Statistica Sinica **11**(2001), 53-61

# ON THE COMPOSITE APPROACH TO DROPOUTS IN CLINICAL TRIALS

Weichung Joseph Shih and Hui Quan

*University of Medicine and Dentistry of New Jersey and Merck Research Laboratories*

*Abstract:* A major problem in the analysis of clinical trials is missing data from patients who drop out of the study before the predetermined schedule. Shih and Quan (1997) proposed a "composite" (or "pattern but not mix") approach to the problem, which makes simultaneous inference on the conditional mean of the completers and the probability of completion or dropping out of the trial. Their method involves the combination of test statistics for the conditional mean and dropout probability. In this paper, we justify the asymptotic independence of the test statistics used in Shih and Quan.

*Key words and phrases:* Asymptotic independence, missing data, pattern-mixture modeling, Weiler's test.

## 1. Introduction

Consider a randomized clinical trial that compares two treatment groups with respect to a continuous outcome, $y$, such as blood pressure, cholesterol level, CD4 counts, or bone mineral density. Repeated measurements are often taken during the study, but the outcome at the end of the study (as opposed to, say, the rate of change) is the main interest here. A well-known problem in the analysis of such experiments is the data missing from subjects who drop out before the pre-determined study termination. Ethically, the design of a clinical trial has to permit patients to withdraw for unfavorable reasons such as death, adverse reactions, intolerable treatment or procedure, lack of improvement, etc. But the design and conduct of every clinical trial can and should encourage patients with favorable outcomes (such as early recovery) to remain in the study to achieve the full benefit of treatment. Hence, the great majority of patients in a well-designed and conducted study can usually be classified as "completers" or "dropouts for unfavorable reasons". See Shih and Quan (1997) - or SQ henceforth.

Let $Y \sim f_\theta(Y)$ be the complete-data of the outcome measure. When $Y$ is incompletely observed, we write $Y = (Y_o, Y_m)$, where $Y_o$ is observed and $Y_m$ is

missing. Further, let $M$ denote the missing-data indicator that corresponds to $Y$. We may factor the joint distribution of $Y$ and $M$, $f(Y, M)$, as

$$f(Y, M) = f_\Psi(Y|M) f_\pi(M). \tag{1}$$

The first factor in (1) says that the data are stratified by the missing data pattern. The so-called "pattern-mixture" approach (Glynn, Laird and Rubin (1986), Little and Rubin (1987) and Little (1993)) is to integrate estimates of $\Psi$ and $\pi$ to make inference on $\theta$, i.e., the over-pattern, hypothetical data parameter.

SQ argued in the context of clinical trials that when the over-pattern and pattern-specific parameters differ, the conditional parameter in $\Psi$, together with the marginal missing indicator parameter $\pi$, are of substantive interest to the medical investigation in their own right. It makes clinical sense to regard the patients who completed the trial and those who dropped out (for unfavorable reasons) as different populations. SQ advocated a composite analysis that considers the chance of dropping out as an important outcome, together with the expected final response for patients who remain through the prescribed treatment course. Since SQ's approach does not "mix" the estimate of $(\Psi, \pi)$ to obtain that of $\theta$, rather focuses on $(\Psi, \pi)$, the method may be called the "Pattern but not mix" approach.

In comparing treatment groups, SQ proposed the use of regression or, in large data base cases, matching by propensity scores, to adjust for the possible baseline differences in the completers, and multiple-test procedures for the composite of joint and individual hypotheses pertaining to the parameters in $(\Psi, \pi)$. It is important to incorporate necessary covariates, such as the initial severity of the disease, the baseline value of $y$, and more, to make the treatment groups as comparable as possible for the completers. Hence, the method extends to models with covariates $W = (X, \Gamma)$ that include treatment groups $(X)$, the baseline measure of the outcome and other characteristics of the patient $(\Gamma)$:

$$f(Y, M|W) = f_\psi(Y|M, W) f_\pi(M|W). \tag{2}$$

For simplicity, and without confusion, we use the same notation for parameters in (1) and (2). Notice that in (1) or (2), the marginal distribution of the missing indicator $M$ may depend on covariates $W$, but does not depend on the response variable $Y$. This makes the modeling of $M$ much easier than the so-called selection-model approach, which factors $f(Y, M|W)$ as $f(M|Y, W)\ f(Y|W)$. In the selection model, the conditional distribution of $M$ depends on $Y$, which could involve the unobserved part of $Y$.

