Choosing Estimands in Clinical Trials with Missing Data

Craig H. Mallinckrodt[1]

1. Lilly Research Labs, Eli Lilly and Co, Indianapolis, IN, USA.

Abstract
Recent research has fostered new guidance on preventing and treating missing data. Consensus exists that clear objectives should be defined along with the causal estimands, trial design and conduct should maximize adherence to the protocol specified interventions, and a sensible primary analysis should be paired with plausible sensitivity analyses. An estimand is simply what is to be estimated. Two general categories of estimands are: effects of the drug as actually taken (de-facto) and effects of the drug if taken as directed (de-jure). De-jure and de-facto estimands each have strengths and limitations. An iterative process including objectives, estimands, design, analysis, and sensitivity analyses can be used to guide protocol development. Objectives should reflect the diverse needs of regulators, payers, prescribers, patients, care givers, sponsors, and other researchers. Although design and analysis considerations should not dictate choice of estimand, these considerations should not be ignored. For example, maximizing adherence reduces sensitivity to missing data assumptions for de-jure estimands, but may reduce generalizability of results for de-facto estimands if the methods used to maximize adherence in the trial are not feasible in clinical practice. Both de-jure and de-facto estimands are often needed to understand drug benefit and de-jure estimands will often be the focus of safety evaluations. Newer approaches such as reference based controlled imputation provide useful options for analyses of de-facto estimands. A sequential testing approach starting with a de-jure estimand(s) followed by a de-facto estimand(s) may be useful in assessing drug benefit.

Key Words: Missing Data, Clinical Trials, Estimands

1. Introduction

Missing data is an ever present problem in clinical trials that can bias treatment group comparisons and inflate rates of false negative and false positive results (1-8). Fortunately, missing data has been an active area of investigation with many advances in statistical theory and in our ability to implement that theory (1, 3, 5, 7, 8). This research set the stage for new and updated guidance for preventing and handling missing data in clinical trials. Most notably, an expert panel from the National Research Council that was commissioned by FDA issued an extensive set of recommendations that is clearly influencing practice (3).

The NRC recommendations set forth an overarching framework for tackling the problem of missing data. Key pillars of this framework include: 1) clear specification of trial objectives, including defining the causal estimands; 2) trial design and conduct to maximize adherence to protocol defined interventions; 3) and, a sensible primary analysis supported by plausible sensitivity analyses that assess robustness of results to assumptions about the missing data.

The need for clarity in objectives and estimands is driven by ambiguities arising from the missing data. Data may be intermittently missing or missing due to dropout. Patients may or may not be given rescue medications. Assessments after withdrawal from the initially randomized medication or after the addition

of rescue medications may or may not be taken, and if taken, may or may not be included in the analyses (9).

Conceptually, an estimand is simply what is being estimated (3). Given the emphasis on clear objectives, choice of the primary estimand has been the subject of considerable discussion (1, 3, 4, 8-16). With the variety of clinical trial scenarios and missing data possibilities, consensus on a universally best estimand is neither realistic nor desirable. Therefore, recent attention has turned to how to choose estimands.

Leuchs et al (11) proposed a process chart that begins with defining the primary estimand, followed by design, analysis, and sensitivity analyses. In a letter to the editor in response to the Leuchs' paper the PSI/EPSI working group (WG) advocated an additional step prior to choosing the primary estimand. The WG proposal began by considering objectives and then proceeded to the other steps in an iterative manner so that interactions between the various components could be considered (12).

The purpose of the present paper is to elaborate upon the proposal outlined by the WG in their brief letter to the editor. The paper is organized by first providing an overview of objectives and estimands. Next, fundamental considerations for estimands are discussed along with data, design, and analysis considerations. Key points are then illustrated by way of an example. Lastly, as part of an overall discussion, a sequential testing approach is outlined as a conceptual alternative to choosing a single estimand.

