Published in final edited form as:

Stat Methods Med Res. 2018 April ; 27(4): 1067–1075. doi:10.1177/0962280216651307.

A novel complete-case analysis to determine statistical significance between treatments in an intention-to-treat population of randomized clinical trials involving missing data

Wei Liu¹ and Jinhui Ding²

¹Rockville, MD, USA

²Bioinformatics Core, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

Abstract

The application of the principle of the intention-to-treat (ITT) to the analysis of clinical trials is challenged in the presence of missing outcome data. The consequences of stopping an assigned treatment in a withdrawn subject are unknown. It is difficult to make a single assumption about missing mechanisms for all clinical trials because there are complicated reactions in the human body to drugs due to the presence of complex biological networks, leading to data missing randomly or non-randomly. Currently there is no statistical method that can tell whether a difference between two treatments in the ITT population of a randomized clinical trial with missing data is significant at a pre-specified level. Making no assumptions about the missing mechanisms, we propose a generalized complete-case (GCC) analysis based on the data of completers. An evaluation of the impact of missing data on the ITT analysis reveals that a statistically significant GCC result implies a significant treatment effect in the ITT population at a pre-specified significance level unless, relative to the comparator, the test drug is poisonous to the non-completers as documented in their medical records. Applications of the GCC analysis are illustrated using literature data, and its properties and limits are discussed.

Keywords

Drug toxicity; benefit–risk; medical information; impact of missing data; inferential analysis

1 Introduction

The principle of intention-to-treat (ITT) is widely accepted as a gold standard for the analysis of randomized controlled clinical trials.¹ The ITT analysis of the difference between two treatments includes every subject who is randomized.^{2,3} However, missing outcome data are commonly unavoidable in a randomized clinical trial when patients leave the study

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

Corresponding author: Wei Liu, P.O. Box 1121, Rockville, MD 20849, USA. ldwjh08@gmail.com.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

before the protocol-specified completion time.⁴⁻⁶ In the presence of missing data, statisticians have been perplexed by how to perform an ITT analysis to answer the following two fundamental questions on data interpretation: first, how to determine if the treatment effect in the ITT population is significant at a pre-specified level (ITT Puzzle I), and, second, how to make an unbiased estimate of the difference between two treatments in the ITT population (ITT puzzle II).

Currently there are two main approaches for analyzing clinical trial data involving missing data in practice. The first one intends to fill missing data with assumed or simulated values under certain assumptions.⁷⁻¹¹ However, these assumptions are often difficult to verify.⁴ The second approach, commonly known as complete-case analysis, simply uses the observed data often with certain restrictions (such as completers with or without rescue medications, per protocol population, etc.), without accommodating missing data.¹² The current method of completers analysis is problematic because the number of completers is treated as a constant. The National Research Council (NRC) Panel on Handling Missing Data in Clinical Trials provided general principles in the report *The Prevention and Treatment of Missing Data in Clinical Trials*.^{13,14} The NRC report also recommended some mathematical approaches to analysis, intended to restore the balance established by randomization among treatment groups by adjusting for the association among outcome data, covariates, and missingness. In practice, however, there are currently no statistical methods available to solve either one of the ITT puzzles.

The missingness in clinical trials has specific properties that are different from those in simple surveys which form the basis for much of the work on missing data. To understand the reactions of the human body to drugs, note that there are complex biological metabolism pathways or networks.¹⁵⁻¹⁷ There are multiple pathways to cause a clinical sign or organ failure and, however, from another perspective, a single treatment could lead to multiple effects.¹⁸⁻²⁰ In fact, there are various reasons of patient dropouts in a typical clinical trial, including different adverse drug reactions and adverse events (AEs) in various tissues and organs, symptoms of nausea or diarrhea, withdrawal based on medical doctor/investigator's decision, withdrawal of subjects, or and others. It is rare that a single reason causes all the dropouts of subjects. Therefore, it is unlikely that one statistical assumption about the missing mechanism of the population of the non-completers could handle the general missing data issue in the analysis of clinical trials.

Making no assumptions about the missing mechanisms in the present study, we evaluated the impact of missing data on the ITT analysis, based on actually observed data, and proposed a new statistical analysis of clinical trials involving missing data, which we named the generalized complete-case (GCC) analysis to distinguish it from the current completers analysis. This method addresses the ITT Puzzle I. We showed that a statistically significant GCC result implies a significant treatment effect in the ITT population at a pre-specified significance level unless, relative to the comparator, the test drug is toxic to the non-completers.

