valid for testing (17). To derive a general valid test for (17), for any fixed i , let $\hat{\mathbf{b}}_i$ be the estimator of $\mathbf{b}$ based only on the data from the patients who received the ith treatment by a formula similar to (18). Define $u_{ijtk} = y_{ijtk} - \hat{\mathbf{b}}_i'\mathbf{z}_{ijtk}$. Then $\bar{u}_{i\ldots} = \frac{1}{n_i}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}} u_{ijtk}$ is an unbiased and asymptotically normal estimator for $\mu + \alpha_i$ and its variance can be estimated consistently by

$$\hat{V}_i = \frac{1}{n_i(n_i - 1)}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{k=1}^{n_{ijt}}(u_{ijtk} - \bar{u}_{i\ldots})^2.$$

Theorem 6. *Suppose that, conditional on* $(n_{ij1}, \ldots, n_{ijT})$, *the* y_{ijtk}*'s are independent with means* $\mu_{ijt} + \mathbf{b}'\mathbf{z}_{ijtk}$ *and variances* σ^2_{ijt}, *and the* μ_{ij}*'s have the decomposition given in* (16). *Under* (17), *as* $n_{ij} \to \infty$ *for all* i *and* j, $W \to_d \chi^2_{I-1}$, *where*

$$W = \sum_{i=1}^{I}\frac{1}{\hat{V}_i}\left(\bar{u}_{i\ldots} - \frac{\sum_{i=1}^{I}\bar{u}_{i\ldots}/\hat{V}_i}{\sum_{i=1}^{I}1/\hat{V}_i}\right)^2.$$

Consequently, a test of (17) with asymptotic size α rejects H_0 if $W > \chi^2_{I-1,\alpha}$. Inference about α_i's after (17) is rejected can be made using the asymptotic results based on $\bar{u}_{i\ldots}$ and $\hat{V}_i$.

The results in Theorems 5 and 6 can be extended to K-way additive ANCOVA models with $K \geq 3$. We provide a brief discussion here and omit the details. Under any K-way additive ANCOVA model, the LOCF test for the effect of a factor with two levels is asymptotically valid if the design is balanced, and either the covariate averages under the two levels are the same, or the covariate variability across different cells (combinations of factors) is constant. In any case, asymptotically valid tests can be derived along the lines of Theorem 6.

5. Simulation Results

A simulation study was performed to study the finite-sample type I error of the LOCF test and the proposed test in Theorem 6 in an ANCOVA model. The

sample sizes and population parameters were chosen to be similar to a phase II clinical trial in a real application with $I = 2$ treatments, $J = 3$ centers, and $T = 3$ visits. Responses y_{ijtk}'s were generated by the following steps. All parameter values are given in Table 1.

Table 1. Population parameters in the simulation.

i, j	$p_{ij1}, p_{ij2}, p_{ij3}$	$\mu_{ij1}, \mu_{ij2}, \mu_{ij3}$	$\sigma_{ij1}, \sigma_{ij2}, \sigma_{ij3}$
1, 1	0.30, 0.20, 0.50	$-11656.40, -11668.30, -11804.53$	251.2, 249.7, 348.7
1, 2	0.36, 0.36, 0.28	$-11682.40, -11660.80, -11758.14$	396.1, 131.3, 258.0
1, 3	0.08, 0.33, 0.59	$-12012.36, -11650.40, -11434.70$	143.8, 143.8, 591.9
2, 1	0.33, 0.07, 0.60	$-11705.31, -10473.82, -11905.71$	510.4, 486.2, 486.2
2, 2	0.18, 0.27, 0.55	$-12160.21, -11500.81, -11637.39$	371.7, 807.8, 831.5
2, 3	0.18, 0.18, 0.64	$-11767.81, -10762.31, -11720.27$	501.4, 311.6, 716.7

1. Generate covariates z_{1jtk}'s as a random sample of size $n_1 = n_{11}+n_{12}+n_{13}$ from $N(846.6, 514.1^2)$, and z_{2jtk}'s as a random sample of size $n_2 = n_{21} + n_{22} + n_{23}$ from $N(845.2, 367.7^2)$.
2. For each fixed i and j, generate $(n_{ij1}, n_{ij2}, n_{ij3})$ from Multinomial $(n_{ij}, p_{ij1}, p_{ij2}, p_{ij3})$.
3. For each fixed i, j, and t, generate y_{ijtk}'s as a random sample of size n_{ijt} from $N(\mu_{ijt} + bz_{ijtk}, \sigma^2_{ijt})$, where $b = 14.7$.

