Research Article

Statistics in Medicine

Received 31 December 2013, Accepted 9 October 2014 Published online 3 November 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6352

Intention-to-treat analysis with treatment discontinuation and missing data in clinical trials

Roderick Little*† and Shan Kang

Motivated by a recent National Research Council study, we discuss three aspects of the analysis of clinical trials when participants prematurely discontinue treatments. First, we distinguish treatment discontinuation from missing outcome data. Data collection is often stopped after treatment discontinuation, but outcome data could be recorded on individuals after they discontinue treatment, as the National Research Council study recommends. Conversely, outcome data may be missing for individuals who do not discontinue treatment, as when there is loss to follow up or missed clinic visits. Missing outcome data is a standard missing data problem, but treatment discontinuation is better viewed as a form of noncompliance and treated using ideas from the causal literature on noncompliance. Second, the standard intention to treat estimand, the average effect of randomization to treatment, is compared with three alternative estimands for the intention to treat population: the average effect when individuals continue on the assigned treatment after discontinuation, the average effect when individuals take a control treatment after treatment discontinuation, and a summary measure of the effect of treatment prior to discontinuation. We argue that the latter choice of estimand has advantages and should receive more consideration. Third, we consider when follow-up measures after discontinuation are needed for valid measures of treatment effects. The answer depends on the choice of primary estimand and the plausibility of assumptions needed to address the missing data. Ideas are motivated and illustrated by a reanalysis of a past study of inhaled insulin treatments for diabetes, sponsored by Eli Lilly. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: clinical trials; dropouts; incomplete data; intention-to-treat analysis; missing data; treatment discontinuation

1. Introduction

The analysis of randomized clinical trials for comparing treatments is straightforward when all participants take their assigned treatments and have outcomes recorded. Summary measures such as means can be compared to assess treatment effects, and the randomization protects the comparison from measured and unmeasured confounders of treatment differences.

The analysis and interpretation is complicated when individuals discontinue their assigned treatments prematurely, and/or outcome measures are not recorded. It is important to distinguish between these two issues [1]. Treatment discontinuation leads to missing outcome data when a study chooses not to follow up participants who discontinue treatment. However, outcome data can be recorded for individuals who discontinue their assigned treatment, and participants who remain on their assigned treatment can have missing outcomes, as when they miss clinic visits.

This article has three main objectives. The first is to conceptualize treatment discontinuation as a form of noncompliance and apply ideas from the so-called Rubin model of causal inference [2, 3]. Specifically, observed compliance to a treatment is a post-treatment variable, so simply restricting comparisons of treatments to those who comply, as in 'per-protocol analysis', is flawed because observed compliance is a consequence of the treatment. Angrist, Imbens, and Rubin [4] defined strata based on whether participants would comply with each of the treatments being compared and then defined the 'complier-average causal

Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A.
**Correspondence to: Roderick Little, Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A.*
†E-mail: rlittle@umich.edu

effect' (CACE) to be the treatment effect within the stratum of individuals who would comply with all the treatments under comparison. This is a valid causal effect, but estimating it involves a missing data problem, in that compliance status is only known for the treatment actually assigned to a participant. In a seminal paper [5], Frangakis and Rubin generalize these ideas to post-treatment variables other than compliance, by defining 'principal strata' based on values of these post-treatment variables under all the treatments being compared. Thus, for the particular case of compliance, 'principal compliers' are defined as individuals who would comply for any of the compared treatments, if assigned to them. For other discussions, see [6–8]

Analogously, in the setting of treatment discontinuation, we define 'principle discontinuation strata' based on whether individuals would discontinue under each of the treatments being compared. Because we only get to observe discontinuation for the treatment actually assigned, this leads to a problem of missing data in this covariate, as distinct from missing data in the outcome arising when outcomes are not measured after discontinuation.

Our second goal concerns the following recommendation of a recent National Research Council report on missing data in clinical trials [9], which emphasizes the central role of the choice of causal estimand in clinical trial inference with treatment discontinuation or analysis dropouts:

> "The trial protocol should explicitly define (a) the objective(s) of the trial; (b) the associated primary outcome or outcomes; (c) how, when, and on whom the outcome or outcomes will be measured; and (d) the measures of intervention effects, that is, the causal estimands of primary interest. These measures should be meaningful for all study participants, and estimable with minimal assumptions. Concerning the latter, the protocol should address the potential impact and treatment of missing data."