This paper is organized as follows. In section 2, we describe an analysis model for evaluating treatment effects of an ITT population by decomposing an ITT difference into its

observed and unobserved components. In section 3, we introduce the new methodology (GCC analysis) based on the observed treatment difference and evaluate the impact of the unobserved difference on the ITT analysis. We illustrate the application of this method using literature data from both parallel group and crossover randomized clinical trials in section 4, and summarize and discuss the properties and limitations of this method in section 5.

2 Statistical model of treatment effect

2.1 Sample mean distribution

Let n be the number of all randomized subjects who receive a given treatment in a clinical trial (the ITT set), and $n = n_o + n_m$ with n_o being the number of completers and n_m the number of non-completers with regard to the protocol-specified completion time.

Let $y_{o,i}$ be the outcome, a measure of the target symptom, of completer i ($i=1, \dots, n_o$) and $y_{m,j}$ the unobserved outcome of non-completer j ($j=1, \dots, n_m$). One can then decompose the ITT sample mean $\bar{Y}$ into the observed and unobserved components as shown below

$$\begin{aligned}\bar{Y} &= \frac{\sum_{i=1}^{n_o} y_{o,i} + \sum_{j=1}^{n_m} y_{m,j}}{n} = \frac{n_o}{n} \left(\frac{1}{n_o} \sum_{i=1}^{n_o} y_{o,i} \right) + \frac{n_m}{n} \left(\frac{1}{n_m} \sum_{j=1}^{n_m} y_{m,j} \right) \quad (1) \\ &= \bar{p}_o \bar{Y}_o + (1 - \bar{p}_o) \bar{Y}_m\end{aligned}$$

where $\bar{p}_o = \frac{n_o}{n}$ is the observed proportion of completers and $\bar{Y}_o = \frac{1}{n_o} \sum_{i=1}^{n_o} y_{o,i}$ and

$\bar{Y}_m = \frac{1}{n_m} \sum_{j=1}^{n_m} y_{m,j}$ the sample means of completers and non-completers (unobserved),

respectively. By the central limit theorem (CLT) for large n , $\bar{Y}$ follows a normal distribution,

$$\bar{Y} \sim N(\mu, \sigma^2/n) \quad (2)$$

where

$$\begin{aligned}\mu &= p_o \mu_o + (1 - p_o) \mu_m \\ \frac{\sigma^2}{n} &= \frac{1}{n} [p_o \sigma_o^2 + p_o (1 - p_o) \mu_o^2 + (1 - p_o) \sigma_m^2 + p_o (1 - p_o) \mu_m^2 + z \sigma_{om}] \\ \frac{\sigma_{om}}{n} &= \text{cov}(\bar{p}_o \bar{Y}_o, (1 - \bar{p}_o) \bar{Y}_m) = -\frac{1}{n} p_o (1 - p_o) \mu_o \mu_m\end{aligned}$$

The mean and variance are derived for large n and n_o from the conditional distribution

$$\bar{Y}_{(\cdot)} | n_o \sim N\left(\mu_{(\cdot)}, \frac{\sigma_{(\cdot)}^2}{n_{(\cdot)}}\right) \text{ for completers and non-completers, respectively, with } \mu_{(\cdot)} = E(\bar{Y}_{(\cdot)} | n_o) \text{ and } \frac{\sigma_{(\cdot)}^2}{n_o} = \text{Var}(\bar{Y}_{(\cdot)} | n_o) \text{ being unknown constants (independent of } n_o), \text{ and } n_o \sim$$

Binomial (n, p_o) with p_o being the probability of a randomized subject to be a completer (see the Supplementary Appendix A). Note that each component in equation (1) has the form of $\bar{p}\bar{Y}$, and both components likely follow the same kind of distribution as $\bar{Y}$ does.

2.2 Distribution of observed difference

We would like to find the asymptotic distribution of $\bar{p}_o \bar{Y}_o$. An explicit derivation shows that the moment generating function of $\bar{p}_o \bar{Y}_o$ (see the Supplementary Appendix B.1) is

$$M_{\bar{p}_o \bar{Y}_o}(t) \cong \exp \left[p_o \mu_o t + \frac{p_o(\sigma_o^2 + (1 - p_o)\mu_o^2)}{n} \frac{t^2}{2} + \frac{p_o(1 - p_o)}{2} \left(\frac{\mu_o \sigma_o^2 t^3}{n^2} + \frac{\sigma_o^4 t^4}{4n^3} \right) \right] \quad (3)$$

By ignoring the higher order terms, $O(n^{-k})$ with $k \geq 2$, we can see immediately that equation (3) approaches the moment generating function of a normal distribution,

$$M_{\bar{p}_o \bar{Y}_o}(t) \cong \exp \left[p_o \mu_o t + \frac{p_o(\sigma_o^2 + (1 - p_o)\mu_o^2)}{n} \frac{t^2}{2} \right],$$

suggesting

$$\sqrt{n}(\bar{p}_o \bar{Y}_o - p_o \mu_o) \sim N(0, p_o \sigma_o^2 + p_o(1 - p_o)\mu_o^2) \quad (4)$$