A potential difficulty in making simultaneous inference on $(\Psi, \pi)$ is to establish the asymptotic normality and independence of the estimates of the component parameters. We specify the model and derive the estimates of the component parameters in Section 2. The main result is in Section 3: the proof of a

theorem with regularity conditions under which the asymptotic normality and independence of the estimates hold. Justification of the regularity conditions in the context of clinical trials is also given. In Section 4, we apply the theorem to form a joint test for parameters in $(\Psi, \pi)$.

## 2. "Pattern But Not Mix" Approach

Let $m_i$ be 1 if the $i$th patient is a completer, and 0 if a dropout, $i = 1, \ldots, n$. Of course the final response measurement $y_i$ (for the $i$th patient) exists only if $m_i = 1$. Let $x_i$ be the indicator for treatment groups, $x_i = 0$ or 1, and let $\Gamma_i$ denote the vector of covariates (other than the treatment group $x_i$) for the $i$th patient. Consider the following general linear regression models for dropout probability and conditional mean of completers:

$$P(m_i = 0|\Gamma_i, x_i) = g(\beta_0 + \eta'\Gamma_i + \beta x_i), \tag{3a}$$

$$y_i|(m_i = 1, \Gamma_i, x_i) = \lambda_0 + \gamma'\Gamma_i + \lambda x_i + \varepsilon_i, \tag{3b}$$

where, corresponding to (2), $\pi = (\beta_0, \eta', \beta)'$ and $\Psi_1 = (\lambda_0, \gamma', \lambda)'$, a subset of $\Psi$, are unknown regression coefficients. We assume that the errors $\varepsilon_i$ in (3b) are i.i.d. with mean 0 and constant variance $\sigma^2$. Possible choices of $g(u)$ are the logistic, standard normal (probit), and extreme-value c.d.f's (Cox and Snell (1989)). It is also possible to measure time to dropout and to employ survival models for (3a). We assume that for the completers, the treatment groups are comparable with respect to their baseline characteristics after adjusting for the covariates.

Unlike the pattern-mixture model (for sampling surveys), the $y$ value given $m = 0$ part is not well-defined in clinical trials because, for example, the measures of a patient's cholesterol level or symptom score after death, or the glomerular filtration rate (GFR) after kidney dialysis are no longer meaningful for the therapy under study.

We are mainly interested in testing the treatment difference with regard to the dropout rate and the conditional mean, that is, $(\beta, \lambda)$; others are nuisance parameters. For example, if $x = 0$ denotes the control group, $\lambda > 0$ means that the measurement for the completers tends to be greater in the test group than in the control group, and in the case of $g(u) = 1/(1 + e^u), \beta > 0$ means less chance of drop out due to unfavorable events for the test group than for the control group.

Since the inference on $\beta$ only involves $(m_i, x_i, \Gamma_i)$, $i = 1, \ldots, n$, we can obtain the MLE $\hat{\beta}$ of $\beta$ by a usual regression for binary outcomes, such as logistic or probit (see, e.g., Cox and Snell (1989), Agresti (1984)). Let $\hat{\hat{\omega}}^2/n$ be the estimate of the asymptotic variance of $\hat{\beta}$ such that $\sqrt{n}\ (\hat{\beta} - \beta)/\hat{\hat{\omega}}$ has asymptotically a

standard normal distribution. For $\lambda$, we obtain the least-squares estimate $\hat{\lambda}$, conditioning on $m_i$, $i = 1, \ldots, n$, and its conditional variance $\sigma^2/b_n$, where $b_n = (\sum_{i=1}^n m_i x_i^2) - (\sum_{i=1}^n m_i \Gamma_i^{*\prime} x_i)(\sum_{i=1}^n m_i \Gamma_i^* \Gamma_i^{*\prime})^{-1}(\sum_{u=1}^n m_i x_i \Gamma_i^*)$ and $\Gamma_i^* = (1, \Gamma_i')'$; see Section 3. As usual, a consistent estimate of the error variance $\sigma^2$ is the residual mean squares, $\hat{\sigma}^2$.