We agree with the implication of this recommendation that current approaches to handling missing data in clinical trials often fail to define the causal estimand clearly. Protocols propose methods of intention to treat (ITT) analysis or per-protocol analysis, together with weighting, imputation, or maximum likelihood methods to handle missing data, but the underlying estimand is often not explicitly stated. Expanding on this issue, we describe some possible estimands for the ITT population and suggest that the methods for handling treatment discontinuation and analysis dropout depend crucially on the choice of estimand; a method that is appropriate for one estimand is not necessarily appropriate for another.

Our third goal is to provide a nuanced discussion of whether follow-up measures should be attempted for participants who discontinue treatment. The aforementioned National Research Council report [9] recommends that

> "Trial sponsors should continue to collect information on key outcomes on participants who discontinue their protocol-specified intervention in the course of the study, except in those cases for which a compelling cost-benefit analysis argues otherwise, and this information should be recorded and used in the analysis."

While we generally concur with this recommendation, we suggest that the necessity to record outcomes after treatment discontinuation is not universal but depends on the choice of primary causal estimand, and the need and plausibility of assumptions about missing data required to estimate it. This reinforces the need to define the causal estimand and associated methods of analysis carefully.

The following clinical trial, which had substantial missing data and a complex mix of missing-data issues, serves to motivate and illustrate our ideas:

Example: a clinical trial of inhaled insulin treatments for diabetes.

Eli Lilly and Company conducted a study that assessed the efficacy and safety of an inhaled antihyperglycemic medication for patients with type 2 diabetes, compared with the standard therapies based on injected insulin Glargine. This randomized, parallel-group study recruited individuals experiencing lack of control of glucose levels, as measured by the HbA1c laboratory test. There were three treatment arms: the inhaled insulin (Inhaled) arm (n = 222) replaced once-daily insulin Glargine with mealtime inhaled insulin; the Insulin Glargine (IG) arm (n = 223), continued insulin Glargine on an intensified regimen; and a Combined arm (n = 115), which combined oncedaily insulin Glargine with mealtime inhaled insulin. Major objectives were the following:

(a) To test for noninferiority of Inhaled compared with IG;
(b) To test superiority of Combined compared with IG; and
(c) To test superiority of Inhaled compared with IG.

Table I. Inhaled insulin study: discontinuation and missing data in study groups by reason.

Type	Missing/discontinuation status and reason	Combined	IG no rescue	IG rescue	Inhaled no rescue	Inhaled rescue	All
0	Completed	25	47	1	24	18	115
1	Subject decision	10	12	0	30	3	55
2	Physician decision	5	3	0	6	1	15
3	Protocol violation	1	4	0	3	0	8
4	Adverse event	2	0	0	4	0	6
5	Death	1	0	0	1	0	3
6	Sponsor decision to Terminate trial	68	137	9	94	25	333
7	Lost to follow up	3	10	0	13	0	26

IG, Insulin Glargine.

The primary outcome was mean change in HbA1c from baseline to 24 weeks. We analyze here a secondary outcome, change from baseline to 52 weeks, to emphasize missing data issues. Superiority was concluded if the upper limit of the 95% confidence interval for a specified treatment difference was less than zero. Non-inferiority was concluded if this upper limit was less than 0.4%, but greater than or equal to 0.0%.

Measures of HbA1c were obtained at baseline and after 4, 12, 24, 38, and 52 weeks. The primary statistical analysis specified in the protocol was by ITT, including all randomized patients with a baseline and at least one follow-up measure after baseline. Adjustments for missing data were limited to last observation carried forward (LOCF) for missing endpoints when earlier observations on treatment were available.

Rescue treatments were specified in the protocol for individuals whose diabetes was not well controlled in the Inhaled and IG arms. In the IG group, the rescue therapy consisted of a pre-prandial dose of insulin. In the Inhaled group, the rescue medication consisted of the addition of Glargine. No rescue therapy was specified for individuals in the Combined group.

Only 115 of the 560 individuals in the study completed it, mainly because of a sponsor decision to terminate the trial early. Table I shows the distribution of discontinuation types and missing data in the three treatment groups, classified by reason. Of the 48 completers in the IG group, one received rescue therapy, and of the 42 completers in the Inhaled group, 18 received rescue therapy. Follow-up measures were attempted for individuals who took rescue treatments, but in other situations where individuals completely stopped the study, no outcome measures were recorded after discontinuation.

