



Longitudinal Clinical Trials, Dropouts, RCT Target Population, ITT, and Enriched Designs

DRAFT 01AUG2012

Ronald W. Helms^{a,b}

^aRho, Inc., Chapel Hill

^bDepartment of Biostatistics, University of North Carolina

© 2012 Rho, Inc., Chapel Hill NC

Corresponding author contact information:

Ronald W. Helms, Ph.D.

Rho, Inc.

6330 Quadrangle Drive

Chapel Hill NC 27517

Ron_Helms@RhoWorld.com

919-408-8000 ext 6222

Key words: MAR, Missing at Random, MNAR, Missing not at Random, Randomized Clinical Trial, Intent to Treat

Running Title: Target Population and ITT in Longitudinal RCTs

Abstract

1. INTRODUCTION

Dropouts typically afflict longitudinal randomized clinical trials (RCTs) [*citation*], leading to presumptively “missing not at random” (MNAR) incomplete data [*citation*: Little & Rubin], virtually impossible intention-to-treat (ITT) analyses [*citation*], biased estimators and corrupted hypotheses [*citation*: HelmsX4]. Desperate biostatisticians have responded by creating bizarre statistical methods, including various flavors of imputation (e.g., LOCF, BOCF, multiple) [*citation*: NAS Report], retrieval and use of off-study post-dropout data [*citation*], exotic models (e.g., pattern mixture models) [*citation*: Little & Rubin], and a novel pseudo-statistical principle: “punishing” the investigator (sponsor) for dropouts [*citation*: FDA]. This desperation leads to many publications – great for academics – but the confusion can be daunting for both researchers and regulators [*citation*].

Some of the confusion stems from varying definitions, or insufficiently careful definitions, of such terms as “the intention-to-treat principle” [*citation*: from Powerpoint presentation, about varying ITT defs]. We attempt to improve the clarity of the issues by introducing some mathematically and statistically more rigorous definitions of a variety of terms commonly used in RCTs as well as some new entities. We will then draw some logical conclusions from those definitions.

The context is longitudinal RCTs, typically evaluating potential therapies for treatment of chronic diseases or conditions. The context includes RCTs for evaluating both medical procedures and products, including drugs. The product context will typically be drug development programs conducted by pharmaceutical companies or medical center researchers while the medical procedure context will typically be university based research programs programs funded by the National Institutes of Health or other organizations. In this paper we use the more general term *product* in place of *drug*.

2. RATIONALE

This paper's primary goal is to present objective tools that can be used to make precise, mathematical definitions of various aspects of various incomplete data approaches and to focus controversies about the various approaches on a controversy's essential issue or issues. If successful, the tools will help biostatisticians construct rigorous, precise definitions of various concepts, such as "The Intent to Treat Principle" for example, for a specific randomized clinical trial (RCT). Success would also help biostatisticians focus controversies by eliminating distractions arising from extraneous issues and helping participants focus on actual differences that are ultimately, in some cases, simply differences of professional preferences.

We will describe, and present rigorous definitions of a number of patient populations, beginning with the "Patient Population of Interest", or the *PPOI*, that is described briefly as the set of all patients to whom investigators would like to extrapolate the results of their research program (consisting of one or more RCTs). We will define and describe several subpopulations, that is, subsets of the *PPOI*, including:

- An RCT's "Target Population" (*TP*), briefly described as the set of all patients to whom the investigators would like to extrapolate the results of a specific RCT;
- An RCT's "Intent to Treat Analysis Population," AP_{ITT} , that is an extrapolation of the RCT's ITT Analysis Set of Subjects to (almost) the entire *TP*.
- An RCT's "Per Protocol Analysis Population," AP_{PP} , that is an extrapolation of the RCT's Per Protocol Analysis Set of Subjects to (almost) the entire *TP*.

The RCT Target Population, *TP*, is a central, essential entity in this topic. Although investigators would like to extrapolate their research program's results to the entire *PPOI*, we shall see that the *TP* is a proper subset of the *PPOI* – the *PPOI* contains patients who cannot be in the *TP* – and extrapolating a single RCT's results to the *TP* is less controversial than extrapolating to the entire *PPOI*. Thus the *TP* becomes the more important "large" patient population; other patient populations are typically subsets of the *TP*.