2. Objectives and Estimands

Trial objectives are typically driven by the decisions to be made from the trial results. These decisions depend in part on stage of development. Phase II trials are typically used by drug development decision makers to determine proof of concept or to choose doses for subsequent studies. Phase III, confirmatory, studies typically serve a diverse audience and address diverse objectives (11). For example, regulators render decisions regarding whether or not the drug under study should be granted a marketing authorization. Drug developers and regulators must collaborate to develop labeling language that accurately and clearly describe the risks and benefits of approved drugs. Payers must decide if / where a new drug belongs on its formulary list. Prescribers must decide for whom the new drug should be prescribed and must inform patients and care givers what to expect. Patients and care givers must decide if they want to take the drug that has been prescribed.

With a clearly defined objective and estimand it is obvious what parameter is to be estimated as the basis for evaluating the associated objective. Estimands can be divided into two general categories, efficacy or effectiveness. Efficacy may be viewed as the effects of the drug if taken as directed. Effectiveness may be viewed as the effects of the drug as actually taken (1, 6, 10, 11, 15, 16). However, the efficacy and effectiveness nomenclature does not make sense for safety outcomes. A more general terminology is de-jure (if taken as directed) and de-facto (as actually taken), respectively (13).

The NRC guidance (3) discussed five estimands in detail and Mallinckrodt et al (10) proposed a 6th estimand. Focus here is on 3 of those estimands to illustrate key issues. Each of the 3 estimands involves the difference vs control in changes to the planned endpoint of the trial, in all randomized patients, based on the intention-to-treat principle.

1. Estimand 1 is the change due to the treatment regimens as actually taken.
2. Estimand 2 is the change due to the initially randomized treatments as actually taken.
3. Estimand 3 is the change due to the initially randomized treatments if taken as directed.

3. Considerations

3.1. Fundamental considerations

Despite the diversity in potential objectives, some guidance suggests that de-facto estimands are the only estimands of relevance for confirmatory trials (2). Others take a more nuanced view noting that trials addressing symptomatic treatments often have de-jure estimands as primary and trials with hard endpoints such as mortality often have de-facto estimands as primary (14). Even in the same trial multiple estimands may be of interest (1, 10, 11, 12, 15, 16).

De-jure and de-facto estimands each have strengths and limitations. De-jure estimands may be considered counterfactual for groups of patients because treatment effects are assessed as if taken as directed when in any meaningfully sized group some patients will not adhere (3). However, de-jure estimands provide valid estimates of what to expect for patients that do adhere – and in most clinical settings the vast majority do adhere (15).

De-facto estimands may be considered counterfactual for individual patients because treatment effects are assessed from a mix of adherent and non-adherent patients, but each patient is either adherent or not adherent, no patient is both. On the other hand, de-facto estimands may provide valid estimates of what to expect from the group as a whole (2, 3).

3.2.Data considerations

With **e**stimand 1 (de-facto, treatment regimens) data after discontinuation of the initially randomized medication and/or addition of rescue medication are included in the analysis (3). Rescue medications can mask or exaggerate both the efficacy and safety effects of the initially assigned treatments, thereby invalidating causal inferences for the originally assigned medication. Therefore, inference for estimand 1 is in regards to treatment policies or regimens (1, 7, 9, 10, 11, 15, 16). However, the most relevant questions in early research and initial regulatory reviews are often about the causal effects of the investigational drugs, not treatment policies (1, 10, 16).

O'Neill and Temple (14) noted that estimands requiring data after withdrawal of randomized medication and / or initiation of rescue may be more common in outcomes trials where the presence / absence of a major health event are the endpoint and/or the intervention is intended to modify the disease process. Symptomatic trials (symptom severity is the endpoint) typically avoid the confounding from rescue medications by using a primary estimand and analysis that exclude data after initiation of rescue. This trend in the use of follow up data may be as much driven by ethics as by design or inferential interest. For example, in a cancer trial where overall survival is the primary endpoint, if a patient has tumor progression there is no ethical alternative but to try rescue medication. Hence, for a de-facto estimand little choice exists but to use the data after initiation of rescue if overall survival time is the focus. Nevertheless, it may still be possible to assess de-jure estimands for time to progression free survival.