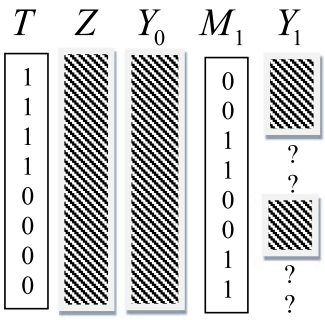
Table I distinguishes five classes of individuals who discontinue treatment because of their outcomes or side effects (Types 1–5), and two classes of missing data that are unrelated to an individual-level decision to discontinue treatment (Types 6 and 7). The most prevalent form of missing data was the sponsor decision to terminate the study (Type 6), which is considered missing data because it was a global decision unrelated to treatment and outcomes of specific participants. This kind of missing data is often called administrative censoring. The treatment discontinuation types 1–5 are likely to be related to side effects or lack of efficacy in controlling HbA1c levels, and this distinction influences the treatment of missing data, as discussed further in the succeeding text. Treatment discontinuation prior to week 52 also leads to missing data for the outcome change in HbA1c levels between week 52 and baseline. We discuss in the succeeding text the chosen method of imputation, LOCF, and propose an alternative choice of outcome measure that avoids the need to impute measures after discontinuation.

The remainder of the article is organized as follows. In the next section, we formalize the distinction between treatment discontinuation and missing outcome data, in a simple setting. In Section 3, we discuss four alternative estimands for the ITT population when there is treatment discontinuation, and in Section 4 apply some of these alternatives to the Insulin trial data. Section 5 presents conclusions and discussion.

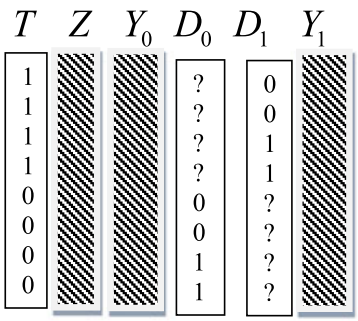
2. Treatment discontinuation and missing outcomes: two distinct missing data problems

For notational simplicity, we consider in this section the comparison of two treatments, $T = 1 =$ new, $T = 0 =$ control. Let Y represent an outcome measure, and let Y_0 denote a baseline measure of Y and Z be other

A. Missing Outcome Data, No Treatment Discontinuation. ? Denotes Missing Values



B. Treatment Discontinuation, Discontinuers Followed . ? Denotes Missing Values



C. Treatment Discontinuation, Discontinuers Not Followed. ? Denotes Missing Values

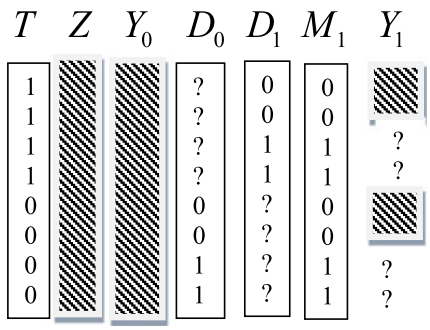


Figure 1. Missing outcomes and treatment discontinuation.

baseline covariates, both of which we assume to be observed for all participants. Let Y_1 denote the trial outcome measure, which could be the change in Y from the baseline value. (In the inhaled insulin example there were also intermediate measures between baseline and end point.) We use a different notation to distinguish treatment discontinuation and missing outcome data. Let D_t denote discontinuation for treatment t, taking value 1 if an individual discontinues treatment t and 0 otherwise, and let M_1 denote the missing data indicator forY_1 , with value 1 if Y_1 is missing and 0 otherwise.

With missing outcome data but no treatment discontinuation, the data are depicted in Figure 1A. Well-studied ideas of missing data in longitudinal studies apply, as discussed in chapters 4 and 5 of [9] or [10]. A key assumption of many methods is missing at random (MAR) [10, 11], which in our setting, corresponds to assuming that the distribution of M_1 depends on the data only through the observed variables Z and Y_0, that is

$$\left[M_1|Y_0, Y_1, T, Z\right] = \left[M_1|Y_0, T, Z\right], \tag{1}$$

where [] denotes distribution. If Y_1 is assigned a distribution, then (1) implies that M_1 and Y_1 are independent given Y_0, T, Z, so it follows that an equivalent assumption is

$$\left[Y_1|Y_0, T, Z, M_1 = 1\right] = \left[Y_1|Y_0, T, Z, M_1 = 0\right], \tag{2}$$

Table II. Classifications by treatment and principal compliance: (A) population proportions; (B) population mean outcomes; (C) observed means, with discontinuers followed; (D) observed means, with discontinuers not followed.