Here is the rationale underlying defining “intent to treat” for a specific RCT in this context. Let A be a well-defined subset of the TP , let θ_A be a well-defined “Parameter of Interest”, typically a “treatment effect” for patients in A , and suppose we would like to create an unbiased estimator of θ_A , namely $\hat{\theta}_A$, computed from data captured in the RCT, such that $E[\hat{\theta}_A] = \theta_A$. In particular suppose A is the intent-to-treat analysis population, AP_{ITT} , and θ_{ITT} is the intent-to-treat treatment effect in AP_{ITT} . Often AP_{ITT} is a proper subset of TP : TP contains patients who are not in AP_{ITT} . We must presume that $AP_{ITT} \neq TP$ implies $\theta_{ITT} \neq \theta_{AP}$ and, therefore, $E[\hat{\theta}_{ITT}] \neq \theta_{TP}$; that is, the intent-to-treat estimator of the treatment effect is a biased estimator of the Target Population treatment effect, which is the true “parameter of interest”. There is some flexibility in how one defines “intent to treat” for a specific RCT. In this context a goal is to define “intent to treat” for the specific RCT in such a way as to minimize the bias in $\hat{\theta}_{ITT}$, which we do by making the approximation $\theta_{ITT} \cong \theta_{AP}$ as close as possible, which is accomplished by making $AP_{ITT} \cong TP$, while maintaining consistency with (1) one’s notion of “intent to treat” and (2) “reality”. Here, “non-reality” typically arises when one defines the AP_{ITT} in such a manner that defining an unbiased $\hat{\theta}_{ITT}$ is impossible – the statistic cannot be computed from data captured in the RCT. ITT controversies seem to arise from varying methods for attempting to maintain consistency with both (1) and (2) to the maximum possible extent.

3. BACKGROUND AND DEFINITIONS

3.1. Patient and Subject

The following definitions are more-or-less standard in the RCT context.

patient: a *patient* is a human being (in our context) who is afflicted with a disease or condition.

candidate patient: a patient who is a candidate for enrollment in an RCT.

subject: a subject is a patient who has been enrolled in an RCT; the conversion from patient to subject occurs upon *enrollment*, which is defined below.

3.2. Pre-enrollment procedures; Enrollment Criteria; Screening Procedures; Enrollment

Specifying the exact moment of enrollment requires care. In this paper we define several terms leading to a definition of enrollment.

pre-enrollment procedures: RCT protocol-specified procedures that are executed beginning with a candidate patient signing informed consent and continuing until a decision is made on whether to enroll the patient in the RCT.

enrollment criteria: a set of well-defined criteria, explicitly specified in an RCT protocol, that a candidate patient must satisfy in order to be correctly enrolled in an RCT. (That is, it is an error to enroll a candidate patient who does not satisfy all the RCT's enrollment criteria.)

screening procedures: One or more RCT pre-enrollment procedures that are executed to determine whether a candidate patient satisfies the RCT's enrollment criteria.

The following definition of *enrollment* is precise, but not typical of current protocols.

enrollment: Enrollment is an RCT protocol-specified milestone that occurs when an investigator makes an explicit, documented declaration that a candidate patient has completed the RCT's screening procedures, that the evaluations indicate that the patient satisfies all the RCT's enrollment criteria, and the patient is declared to be enrolled as a subject in the RCT.

This definition of *enrollment* is not typical of current practice. However we recommend that an RCT's protocol should explicitly specify enrollment as defined above, including explicitly capturing the investigator's declaration in the RCT's case report form (CRF). The declaration should be date-time "stamped". That date-time is when a candidate patient converts to a subject.

Some protocols specify that enrollment occurs when the patient signs informed consent, an act that is well defined and a critical prerequisite to enrollment. However a patient's consenting to enroll is not the same as *enrollment* as defined above.

Assigning an ID number so that data that are captured during pre-enrollment procedures can be stored properly is one of the early pre-enrollment procedures. Some protocols state that assigning an ID number constitutes enrollment, but assigning an ID number not the same as definition of *enrollment* given above.

Enriched design screening procedures specify a *screening period* during which a candidate patient is administered the study product ("active treatment) to determine whether the patient finds the CTM efficacious and tolerable. The decision of whether to enroll the patient follows such screening periods.

3.3. Patient Population of Interest, *PPOI*

We define two preliminary terms:

well defined criteria to specify a patient population: a set of criteria that are sufficiently objective and evaluable that if a patient were selected at random from the set of all patients in a specified geographic area, the probability that an appropriately-trained physician can correctly determine whether the patient satisfies the criteria is at least 99.9% (or choose some other value near 100%).

[well defined] patient population: a set of patients specified by a well defined set of criteria.

Basically, this means that researchers can almost always determine definitively whether a specific patient is in the population or not. The brackets indicate that we generally omit the qualifier "well defined".

Patient Population of Interest (PPOI): A *PPOI* is a well-defined patient population described as follows. In a specific context (e.g., in the context of a development program for a specific product) the *PPOI* is the set of all patients who: (a) are afflicted with a disease or condition of interest in a particular situation that (b) is in a specified list of diseases or conditions, each of which is specified by well defined criteria, (c) who reside in a precisely specified geographic area, and (d) who satisfy additional well defined criteria.