Conceptually, estimand 2 (de-facto, initially randomized treatments) avoids the confounding effects of rescue medications seen with estimand 1 by not allowing rescue medication. However, given the ethical mandate to allow rescue medications and the analytic need to exclude post-rescue data the issue of how to estimate estimand 2 is covered in a subsequent section on analysis considerations.

Estimand 3 (de-jure, initially randomized treatments) is in regards to data the simplest estimand because data after discontinuation or treatment or initiation of rescue are not required.

Most of the discussion on de-jure vs. de-facto estimands has been in the context of assessing drug benefit. However, estimands for assessing drug risk are also important. Consider the following hypothetical example. A drug has the adverse effect of increasing blood pressure. Some patients become hypertensive and discontinue study medication and/or take rescue medication, with subsequent return to normal blood pressure. De-facto estimands would reflect the patients' return to normal, thereby suggesting no change at the planned endpoint of the trial. De-jure estimands would not reflect a return to normal and would reflect increases at endpoint because had the patients been adherent they would likely have continued to be hypertensive. Therefore, for safety assessments de-jure estimands may be particularly relevant.

3.3. Design considerations

Universal agreement exists that trials should aim to maximize adherence to the initially assigned treatments (1, 2, 3, 8, 15, 16). This recommendation is critical because analyses of incomplete data require assumptions about the missing data, which typically are either not reasonable (data are missing completely at random) or cannot be verified from the data (missing at random or missing not at random). Maximizing adherence improves robustness of results by reducing the reliance of inferences on the untestable assumptions about the missing data (1, 3, 5, 7, 8, 15, 16, 17). These considerations have often been in the context of de-jure estimands. However, the impact of maximizing retention on de-facto estimands should also be considered (11).

Increasing adherence may increase benefit from the drug as actually taken. If the measures used to engender adherence in the clinical trial were not feasible in clinical practice the trial could yield biased estimates of effectiveness relative to the conditions under which the drug would be used. In contrast, true values of de-jure estimands are not situation dependent because the setting is inherent to the estimand (when taken as directed) and maximizing retention reduces potential for bias in estimation.

Other factors to consider in assessing generalizability of de-facto estimands are the degree to which randomization, especially in trials with placebo, and / or blinding may influence decisions to discontinue study medication or to discontinue the study (1). Therefore, generalizing results from clinical trials to clinical practice for de-facto estimands suggests that adherence in the trials should be representative of general practice (11).

3.4. Analysis considerations

Analyses of de-jure estimands typically involve the assumption that data are missing at random. While it can be argued that this assumption is often reasonable, or at least a good starting point in clinical trial analyses, it cannot be verified (1, 3, 5, 7, 8, 15, 16, 17); hence, the need to maximize retention and conduct plausible sensitivity analyses.

Analyses of de-facto estimands must either direct inference at treatment regimens or in some manner account for the discontinuation of initially randomized treatments and /or the addition of rescue medications. One approach to assessing de-facto estimands about the initially randomized treatments is to impute the data after initiation of rescue and/or discontinuation of the initially randomized study medication under the assumption that initially randomized active medications have no (or diminished)

effect after discontinuation / rescue (1, 7, 8, 10, 13, 15, 16). This assumption is often reasonable in trials of symptomatic interventions (14, 18).

Such imputations for continuous endpoints have historically been done using baseline observation carried forward (BOCF). For categorical endpoints non-responder imputation (NRI) has often been used wherein all patients that discontinue initially randomized medication and/or use rescue medication are considered non-responders, regardless of the outcome observed at the planned endpoint of the trial. However, single imputation approaches such as BOCF and NRI have a number of disadvantages and more principled multiple imputation based approaches are gaining favor.