		Treatment discontinuation principal stratum						
		$D_1=0$		$D_1=1$		$D_1=0$ or 1		
		$D_0=0$	$D_0=1$	$D_0=0$	$D_0=1$	$D_0=0$	$D_0=1$	ALL
A. Population proportions								
Assigned treatment T	1	$\alpha_1\pi_{00}$	$\alpha_1\pi_{01}$	$\alpha_1\pi_{10}$	$\alpha_1\pi_{11}$	$\alpha_1\pi_{+0}$	$\alpha_1\pi_{+1}$	α_1
	0	$\alpha_0\pi_{00}$	$\alpha_0\pi_{01}$	$\alpha_0\pi_{10}$	$\alpha_0\pi_{11}$	$\alpha_0\pi_{+0}$	$\alpha_0\pi_{+1}$	α_0
	ALL	π_{00}	π_{01}	π_{10}	π_{11}	π_{+0}	π_{+1}	1
B. Population means								
Assigned treatment T	1	$\mu_1^{(00)}$	$\mu_1^{(01)}$	$\mu_1^{(10)}$	$\mu_1^{(11)}$	$\mu_1^{(+0)}$	$\mu_1^{(+1)}$	μ_1
	0	$\mu_0^{(00)}$	$\mu_0^{(01)}$	$\mu_0^{(10)}$	$\mu_0^{(11)}$	$\mu_0^{(+0)}$	$\mu_0^{(+1)}$	μ_0
C. Observed means, discontinuers followed								
Assigned treatment T	1	$\bar{y}_1^{(0+)}$		$\bar{y}_1^{(1+)}$		—	—	$\bar{y}_1$
	0	—	—	—	—	$\bar{y}_0^{(+0)}$	$\bar{y}_0^{(+1)}$	$\bar{y}_0$
D. Observed means, discontinuers not followed								
Assigned treatment T	1	$\bar{y}_1^{(0+)}$		—		—	—	—
	0	—	—	—	—	$\bar{y}_0^{(+0)}$	—	—

that is, the distribution of Y_1 given Z, Y_0 and T is the same for participants with Y_1 missing and participants with Y_1 observed. Under this assumption, the predictive distribution of the missing values can be estimated using the observed data, or weighted estimates can be developed [9]. If MAR is not plausible, and there are no other covariates that characterize the difference between respondents and nonrespondents, there is no information in the data to estimate differences in the distribution $\left[Y_1|Y_0,T,Z,M_1=1\right]$ from the distribution $[Y_1|Y_0,T,Z,M_1=0]$, and the National Research Council report generally advocates a sensitivity analysis.

Our focus here is on treatment discontinuation. Following the principal stratification literature of noncompliance [4, 5, 7, 8, 12, 13], we define principal discontinuation under treatment $T=t$, D_t, taking the value 1 if an individual would discontinue if assigned treatment t, and 0 otherwise ($t=0$ and 1). The variable D_t is a principal stratifier, unaffected by the treatment actually assigned and is observed for the treatment that is actually assigned, but is missing for the treatment that is not assigned. In this formulation, treatment discontinuation leads to missing values in the covariate, principal discontinuation. The data are illustrated graphically in Figure 1B, for situations where outcomes for discontinuers are recorded, and Figure 1C, for situations where outcomes for discontinuers are not recorded. Technically, the mechanism is MAR, because missingness depends only on the known covariate T. However, there are problems of identification, because values of D_t are entirely missing for the treatment not assigned. Clearly, there are also mixed situations not depicted in Figure 1, where outcomes are partly recorded and partly missing, both for discontinuers and continuers.

Table II provides an alternative tabular depiction of the data and defines notation. The row classification is the assigned treatment T, and the column classification is the four combinations of principal treatment discontinuation for the two treatments. Table IIA defines the population proportions, and Table IIB defines the population mean outcomes, in each of the cells. Table IIC indicates the observed sample means when discontinuers are followed, and Table IID indicates the observed sample means when discontinuers are not followed.

3. Alternative estimands for the intention-to-treat population

The International Conference on Harmonization E9 [14] defines ITT as follows:

> "The principle that asserts that the effect of a treatment policy can be best assessed by evaluating on the basis of the intention to treat a subject (i.e. the planned treatment regimen) rather than the actual treatment given. It has

the consequence that participants allocated to a treatment group should be followed up, assessed and analyzed as members of that group irrespective of their compliance to the planned course of treatment."