Our intent is to extrapolate research program+ results to the *PPOI*.

3.4. Parameter of Interest, including Treatment Effect

This paper focuses on clinical outcome variables that satisfy the following:

quantitative, objective clinical measure (or variable): the result of a measurement or evaluation procedure, conducted on a patient or subject, that satisfies the following. Quantitative: the value of the measure is numerical on an ordinal, interval, or ratio scale. Objective: multiple well-qualified investigators conducting the same type of evaluation procedure on the same subject at (hypothetically) the same time will obtain very similar numerical values. The details of reproducibility are specific to the type of measure.

We now define *parameter of interest* and *treatment effect*, which are population parameters that we may wish to estimate and test hypotheses about.

$\theta = \text{Parameter of Interest}$. Consider a quantitative, objective clinical measure or variable, Y , and let θ denote a “Parameter of Interest” related to the distributions of Y in measurements from subjects selected randomly from the *PPOI*. θ is typically a “treatment effect” as defined below. Our illustrations will use scalar-valued parameters but θ might be a vector, i.e., have multiple elements.

$\theta_A = \text{treatment effect of treatment } P \text{ vs. treatment } C \text{ for patients in a subpopulation } A \subset PPOI$. In the present context the following θ is an especially interesting example of a Parameter of Interest. Let A denote a subset (*i.e.*, subpopulation) of the $PPOI$. Let P denote treatment with the product of interest and C denote treatment with a control product (active or placebo). We assume “P” and “C” are well specified (*i.e.*, accurately reproducible in 90+% of pertinent physicians’ offices). Now Let:

$Y_{A,P}$ = quantitative, objective clinical measure from a patient selected at random from A that evaluates the patient’s response to treatment P .

$Y_{A,C}$ = quantitative, objective clinical measure from a patient selected at random from A that evaluates the patient’s response to treatment C .

Note that $Y_{A,P}$ and $Y_{A,C}$ are random variables. We now define:

$\theta_A = \text{the treatment effect of } P \text{ vs. } C \text{ for patients in the subpopulation } A \text{ as } \theta_A = E[Y_{A,P}] - E[Y_{A,C}]$.

Other, more complicated definitions are possible. For example, θ_A could be defined in the context of a statistical model that includes covariates.

A special case of the θ_A above arises when A is actually the entire Patient Population of Interest, $A = PPOI$, and the parameter of interest is the treatment effect for the entire $PPOI$, namely θ_{PPOI} . We will return to this topic after defining some additional interesting subpopulations.

3.5. RCT Target Population and Analysis Populations

RCT subjects are members of a $PPOI$ but not all $PPOI$ members are eligible to participate in the RCT. For example, some $PPOI$ members live so far from any participating clinical center that, as a practical matter, they could not participate in the RCT. Others might live near a center but for various reasons could not or would not participate in the RCT even if they were recruited.

An RCT protocol contains (hopefully well-defined) specifications that define subsets of patients and subjects. Applying one of these sets of specifications to all members of the pertinent *PPOI* defines a subpopulation of the *PPOI*.

RCT Target Population, TP: An RCT's Target Population (RCT's *TP*) is the subset of *PPOI* patients ($TP \subset PPOI$) who satisfy protocol-specified enrollment criteria for the trial.

Notice that this definition extrapolates from the RCT and its subjects to the entire Patient Population of Interest and selects that subset of *PPOI* patients who would satisfy the RCT's enrollment criteria.

An RCT protocol typically contains (hopefully well-defined) sets of criteria that specify several "analysis sets" such as an "ITT analysis set" and a "per protocol analysis set". Sometimes the criteria do not indicate whether they are specifying a subset of subjects or a subset of data from subjects. The following definitions make these points explicit in the names of the subsets.

RCT analysis set of subjects: is a subset of the study's subjects who satisfy specifications in the RCT's protocol and whose data will be used in a specific analysis. The "ITT Analysis Set of Subjects" and the "Per Protocol Analysis Set of Subjects" are two typical examples that will be defined below.

An analysis set of subjects is sometimes called an "analysis population" but we avoid that term; here, a population is a subset of the much larger *PPOI* or *TP*.

RCT analysis set of data: is the subset of the study's data consisting of all data from subjects in an analysis set of subjects. "ITT Analysis Set of Data", consisting of all data from the ITT Analysis Set of Subjects, and the "Per Protocol Analysis Set of Data", consisting of all data from the "per protocol analysis set of subjects", are two typical examples.

The criteria that specify an analysis set of subjects, when extrapolated to the *PPOI* or *TP*, defines an RCT analysis population, $AP \subset TP \subset PPOI$. Typical analysis populations implied by analysis sets of subjects include AP_{ITT} , the ITT Analysis Population, AP_{PP} , the Per Protocol Analysis Population.