Note that the means of y_{ijtk}'s were chosen so that $\mu_{1j} = \mu_{2j}$ for all j, which implies that the model is additive with no treatment effect, that is, (17) is true. The actual type I errors of the LOCF test and the proposed test were evaluated by simulation with 5,000 runs. Results with different choices of n_{ij}'s are summarized in Table 2. The results indicate that the proposed test has size close to the nominal level 5% regardless of whether the sample sizes are balanced or not. The results also show that the LOCF test has the right size when the sample sizes are almost balanced, but a wrong size when the sample sizes are unbalanced. These simulation results generally support our asymptotic results.

Table 2. Type I errors of the LOCF and proposed tests.

Sample Size		Type I Error	
n_{11}, n_{12}, n_{13}	n_{21}, n_{22}, n_{23}	LOCF Test	Proposed Test
30, 33, 36	27, 33, 33	0.0548	0.0544
60, 66, 72	27, 33, 33	0.1118	0.0520
15, 17, 18	27, 33, 33	0.0170	0.0518
30, 33, 36	54, 66, 66	0.0170	0.0496
30, 33, 36	14, 17, 17	0.1078	0.0576

Appendix

Proof of Theorem 1.

(ii) Let $u_{itk} = y_{itk} - \mathbf{b}'\mathbf{z}_{itk}$. The result follows from Lemma 1 of Shao and Zhong (2003) and the fact that

$$\begin{aligned}\text{MSE} &= \frac{1}{n_1+n_2-q-2}\sum_{i=1}^{2}\sum_{t=1}^{T}\sum_{k=1}^{n_{it}}\Big[(u_{itk}-\bar{u}_{i..})-(\hat{\mathbf{b}}-\mathbf{b})'(\mathbf{z}_{itk}-\bar{\mathbf{z}}_{i..})\Big]^2\\ &= \frac{1}{n_1+n_2-q-2}\sum_{i=1}^{2}\sum_{t=1}^{T}\sum_{k=1}^{n_{it}}(u_{itk}-\bar{u}_{i..})^2+o_p(1).\end{aligned}$$

(iii) Note that

$$(\bar{y}_{1..}-\hat{\mathbf{b}}'\bar{\mathbf{z}}_{1..})-(\bar{y}_{2..}-\hat{\mathbf{b}}'\bar{\mathbf{z}}_{2..})=\sum_{t=1}^{T}\sum_{k=1}^{n_{1t}}a_{1tk}y_{1tk}-\sum_{t=1}^{T}\sum_{k=1}^{n_{2t}}a_{2tk}y_{2tk}.$$

By an argument similar to that of Theorem 1 of Shao and Zhong (2003) and Theorem 3.12 of Shao (2003), we have

$$\Big[(\bar{y}_{1..}-\hat{\mathbf{b}}'\bar{\mathbf{z}}_{1..})-(\bar{y}_{2..}-\mathbf{b}'\bar{\mathbf{z}}_{2..})\Big]\Big/\sqrt{\Big(\sum_{i=1}^{2}\sum_{t=1}^{T}\sum_{k=1}^{n_{it}}a_{itk}^2\Big)\sigma^2+\phi_1^2+\phi_1^2}\to_d N(0,1),$$

where $\phi_i^2=\text{Var}(\sum_{t=1}^{T}\sum_{k=1}^{n_{it}}a_{itk}(\mu_{it}+\mathbf{b}'\mathbf{z}_{itk}))$. Then the result follows if $\phi_i^2=\tau_i^2w_i^2$ holds for $i=1,2$. We only prove the case when $i=1$ as an illustration. Since the covariate value of a patient does not vary with visit, we can rewrite $\mathbf{z}_{itk}$ as $\mathbf{z}_{ij}$ and a_{itk} as a_{ij}, where $j=1,\ldots,n_1$. For $t=1,\ldots,T$, let $X_{1jt}=1$ if and only if a_{1j} is from a patient who dropped out after the tth visit, and $X_{1jt}=0$ otherwise. Clearly, $X_{1j}=(X_{1j1},\ldots,X_{1jT})\overset{\text{i.i.d.}}{\sim}\text{Multinomial}(1;p_{11},\ldots,p_{1T})$, $j=1,\ldots,n_1$. Hence