This definition has two aspects, (1) the population for which the inference is defined, and (2) the choice of estimand measuring the treatment effect in that population. The ITT population is all randomized participants, with individuals classified according to the treatment randomized as opposed to (say) the treatment actually received [15]. Under a strict interpretation of the ICH definition, the ITT estimand is a summary of the *effect of randomization to treatment* (ERT), such as the difference in means in the last column of Table IIB:

$$\delta_{\text{ERT}} = \mu_1 - \mu_0, \tag{3}$$

which averages over principal discontinuation strata, that is, ignores discontinuation status. Carpenter, Roger, and Kenward [16] call the ERT the *de facto* estimand. When all discontinuers are followed up and outcomes Y_1 obtained, as in Table IIC, δ_{ERT} can be directly estimated by the sample means of Y_1 in the two treatment groups:

$$\hat{\delta}_{\text{ERT}} = \bar{y}_1 - \bar{y}_0. \tag{4}$$

On the other hand, when outcomes of discontinuers are not obtained, as in Table IID, δ_{ERT} is no longer directly estimable and requires assumptions about the distributions of final outcomes of discontinuers in the two treatment groups. Obviously, if the ERT is of primary interest, a direct estimate is only available if outcomes are recorded for participants who discontinue treatment. This is a formal justification of the previously-cited National Research Council recommendation [9] to seek to measure outcomes for individuals who discontinue treatment. Indeed, if no discontinuers are followed up, the only empirical information for imputation of outcomes of discontinuers lies in the outcomes of continuers, so strong and perhaps dubious assumptions are required to estimate the ERT.

A key feature of ERT is that it incorporates the effects of any treatments the participants received between discontinuation of their assigned treatment and the final measurement Y_1. An advantage of this feature is that it is arguably addressing a realistic scenario. The disadvantage is that the treatment effect may include the effects of treatments other than the treatment under study, so the study is in effect assessing a 'treatment regimen' that involves these other treatments. In many studies, the main interest is not a treatment regimen but rather in the particular effects of the new treatment.

The ERT is not the only possible estimand that can be applied to the ITT population. Alternatives define and estimate an ITT estimand for the counterfactual situation in which treatments other than the actual treatment are given after discontinuation. Examples include the estimand if discontinuers had remained on their assigned treatments [estimand under assigned treatment (EAT); this corresponds to what [16] call the *de jure* estimand]; or the estimand if discontinuers had taken a control treatment (Estimand under Control Treatment, ECT). This approach requires a method for imputing the outcomes of discontinuers *whether or not outcomes of discontinuers are recorded*, since, at least for some, possibly unknown, set of participants, the outcomes are counterfactual; on the other hand, observing the actual outcomes of discontinuers may provide useful input into the imputation model. The utility of this approach ultimately rests on the underlying scientific plausibility of the estimand and associated method for dealing with missing data. For example, the EAT may not be a plausible estimand if individuals drop out of the treatment for dose-related adverse events. In a study on treatments for Alzheimer disease, Little and Yau [17] estimate the EAT and ECT as part of a sensitivity analysis.

Can the EAT and ECT be considered ITT estimands? They are defined for all participants according to the group to which they are randomized, but they do not meet the aforementioned ITT principle in the sense that not all participants are 'followed up and assessed'. However, in practice, many clinical studies fail to follow up and record outcomes for all participants and still claim to do an ITT analysis; so strict application of the principle is often an unattainable ideal in practice. Semantics aside, in situations where outcomes are (implicitly or explicitly) imputed for discontinuers, it seems important to be clear about their assumed treatment post-discontinuation—the actual treatment received (as in ERT), the assigned treatment (as in EAT), or some other treatment, such as the control treatment (as in ECT).

If data after discontinuation are missing, how plausible are alternative imputation methods for these various estimands? We consider the situation depicted in Table II. Methods based on the MAR assumption assume that

$$[Y_1|Y_0, Z, T, D_1 = 1] = [Y_1|Y_0, Z, T, D_1 = 0],$$

which may be plausible if the target estimand is the EAT but seems unjustified if the target estimand is the ERT or ECT, since imputing the outcomes of discontinuers based on continuers seems unjustified when the nature of treatment after discontinuation has changed [17]. Both LOCF and BOCF impute $Y_1 = Y_0$ for discontinuers, which is only realistic if the average outcome for discontinuers to $T = 1$ does not change between the baseline and final value. This seems a strong assumption that may be questionable in many settings, and single imputation as in these methods has the potential to lead to underestimates of uncertainty.