In general, since the $m_i$'s in $b_n$ are random, we need special care when using $(\hat{\beta}, \hat{\lambda})$ to make simultaneous inference on $\beta$ and $\lambda$. When $\varepsilon_i$ are i.i.d. $N(0, \sigma^2)$, we know that $[(\hat{\lambda} - \lambda)|m_i, i = 1, \ldots, n]$ is distributed as $N(0, \sigma^2/b_n)$. Thus $\frac{(\hat{\lambda}-\lambda)}{\sigma} b_n^{1/2} | m_i \sim N(0,1)$ independently of the $m_i, i = 1, \ldots, n$, and $\frac{\hat{\lambda}-\lambda}{\sigma} b_n^{1/2} \sim N(0,1)$, unconditionally on $m_i$. For general error structures, we show in the following section that, under some appropriate regularity conditions, $\frac{\hat{\lambda}-\lambda}{\sigma} b_n^{1/2}$ is asymptotically distributed $N(0,1)$, unconditionally on $m_i$ , as $n$ becomes large.

## 3. Asymptotic Independence of $\hat{\beta}$ and $\hat{\lambda}$

Referring to (3b), let $\Gamma_i^* = (1, \Gamma_i')'$ and $\phi^* = (\lambda_0, \gamma')'$. Furthermore, let $y_i^* = y_i$ if $m_i = 1$ and $y_i^* =$ missing value if $m_i = 0$. Let

$$X_n^* = \begin{bmatrix} m_1 \Gamma_1^{*\prime} & m_1 x_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ m_n \Gamma_n^{*\prime} & m_n x_n \end{bmatrix}_{n \times p}, \quad Y_n^* \begin{bmatrix} y_1^* \\ \cdot \\ \cdot \\ y_n^* \end{bmatrix}_{n \times 1} \quad \text{and} \quad \Psi = \begin{bmatrix} \phi^* \\ \lambda \end{bmatrix}_{p \times 1}. \tag{4}$$

The least-squares estimate $\hat{\Psi}_1$ of $\Psi_1$ (conditioning on $X_n^*$, hence $m_i, i = 1 \ldots, n$) is $\hat{\Psi}_1 = (X_n^{*\prime} X_n^*)^{-1} X_n^{*\prime} Y_n^*$, where zero times anything (including the missing value) is defined as zero. Let

$$S_n = X_n^{*\prime} X_n^* = \begin{bmatrix} \sum_{i=1}^n m_i \Gamma_i^* \Gamma_i^{*\prime} & \sum_{i=1}^n m_i \Gamma_i^* x_i \\ \sum_{i=1}^n m_i \Gamma_i^{*\prime} x_i & \sum_{i=1}^n m_i x_i^2 \end{bmatrix}_{p \times p} \tag{5}$$

and $(u_{n1}, \ldots, u_{nn})$ be the last row of $S_n^{-1} X_n^{*\prime}$. The $(p,p)$-th element of $S_n^{-1}$ is $b_n^{-1}$, where $b_n = (\sum_{i=1}^n m_i x_i^2) - (\sum_{i=1}^n m_i \Gamma_i^{*\prime} x_i)(\sum_{i=1}^n m_i \Gamma_i^* \Gamma_i^{*\prime})^{-1}(\sum_{i=1}^n m_i x_i \Gamma_i^*)$.