In BOCF and NRI, the assumption of no change from baseline is made in order to ascribe no pharmacologic benefit from the drug if it is not taken. However, these approaches ignore the potential changes from non-pharmacologic sources that are often seen in trials (study effect, placebo effect) and would therefore be valid only in those situations where there was no change in a placebo group over time (1, 10, 16). The potential bias in BOCF estimates can be large, resulting in inflated type I error rates or loss of power in testing de-facto estimands (19). In addition, BOCF makes no sense in situations where the therapeutic aim is to prevent worsening. In these situations, carrying the baseline observation forward ascribes a good outcome to patients that discontinue (1). Moreover, as a single imputation technique, BOCF assigns the same change score (zero) to every patient that discontinues, which results in underestimates of variance and standard error. Therefore, BOCF is generally not a useful analytic approach (1, 3, 5, 7, 8, 10, 15, 16, 17).

Multiple imputation-based approaches to test de-facto estimands have come into the literature recently. These methods have been referred to as controlled imputation or more specifically reference-based controlled imputation (1, 7, 8, 13, 15, 16). Full descriptions of these approaches go beyond the present scope. However, the general approach is to use multiple imputation in a manner that accounts for the change in / discontinuation of treatment. In so doing, patients that discontinue from an experimental arm have values imputed as if they were in the reference (e.g., placebo arm). Depending on the exact implementation, imputed values can either reflect no pharmacologic benefit from the drug immediately upon discontinuation / rescue, a decaying benefit after discontinuation / rescue, or a constant benefit after discontinuation / rescue (7, 8, 13, 15, 16).

In contrast to BOCF, reference-based imputation via multiple imputation accounts for the uncertainty of imputation, accounts for study / placebo effects and is valid regardless of whether the therapeutic aim is improvement or prevention of worsening (1, 7, 8, 13, 15). Reference based imputation has been shown to reduce bias and provide better control of type 1 error compared with BOCF (19).

4. Example

The example data comes from clinical trials in depression where the therapeutic aim is the symptomatic treatment of chronic illness. Original results from these trials were previously reported (20, 21, 22, 23). These 4 studies arose from 2 protocols and were included in a New Drug Application for an antidepressant. Within each protocol investigators were randomly assigned to study A or B, thereby creating two identical and adequately powered confirmatory trials run under each protocol.

In addition to studies A and B within each protocol being identical, protocols 1 and 2 were nearly identical. Each trial had four treatment arms with approximately 90 patients each that included two doses of an experimental medication (subsequently granted marketing authorizations in most major

jurisdictions), an approved medication commonly accepted as a standard of care when the trials were planned (paroxetine), and placebo. Assessments on the Hamilton 17-item rating scale for depression [(HAMD17) 24] were taken at baseline and weeks 1, 2, 4, 6, and 8 in each trial. Inclusion and exclusion criteria were nearly identical in the two protocols.

Some aspects of dosing differed between protocols for the experimental drug. Protocol 1 used fixed dosing, whereas in protocol 2 the experimental drug arms had dose titration to minimize initial side effects and had blinded dose reductions for patients with safety or tolerability concerns. These dose changes were implemented in a blinded manner and were applied only to the arms of the experimental drug. Therefore, the present focus is on the standard of care and placebo arms in protocols 1 and 2 because these arms had identical dosing and assessment during the 8-week acute treatment period. The only differences between protocols were that protocol 2 included a 6-month double-blind extension phase following the 8-week acute phase, whereas protocol 1 did not have the extension phase; and, protocol 2 was implemented in Eastern Europe whereas protocol 1 was implemented in the US. A re-analysis focusing on the experimental arms with various doses and dosing regimens across studies and including other studies has been previously published (10).

The identical dosing and assessment schedule across protocols for the standard of care and placebo arms allowed illustration of the association between design features and completion rates, and in turn illustrate what impact completion rates have on estimates of parameters for de-facto and de-jure estimands.