If the ITT estimand of interest is the ECT, it seems preferable to impute missing data in the treatment group using the conditional distribution of Y_1 given Y_0 for individuals in the control group [17, 18]. Note that this sample includes both principal continuers and discontinuers to $T = 1$, who cannot be distinguished in the control group, so this involves the assumption that

$$[Y_1|Y_0, Z, T = 1, D_1 = 1] = [Y_1|Y_0, Z, T = 0], \tag{5}$$

This assumption may be weakened by including and conditioning on covariates Z other than Y_1. A similar approach in the more general situation when there are intermediate measures is one of the options in the sensitivity analysis in [17].

These imputation approaches all involve assumptions, which may be controversial. Our final measure of treatment effect for the ITT population avoids the need to specify treatments and impute outcomes after discontinuation. We define an 'on-treatment summary' (OTS) outcome based on measures recorded prior to discontinuation. An example in a symptomatic trial is to obtain repeated outcome measures until treatment discontinuation or the end of the study and measure the outcome as area under the curve of the repeated measures until treatment discontinuation or the end of the study, with the baseline level set at Y_0. Another example is to define a composite binary success/failure measure, treating discontinuation as a failure.

The advantage of an OTS measure is that it can be defined for all individuals in the group to which they are randomized, yet it avoids assumptions required to impute measures after treatment discontinuation. The disadvantage is that the outcome measure does not reflect any effects of treatment manifested after discontinuation. For example, in studies of chronic disease treatments, survival time is often the primary measure of interest, but time of death is not an OTS measure because it is not known for survivors censored at the time of treatment discontinuation. To sum up, the utility of an OTS measure rests on its scientific value for assessing the treatments.

4. Case study: clinical trial of insulin treatments for diabetes

In the insulin study, rescue treatments (a preprandial dose of insulin in the IG group, or addition of Glargine in the Inhaled group) were specified in the protocol for individuals discontinuing treatment in these arms. When estimating the ERT, these rescue therapies become part of a 'treatment regimen'. Because a substantial proportion of completers in the Inhaled group took the rescue therapy (Table I), the outcomes in this arm do not reflect the sole effect of the Inhaled treatment. A strength of this study is that the nature of the treatment regimen is clarified by specifying allowable rescue treatment(s) in the study protocol. However, a substantial proportion of individuals discontinued treatment despite the availability of rescue therapies (Reasons 1–5 in Table I), and were not followed, leaving a sizeable missing data problem for our secondary outcome.

The columns labelled A in Table III show the results from applying the analysis specified in the protocol, with outcome the change in HbA1c between baseline and week 52, with missing values imputed by the LOCF method. An analysis of covariance (ANCOVA) model is used with two dummy variables for treatment group. Insulin secretagogue strata, country, and HbA1c at baseline are covariates. There are 559 patients in ITT dataset, and 28 of them are not used because of missing data. The 95% confidence interval of 'Inhaled'-'IG' is (−0.380, 0.027) so noninferiority of Inhaled is established, because the upper limit is less than 0.4%, and superiority of Inhaled is not established because the interval includes zero. The 95% confidence interval of 'Inhaled+IG'-'IG' is (−0.626, −0.137), so superiority of the combined treatment over IG is established.

Table III. Three ITT ANCOVA Analyses of Diabetes Data.

	A. Outcome=Change from baseline to week 52		B. Outcome=Transformed Proportion of 52 weeks when on treatment and HbA1c $\leqslant$ 7.5%			
	All types of missing data treated by LOCF imputation		B1. Admin censoring or loss to follow up treated by MI		B2. Admin censoring or loss to follow up treated by LOCF	
Regressor	Estimate (95% CI)	*p*-value	Estimate (95% CI)	*p*-value	Estimate (95% CI)	*p*-value
Intercept	−0.50 (−0.83, −0.17)	0.003	0.48 (0.38, 0.59)	<0.001	0.52 (0.41, 0.63)	<0.001
'Inhaled'-'IG'	−0.18 (−0.38, 0.03)	0.090	0.08 (−0.01, 0.17)	0.068	0.09 (−0.01, 0.18)	0.067
'Inhaled+IG'-'IG'	−0.38 (−0.63, 0.14)	0.002	0.12 (0.02, 0.23)	0.021	0.14 (0.02, 0.25)	0.017
Taking insulin secretagogue	−0.08 (−0.28, −0.13)	0.847	−0.03 (−0.12, 0.06)	0.478	−0.05 (−0.14, 0.05)	0.333
Baseline (centered to 0)	−0.47 (−0.57, −0.37)	<0.001	−0.24 (−0.28, −0.20)	<0.001	−0.26 (−0.30, −0.21)	<0.001
Country (DF=10)	—	0.585		0.158		0.009

MI, multiple imputation; LOCF, last observation carried forward.