**Lemma.** *If (c.1) $E_{y|m}(\varepsilon_i^{2+\delta}) < \infty$ for some $\delta$ in $(0,1]$, then*

$$|Pr(\frac{\hat{\lambda} - \lambda}{\sigma} b_n^{1/2} < t) - \Phi(t) \leq C E_m(\sum_{i=1}^n |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2} \tag{6}$$

*for some constant $C$, where $\Phi(t)$ is the c.d.f. of the standard normal distribution.*

**Proof.** It can be seen that the least-square estimate $\hat{\lambda}$ of $\lambda$ is $\hat{\lambda} = \sum_{i=1}^n u_{ni} y_i^* m_i$ and its (conditional) variance is $\sigma^2 \sum_{i=1}^n u_{ni}^2 = \sigma^2 / b_n$. Note that $u_{ni} = 0$ when

$m_i = 0$. Note that $\hat{\lambda} - \lambda = \sum_{i=1}^{n} u_{ni}\varepsilon_i m_i$ and that $\varepsilon_i$ are i.i.d. for those with $m_i = 1$. If $E_{y|m}(\varepsilon_i^{2+\delta}) < \infty$ for some $\delta$ in $(0, 1]$, then by the Berry-Essen Theorem (Chow and Teicher (1978)) there is a constant $C$ such that

$$\sup_t |Pr(\frac{\hat{\lambda}-\lambda}{\sigma} b_n^{1/2} < t | m_i, i=1,\ldots,n) - \Phi(t)| \leq C(\sum_{i=1}^{n} |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2}, \ -\infty < t < \infty$$

where $\Phi(t)$ is the c.d.f. of the standard normal distribution. Thus

$$\begin{aligned}
&|Pr(\frac{\hat{\lambda}-\lambda}{\sigma} b_n^{1/2} < t) - \Phi(t)| \\
&= |E_m[Pr(\frac{\hat{\lambda}-\lambda}{\sigma} b_n^{1/2} < t | m_i, i=1,\ldots,n) - \Phi(t)]| \\
&\leq E_m|Pr(\frac{\hat{\lambda}-\lambda}{\sigma} b_n^{1/2} < t | m_i, i=1,\ldots,n) - \Phi(t)| \\
&\leq C E_m(\sum_{i=1}^{n} |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2}.
\end{aligned}$$

**Theorem.** *Given* (c.1) *and* (c.2) $S_n/n$ *converges almost surely* (*a.s.*) *to a positive definite matrix* $\sum$, (c.3) *All covariates are bounded by a constant* $B$, $\frac{\hat{\lambda}-\lambda}{\sigma} b_n^{1/2} \to N(0,1)$ *in distribution, unconditionally on* $m_i, i = 1, \ldots, n$.

**Proof.** We first show that $(\sum_{i=1}^{n} |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2} \to 0$ a.s. as $n \to \infty$. Let $a_n' = \{a_{nj}; j = 1, \ldots, p\}$ be the last row of $S_n^{-1}$, $a'$ be the last row of $\sum^{-1}$. From condition (c.2), $na_n' \to a'$ almost surely. Then $b_n = (\sum_{i=1}^{n} m_i x_i^2) - (\sum_{i=1}^{n} m_i \Gamma_i^{*\prime} x_i)\ (\sum_{i=1}^{n} m_i \Gamma_i^* \Gamma_i^{*\prime})^{-1}\ (\sum_{i=1}^{n} m_i x_i \Gamma_i^*) \leq \sum_{i=1}^{n} x_i^2 \leq n$. Therefore,

$$\begin{aligned}
&(\sum_{i=1}^{n} |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2} = O[n^{(2+\delta)/2}(\sum_{i=1}^{n} |u_{ni}|^{2+\delta})] = O[n^{(2+\delta)/2}(\sum_{i=1}^{n} |a_n' z_i|^{2+\delta})] \\
&= O[n^{(2+\delta)/2}(1/n)^{2+\delta}(\sum_{i=1}^{n} |a' z_i|^{2+\delta})] = O[n^{-\delta/2}(\frac{1}{n}\sum_{i=1}^{n} |a' z_i|^{2+\delta})], \qquad (7)
\end{aligned}$$

where $z_i = m_i(\Gamma_i^{*\prime}, x_i)'$. Following (c.3) and (7), $(\sum_{i=1}^{n} |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2} \to 0$ a.s. as $n \to \infty$.