The asymptotic mean and variance of $\bar{p}_o \bar{Y}_o$ in equation (4) can also be obtained directly, using the same approach for deriving the mean and variation of $\bar{Y}$ in equation (2). A computer simulation provides further evidence supporting the asymptotic normal distribution of $\bar{p}_o \bar{Y}_o$ in equation (4) (see the Supplementary Appendix B.2). This distribution is identical to that of a weighted responder mean reported by Chen and Li.²¹

An unbiased estimator of $\text{Var}(\bar{p}_o \bar{Y}_o) = \sigma_{pY}^2 = \frac{1}{n}[p_o \sigma_o^2 + p_o(1 - p_o)\mu_o^2]$ is

$$\hat{\sigma}_{pY}^2 = \frac{1}{n} \left(\frac{n\bar{p}_o - 1}{n - 1} s_o^2 + \frac{n}{n - 1} \bar{p}_o(1 - \bar{p}_o) \bar{Y}_o^2 \right) \quad (5)$$

where s_o^2 is the (conditional) sample variance of y_i , $i = 1, \dots, n_o$ (see the Supplementary Appendix B.3). Both $\bar{Y}$ and $\bar{p}_o \bar{Y}_o$ follow asymptotically normal distributions, as does the distribution of $(1 - \bar{p}_o) \bar{Y}_m$, whose asymptotic mean $(1 - p_o) \mu_m$ and variance $\frac{1}{n}[(1 - p_o) \sigma_m^2 + p_o(1 - p_o) \mu_m^2]$ are defined similarly to those of $\bar{p}_o \bar{Y}_o$.

2.3 Treatment mean difference

Suppose that there are two treatments, a candidate treatment T and a comparator C , in a randomized parallel group study with sample sizes n_t and n_c , respectively. The number of completers is $n_{o,t}$ in the T arm and $n_{o,c}$ in the C arm. From equation (1), we can write the difference between the two treatments as

$$\bar{d} = \bar{Y}_t - \bar{Y}_c = \bar{d}_o + \bar{d}_m \quad (6)$$

where $\bar{d}_o$ is the observed difference and $\bar{d}_m$ the unobserved difference. The relationship in equation (6) is also true for a crossover study, which uses paired data of completers to estimate a treatment difference, where $n_t = n_c$ and $n_o = n_{o,t} = n_{o,c}$. Therefore, we write the observed difference as

$$\bar{d}_o = \begin{cases} \bar{p}_{o,t} \bar{Y}_{o,t} - \bar{p}_{o,c} \bar{Y}_{o,c} & \text{parallel group} \\ \frac{n_o}{n} \left[\frac{1}{n_o} \sum_{i=1}^{n_o} (y_{ot,i} - y_{oc,i}) \right] & \text{crossover} \end{cases} = \bar{p}_o \bar{Y}_{o,tc} \quad (7)$$

where $\bar{Y}_{o,tc} = \frac{1}{n_o} \sum_{i=1}^{n_o} (y_{ot,i} - y_{oc,i})$. The unobserved difference $\bar{d}_m$ can be expressed in a similar way. According to equations (2), (4), and (6), the asymptotic distributions of $\bar{d}$, $\bar{d}_o$, and $\bar{d}_m$ are normal.

1. Asymptotic distribution of $\bar{d}$

$$\bar{d} \sim N(d, \sigma_d^2) \quad (8)$$

where the expected mean $d = \mu_t - \mu_c = d_o + d_m$ and variance $\sigma_d^2 = \frac{\sigma_t^2}{n_t} + \frac{\sigma_c^2}{n_c}$

2. Asymptotic distribution of $\bar{d}_o$

$$\bar{d}_o \sim N(d_o, \sigma_o^2) \quad (9)$$

where

$$\begin{cases} d_o = p_{o,t} \mu_{o,t} - p_{o,c} \mu_{o,c} & \text{and} & \sigma_o^2 = \frac{\sigma_{o,t}^2}{n_t} + \frac{\sigma_{o,c}^2}{n_c} & \text{parallel group} \\ d_o = p_o \mu_{o,tc} & \text{and} & \sigma_o^2 = \frac{\sigma_{o,tc}^2}{n_t} & \text{crossover} \end{cases}$$

For the asymptotic distribution of $\bar{d}_m$, the mean d_m and variance σ_m^2 are defined similarly to equation (9). Since the $\bar{d}_m$ is not observed, our analysis can depend only on $\bar{d}_o$. We propose a GCC analysis based on $\bar{d}_o$ in the next section.