Previously, protocol authors have not considered writing analysis set criteria that could easily be extrapolated to the *TP* to define patient subpopulations. One consequence of this paper could possibly be a change in this practice.

Table 1 describes the typical subpopulation relationships and the types of criteria that define increasingly smaller subpopulations.

Table 1. Typical $PPOI \subset RCT TP \subset RCT AP_{ITT}$

Population	Criteria: All patients who:
<i>PPOI</i>	Live in a specified geographic area, and
	Have a well specified disease or condition in a specified list, and
	Satisfy other criteria (subset of RCT Enrollment Criteria)
RCT <i>TP</i> (Target Pop.)	Are <i>PPOI</i> patients and satisfy <u>all</u> [remaining] RCT Enrollment Criteria
RCT AP_{ITT} (intent to Treat)	Are <i>TP</i> patients and: Are available to RCT clinics, and
	Would enroll in RCT, and
	Satisfy <i>MPE</i> ("Maximum Possible Extent) <i>criteria</i> , e.g., would remain in the RCT thru randomization and 1 st treatment-masked dose.
RCT AP_{PP} (Per Protocol)	Are AP_{ITT} patients and would remain in the RCT thru entire treatment-masked evaluation period (<i>i.e.</i> , are "completers").

3.6. Practicality: RCT Target Population is a Proper Subset of the $PPOI$, $TP \neq PPOI$

An RCT TP is an extension from RCT patients or subjects to the entire $PPOI$; that is, $TP =$ set of all $PPOI$ patients who satisfy the RCT's enrollment criteria. And, as investigators would like to extrapolate the research program's results to the entire $PPOI$, they make each RCT's enrollment criteria match $PPOI$ criteria as closely as possible.

Even if the criteria that define the TP are identical to the $PPOI$ criteria the two populations are not identical. For example, some $PPOI$ patients live too far from any of the RCT's clinical centers to participate in the RCT. Other $PPOI$ patients who live near an RCT clinic would not or could not participate in the RCT. As a practical matter one cannot know all the differences between the TP and the $PPOI$. Consequently, in fact, the TP becomes the more important "larger" patient population. We then hope that the corresponding parameters of interest are approximately equal, $\theta_{TP} \cong \theta_{PPOI}$.

4. RIGOROUS DEFINITIONS FROM THE INTENT-TO-TREAT (ITT) PRINCIPLE

The statistical literature contains varying descriptions of "the" intent-to-treat principle (Hollis and Campbell, 1999; Schulz, *et al.*, 2010) but few if any are sufficiently precise for application in the present context. One description that captures the essence is, "The analysis must incorporate data from all subjects the investigators intended to treat (study) as documented in the protocol."

Recall the ITT rationale described above. We will define an ITT Analysis Set of Subjects and we will define the ITT Analysis Population, AP_{ITT} by extending the criteria that define ITT Analysis Set of Subjects to the entire TP and acknowledge that $AP_{ITT} \subset TP$, possibly a proper subset, $AP_{ITT} \neq TP$. θ_{TP} is the treatment effect of interest for the entire TP ; when restricted to the AP_{ITT} subpopulation this becomes θ_{ITT} . Because $AP_{ITT} \neq TP$ we must presume $\theta_{ITT} \neq \theta_{AP}$. We

attempt to make $\theta_A \cong \theta_{TP}$ by specifying ITT criteria that make $AP_{ITT} \cong TP$, consistent with (1) one's notion of "intent to treat" and (2) "reality" as described above.

4.1. Definitions of the Intent-To-Treat Sets

We now pose criteria that definition the intent-to-treat sets.

The *ITT Analysis Set of Subjects* includes all the RCT's subjects who are members of the *TP* and who satisfy *MPE* (maximum possible extent) *criteria* described below.

The *ITT Analysis Set of Data* includes all data from all RCT subjects who are members of the ITT Analysis Set.

4.1.1. *TP* Membership Issues

First consider the phrase, "all the RCT's subjects who are members of the *TP*". To be enrolled in the RCT a patient must appear to satisfy all the RCT's enrollment criteria, *i.e.*, appear to be a member of the *TP*. There is this troubling phrase, *appear to be*. ***

Screening procedures are imperfect. Sometimes investigators enroll a candidate patient after screening procedures indicate the patient satisfies all enrollment criteria but the investigator discovers later, perhaps after randomization, that the subject does not satisfy the enrollment criteria, *i.e.*, is not a *TP* member. Although some biostatisticians would argue that such subjects should be included in the ITT Analysis Set of Subjects, the fact is that their enrollment was simply an error, resulting from an imperfect screening procedure. A patient's membership in the *TP* is a patient's characteristic before enrollment and that is, in principle, discernible before randomization.