We next show that $E_m \sup_n (\sum_{i=1}^{n} |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2}$ is bounded above:

$$\begin{aligned}
(\sum_{i=1}^{n} |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2} &\leq n^{(2+\delta)/2}(\sum_{i=1}^{n} (u_{ni}|^{2+\delta}) = n^{(2+\delta)/2}(\sum_{i=1}^{n} |a_n' z_i|^{2+\delta}) \\
&\leq n^{(2+\delta)/2} B^{2+\delta} \sum_{i=1}^{n} (p \times \sup_j |a_{nj}|^{2+\delta}) \\
&= n^{(2+\delta)/2} B^{2+\delta} n (p \times \sup_j |a_{nj}|)^{2+\delta}.
\end{aligned}$$

Since $S_n^{-1}$ is of the order of $1/n$ from (c.2), $a'_n$ is also of the order of $1/n$. Thus, $n^{2+\delta/2}(|a_{nj}|^{2+\delta})$ is of the order of $n^{-\delta/2}$ for all $j$ and $n$. It follows that

$$E_m \text{sup}_n (\sum_{i=1}^{n} |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2} \leq B^{2+\delta} p^{2+\delta} E_m \text{sup}_n (n^{2+\delta/2} \text{sup}_j |a_{nj}|^{2+\delta}) \leq \infty.$$

Thus, by the Lebesgue Dominated Convergence Theorem,

$$E_m (\sum_{i=1}^{n} |u_{ni}|^{2+\delta}) b_n^{(2+\delta)/2} \to 0 \text{ as } n \to \infty. \tag{8}$$

The theorem is proved following (6).

**Remark.** Adjusting the baseline covariates is an essential consideration in the analysis of the completers to reduce the potential impact of possible differential patient characteristics between the treatment groups. Therefore, it is perhaps not surprising to see that the above regularity conditions all involve baseline covariates. Justifications of these conditions in the context of clinical trials can be made as follows. First, it is easy to see that condition (c.3) holds since baseline covariates, such as age, sex, initial severity scores or clinical/physiologic measurements are always bounded. Second, for (c.2), since patients in a clinical trial are supposed to be random samples of a patient population, we can regard covariates $\Gamma_i$ as i.i.d. random variables. The $x_i$ is a random indicator for treatment assignment by the study design. Hence the components of the $S_n/n$ matrix are essentially sample means of the corresponding quantities in (5), thus will converge. Furthermore, in most clinical trials the dropouts would not be a major portion of the whole patient sample and the proportion should be roughly constant as the sample size gets larger (otherwise the usefulness of the study is questionable). Consequently it is conceivable that the covergence of $S_n/n$ is to a positive definite matrix, and the convergence rate of $\frac{\hat{\lambda}-\lambda}{\sigma} b_n^{1/2}$ to $N(0,1)$ should be similar to that when there is no dropout, i.e., at the rate of $n^{-\delta/2}$.

**Corollary.** *Under* (c.1)-(c.3) *and the regularity conditions that* $\sqrt{n}(\hat{\beta}-\beta)/\hat{\bar{\omega}}$ *has asymptotically a standard normal distribution,* $\hat{\beta}$ *and* $\hat{\lambda}$ *are asymptotically independent.*