3 A GCC analysis

3.1 Test hypothesis

In a clinical trial, the null hypothesis of a statistical test is usually that the effect of treatment T is either less than or equal to that of treatment C :

$$\begin{aligned} H_0: \mu_t - \mu_c &\leq 0 && \text{for a superiority test, or} \\ H_0: \mu_t - \mu_c &\leq -M \text{ (or } H_0: \mu_c - \mu_t \geq M) && \text{for a non-inferiority test} \end{aligned}$$

where $M > 0$ is a pre-specified non-inferiority margin. We can combine the above two hypotheses as $H_0: \mu_t - \mu_c - \delta$ with $\delta = 0$ for a superiority test and $\delta = M$ for a non-inferiority test.

The null hypothesis of a GCC test is $H_0: d_o - \delta$. For simplicity and without loss of generality, let $\delta = 0$ for the remainder of this paper. A test of the hypothesis can be carried out as described in the following subsection.

3.2 Hypothesis test

Let $T_{d_o} = \frac{\bar{d}_o}{SE(\bar{d}_o)}$ be the test statistic of a GCC analysis where $SE(\bar{d}_o)$ is the standard error of $\bar{d}_o$. In a hypothesis test, the estimate of $SE(\bar{d}_o)$ is restricted under the null hypothesis, $SE(\bar{d}_o | H_0)$.

In a cross-over study, according to equation (7) and under the null hypothesis

$$H_0: d_o = E\left(\frac{1}{n} \sum_{i=1}^{n_o} (y_{ot,i} - y_{oc,i})\right) = 0, \text{ one can see immediately that } \bar{d}_o \sim N\left(0, \frac{p_o \sigma_o^2}{n}\right) \text{ and that}$$

the estimate of $SE(\bar{d}_o)$ is

$$SE(\bar{d}_o) = \frac{\sqrt{\bar{p}_o s_o^2}}{n} \quad (10)$$

In a parallel group study, the null hypothesis $H_0: d_o = E\left(\frac{1}{n_t} \sum_{i=1}^{n_{o,t}} y_{ot,i} - \frac{1}{n_c} \sum_{j=1}^{n_{o,c}} y_{ot,j}\right) = 0$

means that $p_{o,t} \mu_{o,t} = p_{o,c} \mu_{o,c}$. Under this restriction and equations (9) and (5), an un-pooled estimate of $SE(\bar{d}_o)$ is

$$SE(\bar{d}_o) = \sqrt{\frac{\hat{\sigma}_{o,t}^2}{n_t} + \frac{\hat{\sigma}_{o,c}^2}{n_c}} \quad (11)$$

$$= \sqrt{\frac{s_{o,t}^2}{n_t} \left(\frac{n_t \bar{p}_{o,t} - 1}{n_t - 1} \right) + \frac{s_{o,c}^2}{n_c} \left(\frac{n_c \bar{p}_{o,c} - 1}{n_c - 1} \right) + \left[\frac{1 - \bar{p}_{o,t}}{(n_t - 1) \bar{p}_{o,t}} + \frac{1 - \bar{p}_{o,c}}{(n_c - 1) \bar{p}_{o,c}} \right] \psi^2}$$

where ψ is a pooled estimate for $p_{o,t} \mu_{o,t} = p_{o,c} \mu_{o,c}$

$$\psi = \frac{1}{n_t + n_c} \left(\sum_{i=1}^{n_{o,t}} y_{ot,i} + \sum_{j=1}^{n_{o,c}} y_{ot,j} \right) = \frac{1}{n_t + n_c} (n_t \bar{p}_{o,t} \bar{Y}_{o,t} + n_c \bar{p}_{o,c} \bar{Y}_{o,c})$$

We also provide a pooled estimate of $SE(\bar{d}_o)$ as seen in Supplementary Appendix D.

3.3 Validation of the test

We are interested in the inference of the *GCC* analysis about treatment effects in the ITT population. Our approach is to investigate the impact of missing data on the ITT analysis, provided that the null hypothesis of a *GCC* test $H_0: d_o = 0$ is rejected at a pre-specified significance level of $\alpha/2$. Let the null hypothesis of an ITT analysis be $H_0: d = 0$, and let $SE(\bar{d})$ and $SE(\bar{d}_m)$ be the standard errors of $\bar{d}$ and $\bar{d}_m$, respectively.