Increasing the comprehensiveness of the screening procedure would increase the rate of detection of non-*TP* members and lower the rate of enrolled non-*TP* members. But there are practical limits to

the screening procedure comprehensiveness and, as a practical matter, some non-*TP* patients will be enrolled and randomized.

The investigators have no “intention to treat” a non-*TP* patient, regardless of when the non-membership is discovered.

If non-*TP* subjects are included in analyses, the resulting estimator of the Parameter of Interest will pertain to some patient population other than the *TP* and consequently, we must presume, will be biased. That is, including non-*TP* members in the analysis set explicitly introduces a bias into the estimator. To avoid such a bias we limit the ITT Analysis Set of Subjects to subjects who are *TP* members, even if non-membership is discovered after randomization. The mathematical statement of this procedure is: $AP_{ITT} \subset TP$. This procedure is inconsistent with some current practice and is potentially controversial.

4.1.2. Effect of TP membership, $AP_{ITT} \subset TP$, on Randomization as a Basis for Inference

Many statisticians rely on randomization and randomization tests as a foundation for statistical inference. How does excluding post-randomization-discovered non-*TP* members from the ITT Analysis Set of Subjects affect the random treatment assignment of other subjects in the ITT Analysis Set of Subjects? The answer depends somewhat on the actual randomization procedure.

4.1.2.1. Unrestricted Randomization

Suppose the randomization is unrestricted; all subject random assignments are stochastically independent. Omitting excluding post-randomization-discovered non-*TP* members from the ITT Analysis Set of Subjects cannot no affect the random assignment of remaining subjects to treatment and, therefore, has no effect on using the randomization as a basis for inference.

4.1.2.2. *Randomization in Blocks*

Now consider a simple randomization-in-blocks procedure that is often used to balance the numbers of subjects in treatment groups within clinical centers. (We illustrate with 4-subject blocks to keep the illustration small; typical blocks are larger.) The six possible sequences of assignments of subjects 1-4 within a block are: PPCC, PCPC, PCCP, CPCP, CPPC, CCPP. When the time comes to randomize the first subject in a block we choose one of the six possible sets of assignments at random, without restriction, each with probability 1/6. Choosing a block without restriction does not mean subjects are randomized without restriction; indeed we restrict the randomization to only 6 of the 16 possible sequences of treatment assignments. Subsequent subjects (after the first) in the block are assigned to assignments from the sequence that was selected when the first subject was assigned.

Blocked random treatment assignments are typical in RCT research but the restriction on randomization is rarely incorporated into the statistical analysis. In contrast, agricultural statisticians have incorporated “randomized block design” elements, including d.f. adjustments, into ANOVAs for about a century.

After all 4 subjects in a block have been assigned suppose one of them is discovered to be a non-*TP* member and removed from the ITT Analysis Set of Subjects. Indeed, suppose this happened at least once to every block in the RCT. What effect does this have on the use of randomization as the basis for inference?

Most statisticians would agree that if we randomly (without restriction and using equal probabilities) removed one subject from each block this procedure would have no effect on the use of randomization as the basis for inference. In contrast, if we systematically removed the first “P subject” from each block most statisticians would probably decline to use the randomization as a

basis for inference. The question arises: is removing post-randomization-discovered non-*TP* members more nearly comparable to completely random removal or systematic removal?

One could argue that the screening procedure's incomprehensiveness or lack of ability to detect non-*TP* members stems from random noise in evaluating patient characteristics that define *TP* membership. Capturing additional post-randomization patient characteristic data – more information (also afflicted with random variability) – leads to detection of non-*TP* membership. Post-randomization detection is based on random variables that are realized both before and after randomization. One could argue that the post-randomization detection is more like the random removal than the systematic removal and therefore has minimal or no effect on using randomization as a basis for inference.

Additional analyses of these issues will be useful but are beyond the scope of this paper.

4.2. *MPE* (“to the Maximum Possible Extent”) Criteria

The definition of the ITT Analysis Set of Subjects includes all the RCT's *TP* subjects “to the maximum possible extent” (*i.e.*, those who satisfy “*MPE criteria*”). *TP* subjects who do not satisfy the *MPE* criteria are excluded. Many differences of interpretation of “The Intent to Treat Principle” can be identified as differences in interpreting *MPE criteria*. We do not champion a particular interpretation; rather we hope that identifying the precise point of differing interpretations can help objectify the discourse.

As a first illustration consider a *TP* subject who drops out between enrollment and randomization. Many biostatisticians would exclude such subjects from the ITT Analysis Set of Subjects, *i.e.*, one relatively noncontroversial *MBE* criterion is: Include subjects who were randomized.