**Proof.**

$$\begin{aligned}
&Pr[\sqrt{n}(\hat{\beta}-\beta)/\hat{\bar{\omega}} < u, (\hat{\lambda}-\lambda) b_n^{1/2}/\sigma < v] - \Phi(u)\Phi(v) \\
&= EI[\sqrt{n}(\hat{\beta}-\beta)/\hat{\bar{\omega}} < u] I[\hat{\lambda}-\lambda) b_n^{1/2}/\sigma < v] - \Phi(u)\Phi(v) \\
&= EI[\sqrt{n}(\hat{\beta}-\beta)/\hat{\bar{\omega}} < u] \{I[(\hat{\lambda}-\lambda) b_n^{1/2}/\sigma < v] - \Phi(v)\} \\
&\quad + E\{I[\sqrt{n}(\hat{\beta}-\beta)/\hat{\bar{\omega}} < u] - \Phi(u)\}\Phi(v).
\end{aligned}$$

The second term converges to zero, and

$$\begin{aligned} & EI[\sqrt{n}(\hat{\beta}-\beta)/\hat{\tilde{\omega}} < u]\{I[\hat{\lambda}-\lambda)b_n^{1/2}/\sigma < v] - \Phi(v)\} \\ &= EI[\sqrt{n}(\hat{\beta}-\beta)/\hat{\tilde{\omega}} < u]\{Pr[\hat{\lambda}-\lambda)b_n^{1/2}/\sigma < v|m_i] - \Phi(v)\} \\ &\leq CE_m(\sum_{i=1}^{n}|u_{ni}|^{2+\delta})b_n^{(2+\delta)/2} \to 0 \end{aligned}$$

by (6) and (8).

## 4. Application for Analysis and Study Design

As noted previously, we have essentially a "two endpoints" problem that needs to be addressed simultaneously: parameter $\lambda$ represents the treatment effect for completers and $\beta$ compares the dropout rates between treatment groups. After showing that $\hat{\beta}$ and $\hat{\lambda}$ are asymptotically independent, the task of combining becomes easy. Several simultaneous tests that combine $\hat{\beta}$ and $\hat{\lambda}$ were given in SQ, including the generalized, large sample version of Weiler's test (1964), Bonferroni-type adjustment of $p$-values, and the weighted test. The purpose of having a combined test is mainly to increase power and, at the same time, control the overall type-I error rate. See examples in SQ. The sample size calculation at the design stage can also be considered using the weighted test by specifying $(\beta, \lambda)$ and their assumed variances in the alternative hypothesis (see SQ page 1231).

## 5. Discussion

This paper provides the theoretical justification for the previous paper of Shih and Quan (1997), in which the spirit of the "composite" or "pattern but not mix" approach to dropouts problem in clinical trials is discussed in detail. The importance of the results established in this paper, namely the asymptotic normality and independence of the test statistics without assuming the normality of the random error, goes beyond the several combined tests given in SQ. For example, as was suggested in SQ, the dropout probability can be extended to "time to dropout", and the "conditional mean" at the last time-point of the completers can be extended to "slope type" parameters for repeated measures. The extension of the models (3a and 3b) in Section 2 to a survival model and a mixed effects model, or even to a generalized linear model, is future work. The key is to show, with proper conditions, that the estimate of $\lambda$ is asymptotic normal unconditionally on the missing data pattern. Hogan and Laird (1997) recently considered this kind of extension in their approach to the dropout problem. However, in their paper (p.244), even under a mixed effect model for a normally distributed response variable, the asymptotic independence of the estimates had

to be assumed. Our results in Section 3 might provide a framework to prove the asymptotic independence for this extended problem.

As a methodology useful for dealing with incomplete data from clinical trials (see Myers (1999)), we would like to comment on the term "bias", which is often ambiguously applied to "completers only" analysis. First of all, the composite (or pattern but not mix) approach proposed is not "completers only" analysis since, as argued in SQ and here, the dropout probability is an important outcome to be included in the comparison between treatment groups. This is the spirit of the "composite" approach. Second, the completers are a subgroup and, as others have pointed out, would provide a biased estimate of the treatment effect for the whole patient group. However, it would not be biased for the completers subgroup itself, and is what the conditional parameter $\lambda$ is all about. Simply put, $\hat{\lambda}$ may be biased for $\theta$, but is unbiased for $\lambda$.

