

A structured approach to choosing estimands and estimators in
longitudinal clinical trials

Peer-reviewed author version

Mallinckrodt, Craig H.; Lin, Q.; Lipkovich, Ilya & MOLENBERGHS, Geert (2012) A structured approach to choosing estimands and estimators in longitudinal clinical trials. In: PHARMACEUTICAL STATISTICS, 11 (6), p. 456-461.

DOI: 10.1002/pst.1536

Handle: <http://hdl.handle.net/1942/14444>

A structured approach to choosing estimands and estimators in longitudinal clinical trials

Mallinckrodt, CH, Lin Q, Carpenter J, Kenward M, Lipkovich I, Molenberghs G

1. Eli Lilly & Co., Lilly Corporate Center, Indianapolis, IN 46285.

2. Medical Statistics Unit, London School of Hygiene & Tropical Medicine, UK

3. I-BioStat, Hasselt University, Diepenbeek, Belgium, and, Katholieke Universiteit Leuven, Leuven,
Belgium.

Abstract

An important evolution in the missing data arena has been recognition of need for clarity in objectives. The objectives of primary focus can often be categorized as assessing efficacy or effectiveness. The present investigation illustrated a structured framework for choosing estimands and estimators when testing investigational drugs to treat the symptoms of chronic illnesses. Key issues were discussed and illustrated using the confirmatory trials from a new drug application in depression. The chosen primary analysis was a common likelihood-based approach to assess efficacy – mean change to the planned endpoint of the trial assuming patients stayed on drug. Secondly, effectiveness was assessed via placebo multiple imputation (pMI) where the imputation model derived from the placebo group was used to impute missing values for both the drug and placebo groups, thereby assuming that drug-treated patients had no benefit from the drug after discontinuing it. Results provided clear evidence of efficacy for the experimental drug and characterized its effectiveness. Data after discontinuation of study medication were not required for these analyses. Given the idiosyncratic nature of drug development, no estimand or approach is universally appropriate. However, the general practice of pairing efficacy and effectiveness estimands may often be useful in understanding the overall risks and benefits of a drug. The family of controlled imputation approaches in which pMI sits is a flexible and transparent framework for formulating primary analyses of effectiveness estimands and sensitivity analyses for efficacy estimands.

Key words: Missing data, Multiple imputation, Maximum likelihood

Introduction

Missing data is an ever present problem in clinical trials, and therefore has been an active area of research. An important evolution in the missing data arena has been in regards to what parameters are of primary interest. In fact, the first recommendation from the recent National Research Council (NRC) guidance on the prevention and treatment of missing data was that the objectives be clearly specified (1).

The need for clarity in objectives is driven by ambiguities arising from the missing data. For example, there is no single definition of a missing value. Values may be intermittently missing; or, missing due to drop out, with nothing known about treatment or outcome past the point of dropout; or, subjects may withdraw from the initially randomized study medication and may or may not be given alternative (rescue) treatment, with or without further measurements being taken. All these scenarios may happen within a single trial, with differing implications for appropriate handling of the data (2).

The consequences of missing data are also situation dependent. For example, in a clinical trial for depression, if a patient is lost to follow up half way through the trial information needed to understand how well the drug worked for that patient is indeed missing. But, in a trial for a treatment to prevent sudden cardiac death due to premature ventricular contractions (PVC), if a patient has sudden cardiac death due to a PVC midway through the trial, data are again incomplete. However, information about how well the treatment worked for that patient is not missing because it is known that the treatment did not work (3).

Knowing that missingness (dropout) may or may not be associated with changes in treatment raises the important point of whether or not, and if so, how, data after discontinuation of the initial therapy should be used in analyses and inference. This topic was discussed at length in the NRC guidance (1). Therefore, the purpose of the present paper is to build upon the NRC guidance to illustrate a structured approach for choosing estimands and estimators in a specific, common clinical trial scenario.

Setting

The setting addressed here is clinical trials for investigational drugs to treat the symptoms of chronic illnesses. The specific example comes from depression, but could also apply to many psychiatric settings, pain, diabetes, and others. Much of the debate on estimands in these settings centers on efficacy versus effectiveness. Consider efficacy as the effects of the drug if taken as directed; that is, the benefit of the drug expected at the endpoint of the trial assuming patients stayed on drug. Consider effectiveness as the effects of the drug as actually taken, recognizing that in the settings addressed here that patients who discontinue a drug are unlikely to have lasting benefit from it (4). Carpenter, Roger, and Kenward refer to hypotheses about efficacy and effectiveness as *de-jure* and *de-facto* hypotheses, respectively (5).

To motivate this discussion we consider the following phase III studies that were included in a New Drug Application (NDA) for an antidepressant. The six studies arose from 3 protocols. 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. Key features of the studies are summarized in Table 1. In addition to studies A and B within each protocol being

identical, protocols 1 and 2 were nearly identical, with the exception that protocol 2 (studies 2A and 2B) included a 6-month double-blind extension following the 8-week acute phase. In addition, protocol 2 was implemented in Eastern Europe whereas the other studies were implemented in the US. References to the original reporting of these studies are included in the Table.

Table 1. Summary of studies.