A *TP* subject who drops out between enrollment and the time of the first post-randomization evaluation (capture of a *Y* value), *i.e.*, a subject who has no post-randomization data, is a more illuminating illustration.

Statisticians exhibit remarkable variability in their approach to implementing “to the maximum possible extent”. For example, some biostatisticians cheerfully recite an “ITT jingo”: “If you randomize you must analyze” (Helms, 2012a) in support of a position that the ITT Analysis Set of Subjects must include all randomized subjects. On the other extreme, in RCTs evaluating analgesics for chronic pain it is not unusual for a protocol to define the ITT Analysis Set of Subjects as “All subjects who were randomized, received at least one post-randomization dose of clinical trial material [the treatment], and were evaluated on at least one post-randomization occasion [*i.e.*, produced at least one post-randomization primary outcome data point].” (Helms, 2012b) A significant proportion of the audience and speakers in a 2012 JSM session entitled “The Search for Missing Data: The Impact of the National Academies of Science Report” literally laughed out loud when this specification was read (Helms, 2012c).

A reality of clinical research is that despite investigators’ best intentions and best efforts, some *TP* subjects do not produce any data for the ITT Analysis Set of Data. Including such subjects in the ITT Analysis Set of Subjects implies that the ITT Analysis requires data that will not exist and, therefore, is literally impossible. Defining, in advance, an ITT statistical procedure that one knows cannot possibly be performed is an interesting behavior. As noted in the Introduction, biostatisticians have created ingenious and bizarre methods to work around this problem.

We leave the judgments about how to specify the *MPE* criteria for a particular RCT to the RCT’s investigators and biostatisticians.

4.3. Definition of the *ITT Analysis Population*

We extend the *MPE* criteria from the RCT subjects to the *TP* and thus define the *ITT Analysis Population*.

The *ITT Analysis Population*, AP_{ITT} , consists of all *TP* patients who satisfy the *MPE* criteria.

To illustrate, consider the specification: “The [RCT] *ITT Analysis Set of Subjects* consists of all subjects who were randomized, subsequently received at least one dose of study medication, and were subsequently evaluated at least once.” This set contains all RCT *TP* subjects who: (1) were available (live near an RCT clinic, etc.), and (2) agreed to participate in the RCT (consent, etc.), and (3) who satisfied enrollment criteria (implied by *TP* membership), and (4) who were enrolled (became subjects), and (5) who remained in the RCT through (didn’t drop out before) randomization, initial treatment, and at least one post-randomization evaluation. Conditions (4) and (5) are the *MPE* criteria that can be extended to the population level as:

MPE criteria example: (a) a *TP* patient who, if recruited and given the opportunity would enroll in the RCT and (b) if enrolled would remain in the RCT through (not drop out before) randomization, initial treatment, and at least one post-randomization evaluation.

5. FUNDAMENTAL SOURCE OF BIAS, WITH APPLICATION TO ITT AND PER PROTOCOL ANALYSES

Recall the definition of the Parameter of Interest, typically a treatment effect, for the *TP*, namely θ_{TP} , and for a subpopulation *A*, namely θ_A . When $A \neq TP$ we must presume that $\theta_A \neq \theta_{TP}$. We noted above that *TP* is the “real” patient population of interest (*TP*), we let *A* denote a subset of *TP*, and we restate the previous sentence in this context: When $A \neq TP$ we must presume that $\theta_A \neq \theta_{TP}$.

In such a case let

$\widehat{\theta}_A$ denote an unbiased estimator of θ_A , computed from the RCT's data.

Then $\widehat{\theta}_A$ is a biased estimator of θ_{TP} , $E[\widehat{\theta}_A] \neq \theta_{TP}$.

This issue is at the foundation of many longitudinal RCT statistical problems. (Including, we state here without proof, much of the bias arising from likelihood or quasi-likelihood analysis of MNAR incomplete data.)

In particular, recall that the ITT Analysis Population, $AP_{ITT} \cong TP$, with corresponding Parameters of Interest θ_{ITT} and θ_{TP} . The two populations, AP_{ITT} and TP , are not equal: $AP_{ITT} \neq TP$ and consequently we must presume $\theta_{ITT} \neq \theta_{TP}$. The two populations would be equal except for the *MPE* (“to the maximum possible extent”) criteria discussed above. Because $AP_{ITT} \cong TP$ we hope, and some people (including the author) believe that in many cases $\theta_{ITT} \cong \theta_{AP}$ and, therefore, if $\widehat{\theta}_{ITT}$, computed from the RCT ITT Analysis Set of Data, is an unbiased estimator of θ_{ITT} then $\widehat{\theta}_{ITT}$ is approximately unbiased for the parameter of *real* interest, θ_{TP} , namely $E[\widehat{\theta}_{ITT}] \cong \theta_{TP}$.