The main idea of the SQ paper was to argue that the conditional treatment effect for the completers ($\lambda$) is more clinically relevant than the marginal treatment effect for the hypothetical average of "completers plus dropouts" ($\theta$) when dropouts are present. Consider the case of hypertension, for example. When a clinician prescribes a anti-hypertensive treatment to a patient based on a study result where early dropout is a problem, he or she would advise the patient as follows: "Here is a treatment for your hypertension. You need to take the drug for 6 months according to the prescription. If you complete the treatment course for 6 months, I expect your blood pressure to be lowered by 15%. But, according to the clinical study, there is 10% chance that you may not complete the prescribed course because of the side effects associated with the drug. ...." This and the paper of SQ aim to support this kind of statement for clinicians.

We recognize that in the literature, many people have discussed making inference of the hypothetical complete data parameter $\theta$, in relation to the so-called "intention to treat" (ITT) principle. However, it is also recognized that when dropouts occur in a study, the ITT principle is difficult to achieve since no measurement data will be available (unlike the mortality study where ITT originated, since each patient's survival endpoint is always known); see, for example, Lewis and Machin (1993). This and the SQ paper, on the other hand, emphasize testing for the pragmatic incomplete data parameters $\lambda$ (for completers) and $\beta$ (for dropouts). The ITT intends to answer the question: "What would be the treatment effect *if* patients would not dropout?" The latter addresses the question: "What is the treatment comparison in the presence of dropouts?" We believe the latter is also, if not more so, a pertinent question for clinical trials.

Other bias is subtler. That is, the completers form a post-randomization subgroup, as addressed in Shih and Quan (1998), hence may not be balanced with respect to the baseline characteristics between the treatment groups. Essentially,

the incomplete data is semi-observational and needs to be treated as such. We addressed this issue by using baseline covariates in (3b), and assumed that the treatment groups are comparable for the completers with respect to their baseline characteristics after adjusting for the covariates. In SQ, we also suggest using matching by propensity scores, but this requires a very large data set.

## Acknowledgement

The authors are grateful to the referees for their careful review and helpful comments on the initial manuscript, and to Professor Cun-Hui Zhang for discussion regarding the topic.

## References

Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. Wiley, New York.

Chow, Y-S and Teicher, H. (1978). *Probability Theory-Independence, Interchangeability, and Martingales*. Springer-Verlag, New York.

Cox, D. R. and Snell, E. J. (1989). *The Analysis of Binary Data*. 2nd edition. Chapman and Hall, London.

Glynn, R. J., Laird, N. M. and Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In *Drawing Inferences from Self-selected Samples* (Edited by H. Wainer), 115-142. Springer-Verlag, New York.

Hogan, J. W. and Laird, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statist. Medicine* **16**, 239-257.

Lewis, J. A. and Machin, D. (1993). Intention to treat – who should use ITT? *Br. J. Cancer* **68**, 647-650.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis With Missing Data*. Wiley, New York.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *J. Amer. Statist. Assoc.* **88**, 125-134.

Myers, W. (1999). Dealing with dropouts in clinical studies – Biopharmaceutical section roundtables, joint statistical meetings 1998. *Biopharmaceutical Report* **7**, 22-23.

Shih, W. J. and Quan, H. (1997). Testing for treatment difference with dropouts present in clinical trials - A composite approach. *Statist. Medicine* **16**, 1225-1239.

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

Weiler, H. (1964). A significance test for simultaneous quantal and quantitative responses. *Technometrics* **6**, 273-285.

Division of Biometrics, UMDNJ School of Public Health & Cancer Institute of New Jersey, University of Medicine and Dentistry of New Jersey, 195 Little Albany Street, New Brunswick, New Jersey 08901, U.S.A.

E-mail: shihwj@umdnj.edu

Merck Research Laboratories.

E-mail: hui_quan@merck.com

(Received August 1998; accepted August 2000)