For the chosen primary measure, change from baseline HbA1c, LOCF is effectively the same as measuring the outcome as change from baseline to treatment discontinuation or censoring. For the case of discontinuation, this means the length of time on drug is not part of the measure, which seems undesirable because this is an important aspect of the drug's effectiveness. Another comment is that the protocol (like many others) does not make clear the assumed nature of treatments after discontinuation. The LOCF method effectively assumes no change in outcome after discontinuation, but under what treatment? If the assumed measure is the average ERT, LOCF is clearly not realistic in the Inhaled (inhaled insulin) arm, because a common reason for discontinuing treatment in the Inhaled arm was lack of control of diabetes, and in reality, an alternative rescue treatment, specifically Glargine, would be applied in this group to reduce HbA1c levels and bring the diabetes back into control. In summary, we feel that the choice of estimand is unclear, and the plausibility of the LOCF imputation method is doubtful.

Turning to other ITT estimands, the EAT does not appear to be plausible for the Inhaled arm, because individuals tended to discontinue treatment when the treatment failed to control the diabetes, and the counterfactual of what would have happened if they stayed on this treatment is not realistic. The ECT, where the control treatment is Glargine, seems close to ERT and more realistic, since Glargine was the chosen rescue treatment for the Inhaled arm. However, this estimand is really assessing a combined Inhaled + IG regimen, which may not be the primary interest for a trial assessing the effectiveness of inhaled insulin alone. Also, the LOCF imputation method is overestimating the HbA1c levels for discontinuers in the Inhaled arm, for the same reasons as stated previously for the ERT estimand.

Given the difficulties in the aforementioned approaches, we prefer an OTS measure to deal with treatment discontinuation in this study. One such measure is the proportion of the 52 weeks from baseline to endpoint for which the individual was on treatment and had HbA1c levels that are low enough for the diabetes to be considered 'under control', namely below 7.5%. Time post discontinuation is included in the denominator but not the numerator, so there is an implicit penalty for discontinuing treatment, as seems appropriate. This measure could be defined for the individual treatments, if time under control while on rescue treatment was not included, or the treatment regimens, if time under control while on rescue treatment was included.

This OTS approach addresses the issue of treatment discontinuation (Types 1–5 in Table I) but not the issues of missing data because of loss to follow up or premature termination of the study (Types 6 and 7 in Table I). We need to impute the OTS measures for individuals with these missing data types. We do this by a multiple imputation (MI) method, assuming MAR, that conditions on information prior to censoring and allows for the possibility that individuals discontinue between the censoring time and 52 weeks.

Because the pattern of missing data is monotone, the imputation is carried out sequentially from time of termination of study (or loss to follow up) to 52 weeks, conditioning the imputations at each visit on observed or imputed values from previous visits. For each missing visit, the algorithm is as follows:

Step 1: the missing HbA1c values are imputed as draws from their predictive distribution, based on a model that includes treatment, previous recorded HbA1c values, whether rescue is initiated prior to the missing value, and other covariates.
Step 2: Treatment discontinuation is imputed to have occurred at time t for subject i with probability P(ti), where P(ti) is the prediction from a logistic regression model of discontinuation (yes, no) on prior HbA1c values, whether rescue is initiated before, and other covariates.
Step 3: If treatment discontinuation is not imputed to happen at a visit, whether rescue therapy is initiated at this visit is imputed based on a logistic regression of rescue (yes, no) on HbA1c values for this visit and other covariates.

This algorithm is repeated to create 20 multiply imputed data sets, and MI combining rules used to propagate imputation uncertainty (Rubin, 1987). The resulting data sets are analyzed by ANCOVA, with an arcsine square root ($\sin^{-1}\sqrt{\text{p}}$) transformation of the proportions to stabilize the variance and reduce skewness in the outcome distribution. The results are presented in the columns labeled B1 in Table III. The results for the same analysis with LOCF imputation are included in the columns labeled B2 in Table III for comparison.