Some longitudinal RCT protocols define a “Per Protocol Analysis Set of Subjects” to include all subjects who complete all of the scheduled longitudinal post-randomization evaluations; (such subjects are sometimes called “completers”) and the corresponding “Per Protocol Analysis Set of Data” that includes all data from such subjects. Using the methods above we can extend these specifications to create a Per Protocol Analysis Population, $\theta_{PP} \subset \theta_{TP}$, analogous to AP_{ITT} , and a corresponding Per Protocol Parameter of Interest, θ_{PP} . As above we readily acknowledge the Per Protocol Analysis Population is different from the Target Population: $AP_{PP} \neq TP$ and consequently we must presume $\theta_{PP} \neq \theta_{TP}$. Many persons (including the author) believe that in many (perhaps most) cases the differences among AP_{PP} members and TP members are so large that θ_{PP} is probably not approximately equal to θ_{TP} and the bias $E[\widehat{\theta}_{PP}] - \theta_{TP}$ may be too large to ignore.

6. PLACEBO CONTROLLED ENRICHED DESIGNS: RESTRICT THE TARGET POPULATION

Enriched designs add enrollment criteria that restrict the RCT's Target Population, *TP*, to patients for whom the product is (a) tolerable and (b) efficacious. One justification for such a *TP* is described in the following subsection.

6.1. Iterative Nature of Typical Medical Practice for Chronic Conditions

Here is a greatly simplified and abbreviated description of a typical medical practice for patients with a chronic disease or condition. A patient presents to a physician with a set of signs and symptoms and the physician performs diagnostic procedures and arrives at a diagnosis. In some cases the physician may prescribe a treatment regimen consisting of one or more products (often, drugs), specified dose(es), and administration schedule(s). For simplicity of exposition consider a regimen using a single product, dose, and administration schedule. The patient tries the product for a period of time and typically either (1) finds the product to be efficacious (provides effective relief of the condition's symptoms) and tolerable, as indicated by a lack of side effects, or (2) finds the product to be inefficacious or intolerable. In the latter case the patient typically returns to the physician who may modify the treatment regimen (*e.g.*, dosage and/or timing), or may switch the patient to a different product or regimen. The physician and patient typically iterate this process until they arrive at an acceptable regimen. The patient continues on that regimen for a relatively long period of time. An important point is that, in many cases, a chronic-disease patient will ultimately use a regimen only if the product is both tolerable and efficacious. This simplified, generic description covers many situations, but not all by any means.

A typical enriched design's enrollment criteria and Target Population includes only subjects who find the product (at specified dose level(s)) to be tolerable and efficacious and, of course, who meet other enrollment criteria.

6.2. Brief Overview of Enriched Design Procedures

The following is a vastly simplified description of enriched design procedures. The pre-enrollment procedures, conducted to determine whether a candidate patient satisfies all enrollment criteria, include a pre-randomization screening period during which patients are treated with the product at a target dose, over a period long enough to evaluate *TP* membership. (Here, we focus on tolerability and efficacy.) Investigators randomize only subjects who appear to be *TP* members, that is, who satisfy all other enrollment criteria and the specific enrollment criteria that the patient must find the product (at the specified dose) to be tolerable and efficacious. After randomization the design includes a typical longitudinal treatment-masked evaluation period.

Unfortunately, screening procedures are imperfect and, to use the vernacular, "Stuff happens." Despite careful study execution some non-*TP* subjects get randomized.

6.3. Product Treatment Arm Subjects Discovered to be non-*TP* Members

Typically, some subjects in the product treatment arm find the product intolerable or insufficiently efficacious and will drop out. Such a subject is not a member of the *TP*, a fact that is unfortunately discovered after randomization.

As we discussed above, the ITT Analysis Set of Subjects excludes non-*TP* subjects even when the discovery is made after randomization. We also noted that this conflicts with typical current practice.

As we also discussed above, including non-*TP* members' data in the analyses would introduce bias in estimators and corrupt hypotheses; estimates and inferences would pertain to some population other than the Target Population.

6.4. Placebo Treatment Arm Subjects Discovered to be non-*TP* Members

In this section we assume the control treatment is a placebo. Typically, subjects also drop out of the placebo treatment arm, sometimes for intolerability, more often for lack of efficacy (LOE), and for other reasons that we shall ignore in this discussion.

Is a placebo dropout a *TP* member? If this subject had been assigned to the product treatment arm it is possible that the subject would have found the product to be efficacious and tolerable, i.e., this subject might belong to the *TP*. But there is no post-randomization information on this point.