In the original re-analysis (10) estimates of estimand 3 (de-jure, initially randomized treatment) for all four treatment arms were obtained using a direct likelihood based repeated measures approach that assumed data were missing at random. The analysis included the fixed, categorical effects of treatment, investigative site, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariates of baseline score and baseline score-by-visit-interaction. An unstructured (co)variance structure was used to model the within-patient errors. The Kenward-Roger approximation was used to estimate denominator degrees of freedom and adjust standard errors. Analyses were implemented using SAS PROC MIXED (25). The primary comparison was the contrast between treatments at week 8, the endpoint visit.

These studies did not collect data after discontinuation of randomized study drug. Hence, in the original re-analysis estimates of estimand 2 (de-facto, initially randomized treatments) were obtained using the reference-based imputation approach termed placebo multiple imputation (19), which is essentially the same as the copy reference approach (13). As with the de-jure estimand, mean change to Week 8 was the focal point.

More specifically, multiple imputation via SAS PROC MI (25) was used to impute missing outcomes for both drug- and placebo-treated patients using an imputation model derived solely from placebo-treated patients. Analyses of the multiple completed datasets used the same direct likelihood analysis as for the de-jure estimand, with a saturated fixed effects model such that endpoint contrasts, p values, and confidence intervals were equivalent to an ANOVA at the final time point. Rubin's rules (26), as implemented in SAS PROC MIANALYZE (25) were used to combine results and draw inference from the multiple data sets.

To these previous results the present investigation adds BOCF and NRI analyses for estimand 2. For the BOCF analysis, treatment differences were assessed using an ANCOVA model that included treatment, site and baseline severity. For NRI, patients were considered responders if they completed acute treatment (week 8) and improvement was $\geq 50\%$ of baseline severity.

Completion and response rates are summarized in Table 1. Despite the identical dosing, treatment duration, and assessment schedule the rate of missing data for both drug and placebo in protocol 2 was only 1/3 of the rate in protocol 1. Similarly, response rates for placebo were approximately 2-fold greater in protocol 2 than in protocol 1 and the corresponding difference for drug response rate was nearly 2-fold.

Table 1. Completion and response rates from example data.

	Completion Rate (%)		Response Rate (%)	
Study	Pbo	Drug	Pbo	Drug
1A	69.7	70.9	21.4	39.5
1B	61.4	64.3	23.9	35.7
Protocol 1 average	65.6	67.6	22.7	37.6
2A	81.7	89.4	40.9	72.9
2B	92.9	90.7	50.5	60.8
Protocol 2 average	87.3	90.1	45.7	66.9

Mean changes are summarized in Table 2. The average estimates of de-jure estimand 3 (direct likelihood analysis) were similar for the two protocols (1.91 vs. 2.03). Consistent with the higher completion rates in Protocol 2, copy reference and BOCF pooled estimates for de-facto estimand 2 were higher and closer to the corresponding de-jure estimate in Protocol 2 than in protocol 1. Of course, it is impossible to know if the design differences (geographic region, presence / absence of extension treatment) *caused* the disparity in completion rates that drove the differences between studies in estimates of de-facto estimands.

Table 2. Mean changes from example data.

	Direct Likelihood			Copy reference			BOCF		
Study	Delta1	SE	T	Delta	SE	T	Delta	SE	T
1A	2.44	0.96	2.54	1.99	1.11	1.79	2.27	0.90	2.52
1B	1.37	1.23	1.11	0.99	1.13	0.88	0.94	0.99	0.95
Protocol 1 average	1.91			1.49			1.61		
2A	2.90	0.73	3.97	2.71	0.79	3.45	3.11	0.84	3.72
2B	1.16	0.68	1.71	1.00	0.71	1.41	0.69	0.75	0.92
Protocol 2 average	2.03			1.86			1.90		

1. Delta = the difference between treatments in mean change to week

5. Discussion

Consensus exists that the best way to deal with missing data is to prevent it and that a sensible primary analysis should be supported by plausible sensitivity analyses that assess robustness of inferences to violations of missing data assumptions. Consensus also exists on the need for clarity in objectives and estimands. However, given the diverse settings in which clinical trials are conducted it is neither realistic nor desirable to seek consensus on a universally best primary estimand. Therefore, discussion is shifting toward how to choose estimands.