$$\begin{aligned}\phi_1^2 &= \text{Var}\Big(\sum_{t=1}^{T}\sum_{j=1}^{n_1}a_{1j}(\mu_{1t}+\mathbf{b}'\mathbf{z}_{1j})X_{1jt}\Big)\\ &= \text{Var}\Big(\sum_{j=1}^{n_1}a_{1j}\Big(\sum_{t=1}^{T}\mu_{1t}X_{1jt}+\sum_{t=1}^{T}\mathbf{b}'\mathbf{z}_{1j}X_{1jt}\Big)\Big)\\ &= \text{Var}\Big(\sum_{j=1}^{n_1}a_{1j}\sum_{t=1}^{T}\mu_{1t}X_{1jt}+\sum_{j=1}^{n_1}a_{1j}\mathbf{b}'\mathbf{z}_{1j}\Big)\\ &= \text{Var}\Big(\sum_{j=1}^{n_1}a_{1j}\sum_{t=1}^{T}\mu_{1t}X_{1jt}\Big)\end{aligned}$$

$$= \sum_{j=1}^{n_1} a_{1j}^2 \mathrm{Var}\Big(\sum_{t=1}^{T} \mu_{1t} X_{1jt} \Big)$$
$$= \tau_1^2 \sum_{j=1}^{n_1} a_{1j}^2 = \tau_1^2 \sum_{t=1}^{T} \sum_{k=1}^{n_{1s}} a_{1tk}^2 = \tau_1^2 w_1^2.$$

Acknowledgements

This work was partially supported by grant CA53786 from the National Cancer Institute. The authors would like to thank two referees and an associate editor for their helpful comments.

References

Cheng, B. (2004). Some hypothesis testing results in two-way linear models for clinical trials. Ph.D. thesis, University of Wisconsin-Madison.

Dawson, J. D. (1994a). Comparing treatment groups on the basis of slopes, areas-under-the-curve, and other summary measures. *Drug Infor. J.* **28**, 723-732.

Dawson, J. D. (1994b). Stratification of summary statistic tests according to missing patterns. *Statist. Medicine* **13**, 1853-1863.

Dawson, J. D. and Lagakos, S. W. (1993). Size and power of two-sample tests of repeated measures data. *Biometrics* **49**, 1022-1032.

FDA (1988). *Guideline for Format and Content of the Clinical and Statistical Sections of New Drug Applications*. Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.

Heyting, A. and Tolboom, J. T. B. M. and Essers, J. G. A. (1992). Statistical handling of drop-outs in longitudinal clinical trials. *Statist. Medicine* **11**, 2043-2061.

Lavori, P. W. (1992). Clinical trials in psychiatry: Should protocol deviation censor patient data? *Neuropsychopharmacology* **6**, 39-48.

Little, R. (1993). Pattern-Mixure Models for Multivariate Incomplete Data. *J. Amer. Statist. Assoc.* **88**, 125-134.

Searle, S. R. (1987). *Linear Models for Unbalanced Data*. Wiley, New York.

Shao, J. (2003). *Mathematical Statistics*. Second edition. Springer-Verlag, New York.

Shao, J. and Zhong, B. (2003). Last Observation Carry-Forward and Last Observation Analysis. *Statist. Medicine* **22**, 2429-2441.

Shih, W. and Quan, H. (1998). Stratified testing for treatment effects with missing data. *Biometrics* **54**, 782-787.

Ting, N. (2000). Carry-forward analysis. In *Encyclopedia of Biopharmaceutical Statistics* (Edited by S. Chow), 103-109. Marcel Dekker, New York.

Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032, U.S.A.

E-mail: bc2159@columbia.edu

Department of Statistics, University of Wisconsin, Madison, WI 53706, U.S.A.

E-mail: shao@stat.wisc.edu

Centocor, Inc., Malvern, PA 19355, U.S.A.

E-mail: BZhong@CNTUS.JNJ.COM

(Received December 2003; accepted November 2004)