One cannot reasonably conclude that a placebo arm LOE or intolerability dropout was a non-*TP* person. Consequently, the ITT Analysis Set of Subjects includes placebo treatment arm dropout subjects and the ITT Analysis Set of Data includes their available data. If the estimator of the Parameter of Interest depends upon having complete data from every subject in the ITT Analysis Set of Subjects then post-dropout data will have to be invented for such dropout subjects.

6.5. ITT Conundrum

Application of the ITT Principle in this context and in this manner leads to a procedure that some biostatisticians (including the author) will find intolerable: Exclude product treatment arm LOE or intolerability dropouts from the ITT Analysis Set of Subjects but include corresponding placebo treatment arm dropouts.

We have initiated, but not completed, research on several alternatives that many biostatisticians might find more palatable.

One “lesson” from this section is that the reasons why a subject drops out of a study should be carefully evaluated and recorded in the RCT data.

7. ACTIVE CONTROLLED ENRICHED DESIGNS

The description of placebo controlled enriched designs in the preceding section is also applicable to active controlled enriched designs with the following changes.

- The Target Population is modified to include only patients who find both the active control and the product to be tolerable and efficacious (and who satisfy the other enrollment criteria).
- The screening procedure is expanded to include two screening periods, one for the product (as above) and another for the active control. A patient who finds either treatment to be intolerable or inefficacious is deemed not a member of the *TP* and is not enrolled.
- The *MPE* criteria for the ITT Analysis Set of Subjects are the same for both treatment arms. A subject in either treatment arm who drops out for LOE or intolerability is deemed to have been discovered, after randomization, to be a non-*TP* member and is excluded from the ITT Analysis Set of Subjects.

Such a design should eliminate the vast majority of dropouts whose post-dropout missing data are considered or presumed to be MNAR. This could eliminate a major issue that arises in a substantial proportion of longitudinal RCTs for treatment of chronic diseases or conditions.

These procedures do not eliminate other dropouts and the problems raised by their incomplete data. Some proportion of other dropouts will discontinue for reasons that are not related to the study or its treatments. (“My boyfriend moved to Alaska and I’m going with him,” for example.) Many biostatisticians consider post-dropout missing data from such subjects to be MAR and will (1) include such subjects in the ITT Analysis Set of Subjects and (2) include the subject’s

pre-dropout data in the ITT Analysis Set of Data. Analysis of the incomplete data can be accomplished via likelihood or quasi-likelihood methods such as a mixed model for repeated measures, GEE, or similar.

Other dropouts, whose posts-dropout missing data are considered MNAR, remain a problem, but the proportion of such dropouts may be so low that treating them as if their post-dropout missing data were MAR might not introduce unacceptable biases into the results.

Active-controlled designs are typically non-inferiority or equivalence designs, which raise two additional issues. Each of these types of design requires specification of an often difficult-to-specify parameter, δ , the non-inferiority margin or the equivalence margin. This issue has previously been well studied and we do not address it here. Cost is a second issue: typically, non-inferiority designs and equivalence designs require “much larger” numbers of subjects, and are therefore more expensive, than a typical superiority design for the same context. Study design is an exercise in tradeoffs. A designer must weigh the alternatives; for example, is the advantage of essentially eliminating the MNAR dropout problem worth the additional cost of a non-inferiority or equivalence design?

8. CONCLUSIONS

9. REFERENCES

Helms, R.W. (2012a). Personal observation: two of five presenters in the 2012 JSM session (#384) entitled “The Search for Missing Data: The Impact of the National Academies of Science Report” recited this jingo. Each of the two self-described as having a cardiology research background and related their attitude to research in an area with a small proportion of dropouts.

Helms, R.W. (2012b). Personal observation: As a consultant to multiple pharmaceutical companies engaged in analgesic research the author has observed this language, or essentially equivalent language, in a variety of confidential protocols that cannot be cited directly because they are protected by contractual nondisclosure restrictions.

Helms, R.W. (2012c). Personal observation; the author was one of the presenters in the session.

Hollis S. and Campbell, F. (1999). "What is meant by intention to treat analysis? Survey of published randomised controlled trials." *BMJ*. 1999 Sep 11;319(7211):670-4.

Schulz K.F., Altman, D.G., Moher, D.; CONSORT Group (2010). "CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials." *J Clin Epidemiol*. 2010 Aug; 63(8):834-40.

10. ACKNOWLEDGEMENTS

This research was supported, in part, with funds from Rho, Inc. and the U.S. National Institutes of Health through National Institute of Allergy and Infectious Disease Contract HHSN272200800029C.

We are grateful for a referee's careful review and excellent suggestions for improvement and also for review and comments by our colleagues at Rho.