There is a relationship among $\bar{d}$, $\bar{d}_o$, and $\bar{d}_m$ (see Supplementary Appendix C):

$$SE(\bar{d}) \leq SE(\bar{d}_o) + SE(\bar{d}_m) \quad (12)$$

Let $T_d = \frac{\bar{d}}{SE(\bar{d})}$ be the test statistic of the ITT analysis. Given that the *GCC* test is significant,

i.e. $T_{d_o} = z_{1-\alpha^*/2} > z_{1-\alpha/2}$ with $\alpha^* < \alpha$, we will reject the ITT null hypothesis $H_0: d = 0$ if

$$T_d = \frac{\bar{d}_o + \bar{d}_m}{SE(\bar{d})} \mid \left(\bar{d}_o = z_{1-\alpha^*/2} \cdot SE(\bar{d}_o) > z_{1-\alpha/2} SE(\bar{d}_o) \right) \geq \frac{z_{1-\alpha^*/2} \cdot SE(\bar{d}_o) + \bar{d}_m}{SE(\bar{d}_o) + SE(\bar{d}_m)} > z_{1-\alpha/2}$$

The last inequality holds if and only if

$$(z_{1-\alpha^*/2} - z_{1-\alpha/2}) \cdot SE(\bar{d}_o) + \bar{d}_m - z_{1-\alpha/2} \cdot SE(\bar{d}_m) > 0 \quad (13)$$

In the following discussion, assume that a larger value of the outcome represents a more favorable condition of a subject. Now we look into the impact of d_m on $d = d_o + d_m$ (see equation (8)) given $d_o > 0$.

Case 1: $d_m > 0$ —Note that the unobserved part in equation (13), $\bar{d}_m - z_{1-\alpha/2} \cdot SE(\bar{d}_m)$, is actually the lower limit of $\bar{d}_m$'s confidence interval at a significance level α , namely L_{dm} . The situation, $L_{dm} > 0$, indicates that equation (13) holds so one rejects the null hypothesis $H_0: d_m = 0$ at the significance level of $\alpha/2$. In this situation, the missing mechanism appears to be completely random.

Case 2: $d_m = 0$ —One can see that $L_{dm} = \bar{d}_m - z_{1-\alpha/2} \cdot SE(\bar{d}_m) = 0$ is equivalent to $T_{d_m} = \frac{\bar{d}_m}{SE(\bar{d}_m)} \leq z_{1-\alpha/2}$, suggesting that one does not reject the null hypothesis $H_0: d_m = 0$ at the significance level of $\alpha/2$.

Since $d = d_o + d_m$, the special case $d_m = 0$ implies that $d = d_o$ with $\bar{d}_o$ being an unbiased estimate of d . However, the inequality in equation (13) may not hold when $d_m = 0$, so it is possible that the ITT analysis would not reject the null hypothesis $H_0: d = 0$ at a significance level of $\alpha/2$. This is because the presence of the error term, $SE(\bar{d}_m) > 0$, in T_d increases the type II error rate for not rejecting the null hypothesis, given the alternative hypothesis is true, $d = d_o > 0$. Therefore, the GCC test is more powerful than the ITT test when $d_m = 0$. If there are no overwhelming concerns about treatment safety, then the missing mechanism is likely due to a lack of efficacy.

Case 3: $d_m < 0$ —According to equation (8), the circumstance $d_m < 0$ states that $(1 - p_{o,d})\mu_{m,t} < (1 - p_{o,c})\mu_{m,c}$. If $\mu_{m,t} = \mu_{m,c}$, then the dropout rate in the T arm must be significantly lower than that in the C arm, $(1 - p_{o,d}) < (1 - p_{o,c})$. In this situation, we will have no serious concerns about the favorable effect of T over C , even if the ITT analysis is not significant according to equation (13). This is because:

- there is a significantly favorable treatment effect for completers, $d_o > 0$;
- the mean outcome of non-completers in the T arm is equal to or greater than that in the C arm, $\mu_{m,t} \geq \mu_{m,c}$;
- the dropout rate in the T arm is significantly lower than that in the comparator arm, $(1 - p_{o,d}) < (1 - p_{o,c})$

The above reasons support the inference of the GCC analysis about the significant treatment effect in the ITT population. Again, the GCC analysis here is more powerful than an ITT analysis. The missing mechanism is likely to be completely random.