Leuchs et at (11) suggested a process chart that begins with choice of the primary estimand, after which design, analysis and sensitivity analyses could be determined. The PSI / EFSPI working group (WG) on

estimands advocated a refinement of that proposal wherein trial objectives drive choice of estimands in an iterative process to allow design and analysis considerations to be factored into the choice of estimand(s) (12).

In the present paper, fundamental, design, and analysis considerations for choosing estimands were discussed. Example data from four placebo-controlled trials illustrated the potential for association between design features and estimates of de-facto estimands, while estimates of a de-jure estimand were relatively unaffected. The example data also illustrated a reference-based controlled imputation approach to estimate de-facto estimands for the initially randomized treatments. The controlled-imputation family of methods is gaining popularity at a time when use of ad hoc approaches such as BOCF can no longer be justified.

In the example data the large differences between studies in completion rates was associated with large differences in estimates of de-facto estimands - despite identical treatment regimens and assessment schedules during the acute treatment periods. However, it is not known whether the difference in design features cause the differences in results. Moreover, the example data reflect only one clinical scenario, acute treatment of depression, which is one with subjective endpoints and historically high rates of early discontinuation. Therefore, the large differences in completion rate in the example data may be more indicative of what can happen rather than what would be expected to happen broadly across various settings. The results do however suggest the need to carefully consider the generalizability of effectiveness results in this setting.

In the iterative process proposed by the WG design and analysis considerations can influence choice of primary estimand. This is not in conflict with the notion that such considerations should not lead to compromises in the meaningfulness of endpoints and estimands simply as a means of dealing with missing data (27). Rather, jointly considering all aspects of the process can result in focus on estimands that are more meaningful given the circumstances.

For example, trial design and conduct to maximize adherence may often be associated with de-jure estimands in contrast to the more naturalistic settings in which effectiveness is often best assessed (11). If the anticipated rate of adherence is too low such that plausible departures from MAR would likely overturn positive findings for a de-jure estimand then focus may sway more to de-facto estimands than would otherwise have been the case. If the experimental conditions of the trial sufficiently deviate from clinical practice so as to limit generalizability of de-facto estimands then focus may sway more to de-jure estimands than would otherwise be the case.

The WG's letter to the editor also noted the multifaceted nature of clinical trials (12). Confirmatory trials must address the diverse objectives of regulators, health technology assessors / payers, prescribers, patients, caregivers, sponsors, other researchers, etc. Even for a single stake-holder in a single trial it is often important to know what happens when a drug is taken as directed (de-jure estimand) and to know what happens when the drug is taken as in actual practice (de-facto estimand). Therefore, no single estimand is likely to best serve the interests of all stake holders and de-jure and de-facto estimands will both be of interest (1, 10, 11).

The benefit from testing multiple estimands in a single trial can be seen in the following hypothetical example. For this example consider effectiveness as the sum of efficacy and adherence (effectiveness = efficacy + adherence). Drug A and Drug B (or dose A and dose B of a drug) have equal effectiveness but different efficacy and adherence:

	Efficacy	Adherence	Effectiveness
Drug A	High	Low	Average
Drug B	Low	High	Average

The differences in clinical profiles would be important to understand. Drug A might be the best choice for patients with more severe illness. And Drug B might be best for patients with less severe illness and / or safety and tolerability concerns.

In some situations a sequential testing approach may be considered that capitalizes on the logical ordering of estimands (10). For example, for a drug to be effective it must have efficacy. Therefore, the primary estimand to assess drug benefit could a de-jure (when taken as directed) estimand; if successful, then the de-facto (as actually taken) estimand can be tested. For safety outcomes de-jure estimands may generally be of primary interest.