The intercept is lower for MI (0.48) than for LOCF (0.52); hence under the MAR assumption for MI, LOCF imputation results in an overstatement of the transformed proportion of the study period where the diabetes is under control. Results comparing the treatment groups are similar for the two imputation methods, with a marginally significant increase in transformed proportion for the Inhaled group over the IG ($p = 0.08$. 95% CI = (−0.01, 0.16), and a statistically significant increase in transformed proportion for the Combined group relative to the IG group ($p = 0.022$, 95% CI = (0.02, 0.23)). The statistical significance of these results is comparable with those for the analysis specified in the protocol, in column A of Table III; sizes of effects are not directly comparable because of the differences in outcome measures.

5. Conclusions

Our main points are as follows:

(a) Treatment discontinuation is distinguished as a different missing data problem from missing outcome data, involving missing data in covariates defining principle discontinuation strata;
(b) Analyses of the ITT population based on estimands after treatment discontinuation need to be clear about assumptions about the nature of treatments after discontinuation; in particular, the ERT estimand includes any treatments administered after treatment discontinuation and the end of the study, which may complicate the interpretation of treatment effects;
(c) Alternative estimands to ERT can be defined, such as EAT and ECT, that make counterfactual assumptions about treatments after discontinuation. These need to be scientifically plausible, and imputations after discontinuation need to be suitable for the assumed counterfactuals; and
(d) One way of avoiding these difficulties is to define an OTS measure of treatment effect based on data prior to treatment discontinuation. This approach is illustrated using data from the inhaled insulin study.

We have confined attention here to estimands for the ITT population. An issue not addressed is the definition of treatment effects in subpopulations that are the target of per-protocol analysis. Based on principle stratification ideas, one such estimand is the *completer-average causal effect*, which for the data in Table II is

$$\delta_{CACE} = \mu_1^{(00)} - \mu_0^{(00)}, \tag{6}$$

which is analogous to the complier-average causal effect in the compliance literature [4]. The average effects in the other principal discontinuation strata are also potentially of interest, but we suggest that the CACE is usually the primary estimand of interest other than ITT. Estimation of this effect requires imputation of the completer status for treatments other than the treatment actually received, which requires additional assumptions, paralleling the situation with noncompliance; see, for example [7]. Application of our framework to this type of estimand is a topic for future research.

Acknowledgements

This research was supported as part of a Master Services Agreement 12-PAF01969 between the University of Michigan and Eli Lilly, who kindly provided the data set. We greatly appreciate the assistance of Malgorzata Leyk of Eli Lilly in helping us to understand the Insulin study data, and the referees and associate editor for constructive comments.

References

1. Meinert CL. Toward more definitive clinical trials. *Controlled Clinical Trials* 1980; **1**:249–261.
2. Rubin DB. Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics* 1978; **6**:34–58.
3. Holland PW. Statistics and causal inference. *Journal of the American Statistical Association* 1986; **81**:945–970.
4. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables (with discussion and rejoinder). *Journal of the American Statistical Association* 1996; **91**:444–472.
5. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; **58**:21–29.
6. White IR. Uses and limitations of randomization-based efficacy estimators. *Statistical Methods in Medical Research* 2005; **14**:327–347.
7. Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics* 2009; **65**(2):640–649.
8. Frangakis CE, Rubin DB. Addressing complications of intent-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 1999; **86**:365–379.
9. National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press: Washington, DC, 2010.
10. Little RJ, Rubin DB. *Statistical Analysis with Missing Data, 2nd edition*. Wiley: New York, 2002.
11. Rubin DB. Inference and missing data. *Biometrika* 1976; **63**:581–592.
12. Little RJ, Yau L. Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. *Psychological Methods* 1998; **3**:147–159.
13. Peng Y, Little RJ, Raghunathan T. An extended general location model for causal inferences from data subject to non-compliance and missing values. *Biometrics* 2004; **60**:598–608.
14. Food and Drug Administration. *Guidance for Industry*, Vol. E9. Statistical Principles for Clinical Trials, 1998. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073137.pdf.
15. White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. *Clinical Trials* 2012; **9**:396–407.
16. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics* 2013; **23**(6):1352–1371.
17. Little RJ, Yau L. Intent-to-treat analysis in longitudinal studies with drop-outs. *Biometrics* 1996; **52**:1324–1333.
18. Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics* 2013; **12**(6):337–347.