Now we would like to find out what the situation $\mu_{m,t} < \mu_{m,c}$ tells us, given $d_o > 0$ and $d_m < 0$. Although equation (13) may still hold in certain situations especially when $\alpha^* \ll \alpha$, here we

need only to consider the case of $d = 0$. In our view, such a scenario demonstrates that, relative to the comparator, the treatment T significantly alleviates the target symptom of the completers ($d_o > 0$) but severely aggravates that of the non-completers ($\mu_{m,t} < \mu_{m,c}$). This is because a small value of the outcome reflects a subject's unfavorable condition, and a significantly worse target symptom could lead to or be associated with serious medical consequences, either directly or indirectly. Therefore, although the outcomes of non-completers for estimating $\mu_{m,t}$ and $\mu_{m,c}$ are unknown, their medical information would be critical for providing hints concerning whether the non-completers in the T arm are under more unfavorable conditions than those in the C arm, i.e. $\mu_{m,t} < \mu_{m,c}$. In fact, it is not uncommon that subjects whose symptoms are exacerbated during a study would receive rescue medication or additional care. Since this is the only case that a significant GCC result does not imply significant treatment effect in the ITT population, we will evaluate whether the candidate treatment T is poisonous to the non-completers in the analysis of ITT treatment effect.

The above discussions suggest that a statistically significant GCC result implies a significant treatment effect in the ITT population at a pre-specified significance level unless, relative to the comparator, the candidate treatment T is toxic to the non-completers. In addition, equation (13) suggests that rejecting the null hypothesis of the GCC analysis at a high significance level $\alpha^*/2$ with $\alpha^* \ll \alpha$ is a favorable condition for rejecting the corresponding null hypothesis of the ITT analysis. A GCC analysis does not inflate the type I error of an ITT analysis and is at least as powerful as the ITT analysis in certain circumstances, as discussed above and shown by computer simulation (see Supplementary Appendix E).

4 Demonstration of GCC analysis in clinical trials

The application of GCC analysis to clinical trials is demonstrated using literature data. All selected studies are randomized controlled clinical trials. In our GCC analysis, we will focus on checking if the candidate drug is poisonous to the non-completers as documented in their medical records.

4.1 Application to crossover study

A crossover design has two advantages over both a parallel study and a non-crossover longitudinal study: reduced confounding covariates and smaller sample sizes. Thus this design is widely used in neurology and psychiatry studies and pain treatment trials. The statistical analysis of cross-over trials is typically a completers analysis. The data demonstrating the GCC analysis are from a drug abuse potential study by Schoedel et al.²² which was a single-site, randomized, double-blind, double-dummy, single-dose, active-and placebo-controlled, crossover study.

The primary endpoint was the peak effect ($E_{\max}$) of drug liking (at the moment), assessed on a bipolar visual analog scale (VAS) from 0 (strong disliking) to 100 (strong liking) with 50 as the neutral score. For assay sensitivity, the null hypothesis was no difference between the positive control D-amphetamine and placebo. According to the standard study design, all subjects were expected to like the positive control (evidenced by larger mean VAS $E_{\max}$) more than the placebo, which would have been guaranteed by a prior qualification phase.

The GCC test statistic is constructed based on equation (10) and the analysis is significant at a level of 0.05. Table 1 summarizes the results of the reported completers analysis and the GCC test.

As regards the safety information, the authors reported “no serious or significant AEs among the subjects” in the treatment phase and comparable AE rates between D-amphetamine and placebo. Based on these results, we are confident at a significance level of 0.05 that these subjects like D-amphetamine more than the placebo at either dose in the ITT population, supporting the assay sensitivity.

4.2 Application to parallel-group trial

A parallel group design is the most common clinical trial design for confirmatory trials. Data from a noninferiority study reported by Gallwitz et al.²³ are used here for illustration. The study was a multicenter, randomized, two-year, parallel-group, non-inferiority, double-blind clinical trial for type 2 diabetes mellitus. Participants were randomly assigned to take either the test drug, linagliptin (5 mg), or the active comparator, glimepiride (1–4 mg) orally once daily and treated for two years (104 weeks). The primary endpoint was change in glycated hemoglobin A1C (HbA_{1c}) from baseline to week 104. After 104 weeks of treatment, the number of completers was 447 for linagliptin and 458 for glimepiride. The missing rates of the two treatments were similar: 42% in the linagliptin arm and 41% in the glimepiride arm.

The GCC test statistic is constructed using the un-pooled standard error as shown in equation (11) for parallel designed studies. The non-inferiority of linagliptin (5 mg) to glimepiride (1–4 mg) is supported by the GCC analysis (shown in Table 2) since the upper bound of the 97.5% confidence interval (Lina-Glim), 0.19%, is smaller than the pre-selected non-inferiority margin of 0.35%. With respect to the safety information from this publication, subjects in the linagliptin arm had fewer AEs than those in the glimepiride arm:

- fewer drug-related AEs: 118 (15%) versus 300 (39%);
- fewer AEs in the non-completers: 61 out of 189 (32%) versus 90 out of 171 (52%);
- fewer AEs leading to discontinuation: 60 (8%) versus 85 (11%).