To illustrate, consider the example above as being two doses of a drug (A and B) that are compared versus placebo in a confirmatory trial. Based on a de-facto estimand the high dose does not differ from the low dose. Results from a de-jure estimand show a significant advantage for the high dose over the low dose. Based on the de-facto estimand the study conclusion would be that high dose did not differ from low dose. If sequential testing of the de-jure and then de-facto estimand is used the inference would be that high dose had greater efficacy but that advantage was negated by more early discontinuations.

This more nuanced understanding could lead to additional investigation that could lead to more optimized patient outcomes. For example, subgroups of patients who especially benefit from or tolerate the high dose might be identified from the existing data or from a new trial (non-responders to low dose). Or, alternate dosing regimens that might improve the safety / tolerability of the high dose, such as titration, flexible, or split dosing (40 mg every two weeks rather than 80 mg every 4 weeks), could be investigated in subsequent trials.

Several caveats to the sequential testing scheme should be considered. First, the approach as outlined here is conceptual and before being put into practice the operational characteristics, especially in regards to type I error, need to be understood. In addition, trial design and conduct must be suitable for estimation of both estimands. Design alterations to enhance the meaningfulness of a secondary estimand at the expense of the primary estimand should be avoided. However, if it is possible to obtain meaningful estimates of a secondary estimand while focusing trial design and conduct on the primary estimand, sequential testing warrants consideration. One such scenario would be when adherence is maximized using trial design and conduct features that are consistent with clinical practice. In this scenario the trial design and conduct are consistent with the de-jure primary estimand and results from de-facto estimands would be generalizable to clinical practice.

In the sequential testing scheme those that make decisions about individual patients (prescribers, patients, caregivers, etc.) may focus most on the de-jure estimands and secondarily on de-facto estimands. Those that make decisions about groups of patients (e.g., regulators, HTAs) may focus most on the de-facto estimand and therefore require positive findings on both de-jure and de-facto estimands. But all decisions makers will benefit from understanding both de-jure and de-facto estimands.

Conclusions

De-jure and de-facto estimands each have strengths and limitations. An iterative process can be used to choose estimands, beginning with the objectives required to address the needs of regulators, payers, prescribers, patients, caregivers, sponsors, and other researchers. No single estimand is likely to meet the

needs of all stake-holders. The focus on potential bias and sensitivity to assumptions for estimates of de-jure estimands drove the recommendation to maximize adherence. However, it is also important to consider generalizability of results for de-facto estimands if efforts to maximize adherence in the trial are not feasible in clinical practice.

References

1. Mallinckrodt, CH. Preventing and Treating Missing Data in Longitudinal Clinical Trials: A Practical Guide. 2013; Cambridge University Press. New York.

2. Committee for Medicinal Products for Human Use (CHMP). Guideline on missing data in confirmatory clinical trials. 2010. EMA/CPMP/EWP/1776/99 Rev. 1

3. National Research Council (2010). The prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioural and Social Sciences and Education. Washington, DC: The National Academies Press.

4. O'Neill RT and Temple R. (2012). The Prevention and Treatment of missing Data in Clinical Trials: An FDA Perspective on the Importance of Dealing With It. Clinical Pharmacology and Therapeutics. doi:10.1038/clpt.2011.340

5. Molenberghs G, Kenward MG. (2007), Missing Data in Clinical Studies. Chichester: John Wiley & Sons.

6. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, and Mancuso JP. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. Drug Information Journal. 2008; 42:305-319.

7. Molenberghs G, Fitzmaurice G, Kenward M, Tsiatis A, Verbeke G. Handbook of Missing Data Methodology. (2015). CRC Press, Boca Raton

8. O'Kelley M and Ratitch B. Clinical Trials with Missing Data. (2014). Wiley, Chichester

9. Mallinckrodt CH, Kenward MG. Conceptual considerations regarding choice of endpoints, hypotheses, and analyses in longitudinal clinical trials. Drug Information Journal. 2009; 43(4):449-458.

10. Mallinckrodt, CH, Lin Q, Lipkovich I, Molenberghs G. A structured approach to choosing estimands and estimators in longitudinal clinical trials. Pharmaceutical Statistics. 2012, 11 456–461

11. Leuchs, AK, Zinserling J, Brandt A, Wirtz D, Benda N. Choosing appropriate estimands in clinical trials. Therapeutic Innovation and Regulatory Science. February 17, 2015, doi: 10.1177/2168479014567317.

12. Garrett A. Choosing Appropriate Estimands in Clinical Trials (Leuchs et al): Letter to the Editor. Therapeutic Innovation & Regulatory Science. May 6, 2015 doi:10.1177/21684790155860.

13. Carpenter J, Roger J, and Kenward M. Analysis of Longitudinal Trials with Missing Data: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation. 2013. J Bio pharm Stat 23:1352-1371

14. O'Neill RT and Temple R. (2012). The Prevention and Treatment of Missing Data in Clinical Trials: An FDA Perspective on the Importance of Dealing With It. Clinical Pharmacology and Therapeutics. doi:10.1038/clpt.2011.340

15. Mallinckrodt CH, Roger J, Chuang-Stein C, et al. Recent Developments in the Prevention and Treatment of Missing Data. Therapeutic Innovation and Regulatory Science; 2014, 48(1): 68-80.

16. Mallinckrodt CH, Roger J, Chuang-Stein C, et al. Missing data: Turning guidance into action. Statistics in Biopharmaceutical Research; 2013, 5(4): 369-382.
17. Verbeke G, Molenberghs G. (2000). Linear Mixed Models for Longitudinal Data.

18. Kim Y. Missing Data Handling in Chronic Pain Trials. Journal of Biopharmaceutical Statistics. 2011; 21: 2, 311 – 325.

19. Ayela B, Lipkovich I, Molenberghs G, Mallinckrodt CH. (2014). A Multiple -Imputation-Based Approach to Sensitivity Analyses and Effectiveness Assessments in Longitudinal Clinical Trials, Journal of Biopharmaceutical Statistics, 24:2, 211-228,

20. Goldstein DJ, Lu Y, Detke MJ, Wiltse C, Mallinckrodt C, Demitrack MA: Duloxetine in the treatment of depression: a double-blind placebo-controlled comparison with paroxetine. J Clin Psychopharmacol 2004;24: 389-399.
21. Detke MJ, Wiltse CG, Mallinckrodt CH, McNamara RK, Demitrack MA, Bitter I. Duloxetine in the acute and long-term treatment of major depressive disorder: A placebo- and paroxetine-controlled trial. European Neuropsychopharmacology. 2004; 14(6):457-470.

22. Perahia DGS, Wang F, Mallinckrodt CH, Walker DJ. Duloxetine in the treatment of major depressive disorder: A Placebo- and Paroxetine-Controlled Trial. European Psychiatry. 2006; 21 (6):367-378.

23. Mallinckrodt CH, Goldstein DJ, Detke MJ, Lu Y, Watkin JG, Tran PV. Duloxetine: A new treatment for the emotional and physical symptoms of depression. Primary Care Companion J. Clin. Psychiatry. 2003; 5:19-28.

24. Hamilton M: A rating scale for depression. J Neurol Neurosurg Psychiatry 1960, 23: 56-61.

25. SAS Institute Inc. 2008. SAS/STAT® 9.2. User's Guide. Cary, NC: SAS Institute Inc.

26. Little, R.J.A., and Rubin, D.B. (2002). Statistical Analysis with Missing Data, 2nd edition. New York: Wiley.

27. Flemming, TR. Addressing Missing Data in Clinical Trials. Ann Intern Med. 2011;154:113-117.