The GCC result and above safety information support the non-inferiority of linagliptin over glimepiride in the ITT population at a significance level of 0.025. Table 2 summarizes the efficacy results from the authors and the GCC analysis.

5 Summary and discussion

Applying ITT analysis to randomized controlled clinical trials in practice has been challenged by the presence of missing data. This problem is complicated by the unknown consequences of stopping an assigned treatment in a withdrawn subject. Making no assumptions about the missing mechanisms, we proposed a GCC analysis to deal with the ITT puzzle I. The GCC analysis is a *de facto* natured analysis and depends on the observed data in the ITT population. If all randomized subjects are completers, then the GCC analysis becomes an ITT analysis. Taking into consideration that the number of subjects who drop

out from a treatment arm is a random variable, the distribution of the observed difference was derived explicitly as the base for hypothesis testing.

By evaluating the impact of the unobserved difference on the ITT analysis, we showed that a statistically significant GCC result implies a significant treatment effect in the ITT population at a pre-specified significance level unless the candidate drug is toxic to the non-completers, who likely have many more drug-related AEs than the non-completers in the comparator arm do. A GCC analysis does not inflate the type I error of an ITT analysis and is at least as powerful as the ITT analysis in certain circumstances. In addition, if collected, the values of the outcome data in the non-completers prior to dropouts from the test drug arm should be smaller (unfavorable conditions) than those in the comparator's arm. If the test drug is beneficial to some subjects (who are more likely to adhere to treatment) but poisonous to some others (who are more likely to dropout) and the null hypothesis of the GCC test is rejected, then one can only claim the efficacy of the test drug in the completers but not in the ITT population. Should such a case occur, an enriched design may be desirable.

Although the GCC analysis is useful to qualitatively tell the treatment effect in the ITT population (the ITT puzzle I), in general, this method, like all other available statistical methods, is not able to provide an unbiased estimate of the true treatment difference in the ITT population with missing data (the ITT puzzle II), except that the expected treatment difference in the non-completers is zero. This assumption is less restrictive than that of missing at random because the basic component involved in the unobserved treatment difference is the product of the (conditional) unobserved outcome mean and the observed dropout rate. Various techniques used to fill the missing values might play a role in the verification of such an assumption.

Note that the p value of a GCC test whose null hypothesis $H_0: d_o = 0$ has been rejected should not be claimed to be the significance level of the treatment difference in the ITT population, provided that there is no clinical evidence of the test drug being poisonous to the non-completers. For example, if a pre-determined level of significance is 0.025, the GCC test is significant at a level of <0.0001 , and there is no clinical evidence of the drug candidate being poisonous to the non-completers, then one can confidently claim that the treatment difference between the test drug and the comparator in the ITT population is significant at the level of 0.025. One should also be aware that a significant GCC analysis at a pre-determined level could imply the significance of the treatment difference in an ITT population, not the ITT test itself, because in certain circumstances (for example, when $d_m = 0$), a GCC test could be more powerful than an ITT analysis.

The GCC analysis is demonstrated in this paper using literature data from both parallel group and crossover randomized clinical trials for either a superiority or a non-inferiority test. Given that the GCC result is significant, the safety information of the non-completers play a critical role for inferring treatment effects in the ITT population, suggesting that the analyses of the benefit and risk effects of a test drug are nested together. Moreover, a possible scenario is that the ITT analysis is significant but the GCC analysis is not, implying that the beneficial effect of the test drug is accompanied by some adverse effects in the non-

completers. Such a drug is unlikely to be safe to use. In clinical trials, missing data reveal the information of patients' unfavorable responses to treatments. Therefore, when we are considering the issue of missingness while evaluating treatment effects, we are eventually dealing with the benefit–risk problem. The GCC analysis is unique in that it is the first efficacy analysis method that uses safety information to qualitatively tell if the difference between treatments in an ITT population with missing data would be significant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge Drs Yi Tsong, Thomas Permutt, Hsien Ming J Hung, Kallappa Koti, Ram Tiwari, Ling Chen, and Yan Zhou for either giving their critical comments and discussions on an earlier version of the manuscript or providing useful information. The authors are also grateful to the two anonymous referees for their insightful comments. This work utilized the Biowulf clusters, high-performance computing resources at the National Institutes of Health, Bethesda, MD (<http://hpc.nih.gov>).

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials*. 2000; 21:167–189. [PubMed: 10822117]
2. Fisher, L., Dixon, D., Herson, J., et al. Intent to treat in clinical trials. In: Peace, KE., editor. *Statistical issues in drug research and development*. New York, NY: Mercel Dekker, Inc; 1990. p. 331–350.
3. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011; 2:109. [PubMed: 21897887]
4. O'Neill R, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clin Pharmacol Ther*. 2012; 91:550–554. [PubMed: 22318615]
5. Pocock SJ, Abdalla M. The hope and the hazards of using compliance data in randomized controlled trials. *Stat Med*. 1998; 17:303–317. [PubMed: 9493256]
6. Armijo-Olivo S, Warren S, Magee D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. *Phys Ther Rev*. 2009; 14:36–49.
7. Enders, CK. *Applied missing data analysis*. New York, NY: Guilford Publications; 2010.
8. Carpenter JR, Kenward MG, Vansteelandt S. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *J R Stat Soc Series A (Stat Soc)*. 2006; 169:571–584.
9. Donders ART, van der Heijden GJ, Stijnen T, et al. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006; 59:1087–1091. [PubMed: 16980149]
10. Bombardier C, Tugwell P. Controversies in the analysis of long term clinical trials of slow acting drugs. *J Rheumatol*. 1985; 12:403–405. [PubMed: 4045839]
11. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–592.
12. Dziura JD, Post LA, Zhao Q, et al. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med*. 2013; 86:343–358. [PubMed: 24058309]
13. National Research Council. *The prevention and treatment of missing data in clinical trials*. Washington, DC: National Academies Press; 2010.

14. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012; 367:1355–1360. [PubMed: 23034025]
15. Kitano H. Systems biology: a brief overview. *Science*. 2002; 295:1662–1664. [PubMed: 11872829]
16. Guimera R, Amaral LAN. Functional cartography of complex metabolic networks. *Nature*. 2005; 433:895–900. [PubMed: 15729348]
17. Gibson, GG., Skett, P. Introduction to drug metabolism. Cheltenham, UK: Nelson Thornes; 2001.
18. Fry DE. Sepsis, systemic inflammatory response, and multiple organ dysfunction: the mystery continues. *Am Surg*. 2012; 78:1–8. [PubMed: 22273282]
19. Hotchkiss RS, Swanson PE, Freeman BD, et al. Apoptotic cell death in patients with sepsis, shock, and multiple organ dysfunction. *Crit Care Med*. 1999; 27:1230–1251. [PubMed: 10446814]
20. Tilghman SL, Nierth-Simpson EN, Wallace R, et al. Environmental hormones: multiple pathways for response may lead to multiple disease outcomes. *Steroids*. 2010; 75:520–523. [PubMed: 20466011]
21. Chen L, Li QH. Assessing clinically relevant treatment differences using weighted responder means. *Model Assist Stat Appl*. 2010; 5:159–167.
22. Schoedel K, Meier D, Chakraborty B, et al. Subjective and objective effects of the novel triple reuptake inhibitor tesofensine in recreational stimulant users. *Clin Pharmacol Ther*. 2010; 88:69–78. [PubMed: 20520602]
23. Gallwitz B, Rosenstock J, Rauch T, et al. 2-year efficacy and safety of linagliptin compared with glimepiride in patients with type 2 diabetes inadequately controlled on metformin: a randomised, double-blind, non-inferiority trial. *The Lancet*. 2012; 380:475–483.

Table 1

Comparison of differences in peak effects of drug liking visual analog scale from a drug abuse potential study.^a

n_0/n^b	Test contrast	Diff ^c (95% CI) <i>p</i> value	GCC diff (95% CI) <i>p</i> value
44/52	D-Amphetamine 15 mg-placebo	8.8 (2.3, 15.3) <0.01	7.4 (1.6, 13.3) 0.01217
	D-Amphetamine 30 mg-placebo	23.9 (17.4, 30.3) <0.001	20.2 (14.5, 26.0) <0.001

^aData from Schoedel et al.²²

^bNote: n and n_0 denote the number of randomized patients and completers, respectively.

^cThe differences between the two treatments were analyzed using a mixed model.

Table 2

Summary of efficacy results from a parallel group study.^a

Method	Drug, T (n_o/n or n_T) ^b	Control, C (n_o/n or n_T) ^b	Mean (SE), T	Mean (SE), C	T - C (97.5% CI)	p value
GCC	Linagliptin (447/777)	Glimepiride (458/775)	-0.35 (0.04)	-0.53 (0.04)	0.11 (0.04, 0.18)	0.0013
LOCF ^c	Linagliptin (764)	Glimepiride (755)	-0.16 (0.03)	-0.36 (0.03)	0.20 (0.09, 0.30)	0.0004

^aData from Gallwitz et al.²³

^b n and n_o denote the number of randomized patients and completers, respectively; n_T : number of randomized subjects who received at least one dose of treatment, had a baseline measurement, and had at least one on-treatment measurement.

^cLOCF: last observation carried forward.