Study	Treatment Duration	Treatment Arm ^a	Number of Patients ^b	Drug Dose	Completion %
1A [6]	8 weeks	Placebo	89	NA	69.7
		Drug	90	40 mg/day	72.2
		Drug	81	80 mg/day	75.3
		AC	87	20 mg/day	70.9
1B [7]	8 weeks	Placebo	88	NA	61.4
		Drug	84	40 mg/day	70.2
		Drug	86	80 mg/day	63.9
		AC	84	20 mg/day	64.3
2A [8]	8 weeks	Placebo	93	NA	81.7
		Drug	93	80 mg/day	91.4
		Drug	93	120 mg/day	90.3
		AC	85	20 mg/day	89.4
2B [9]	8 weeks	Placebo	99	NA	92.9
		Drug	93	80 mg/day	91.4
		Drug	102	120 mg/day	90.2
		AC	97	20 mg/day	90.7
3A [10]	9 weeks	Placebo	115	NA	77.3
		Drug	121	60 mg/day	69.4
3B [11]	9 weeks	Placebo	136	NA	66.2
		Drug	123	60 mg/day	65.9

^a Drug = experimental drug; AC = active comparator. The same active comparator was used in all studies. More information on the specific drugs tested can be found in the references.

^b Number of patients randomized who had at least 1 postbaseline efficacy observation, and were therefore included in the efficacy analyses.

Choosing Estimands

The NRC guidance on the prevention and treatment of missing data (1) lists the following 5 estimands. The strengths and limitations of each estimand are discussed relative to the example data.

1. *Difference in outcome improvement at the planned endpoint for all randomized*

participants. For estimand 1, data after withdrawal from the initially randomized medication and/or the addition of a rescue medication are included in the analysis (12). The intention-to-treat (ITT) framework is used to compare the initially randomized groups regardless of what treatment subjects actually received.

With ITT, rescue medications can mask or exaggerate both the efficacy and safety effects of the initially assigned treatments (3). Therefore, estimand 1 tests *de-facto* (effectiveness) hypotheses regarding treatment policies (12). However, prior to regulatory approval, causal effects of the investigational drugs are typically the focus, not treatment policies. Therefore, estimand 1 was not appropriate for the present data.

2. *Difference in outcome improvement in tolerators.* This estimand compares the mean outcomes for treatment vs. control in the subset of the population who initially tolerated the treatment. A run-in phase is used to identify patients that meet efficacy and / or safety and tolerability criteria to continue. Those patients that continue are randomized (usually double-blind) to either continue on the investigational drug or switch to control (1).

Randomizing only those patients that initially tolerate the drug would likely lower dropout, thereby reducing the potential for bias. However, estimand 2 draws inference on a subset of patients, but the only way to identify that subset is to treat all patients. Hence, the estimand does not apply to the population treated and is therefore not appropriate for the present data.

3. *Difference in outcome improvement if all subjects tolerated or adhered.* This estimand addresses the expected change if patients adhered to the treatment regimen as directed. It is addressing *de-jure* (efficacy) hypotheses about the causal effects of the initially randomized drug (1).

However, Estimand 3 is a hypothetical parameter – assuming all patients take the drug as instructed, when in fact there will always be patients that can not adhere. Therefore, whenever using estimand 3 it is important to assess the degree to which non-adherence may diminish the actually achieved results. That is, secondary assessments of effectiveness are needed.

4. *Difference in areas under the outcome curve during adherence to treatment.*

5. *Difference in outcome improvement during adherence to treatment.*

These estimands address the effects of treatment on both the outcome measure and the duration of adherence. Hence, there is no missing data due to dropout (1). Estimands 4 and 5 assess *de-facto* hypotheses regarding the initially randomized drug.

However, assessing effects only while a drug is being taken overestimates effectiveness at the planned endpoint, unless the drug has disease modification properties or prolonged activity such

that its benefits persist when the drug is not taken. Given the experimental drug tested in the present data did not have prolonged activity and was not thought to have disease modifying properties, estimand 4 and 5 were not appropriate.

Of the 5 estimands proposed in the NRC guidance (1), estimand 3 is the best choice for the primary analysis of the present data. However, need exists to supplement this efficacy estimand with a secondary effectiveness estimand. None of the five NRC (1) estimands address *de-facto* (effectiveness) hypotheses for the initially randomized medication at the planned endpoint of the trial. Therefore, we propose a 6th estimand which can be particularly relevant in the early evaluations and initial regulatory approvals of new medication in settings similar to those considered here.

6. *Difference in outcome improvement in all randomized patients at the planned endpoint of the trial attributable to the initially randomized medication.* Estimand 6 assesses effectiveness at the planned endpoint of the trial attributable to the initially randomized medications.

The key attributes of the various estimands are summarized in Table 2. Conceptually, as with estimand 1, estimand 6 requires data after withdrawal of randomized study medication until the planned endpoint. However, unlike estimand 1, estimand 6 needs to be free from the confounding effects of rescue medications because inference is to be drawn on the individual randomized study medications, not treatment regimens as in estimand 1. However, ethical considerations often mandate that rescue medications be allowed, particularly after patients discontinue randomized study medication. The issue of how to estimate estimand 6 when the

data needed are problematic to obtain is addressed in the next section. We conclude this section by noting the complementary nature of estimands 3 and 6 for the settings addressed here.

Estimand 3 addresses the primary *de-jure* (efficacy) hypotheses and estimand 6 addresses *de-facto* (effectiveness) hypotheses. Estimands 3 and 6 both make causal inference regarding the initially randomized medication, in all randomized patients, at the planned endpoint of the trial. These estimands will be the focus of the re-analysis of the example data.

Table 2. Proposed estimands and their key attributes.

Estimand	Hypothesis	Inference	Population	Endpoint	Use of data after withdrawal of randomized study medication
1	<i>de-facto</i> (effectiveness)	Treatment policy	All patients	Planned endpoint	Included in primary analysis
2	<i>de-jure</i> (efficacy)	Initially randomized medication	Tolerators	Planned endpoint	Not included in primary analysis
3	<i>de-jure</i> (efficacy)	Initially randomized medication	All patients	Planned endpoint	Not included in primary analysis
4	<i>de-facto</i> (effectiveness)	Initially randomized medication	All patients	Undefined	Not included in primary analysis
5	<i>de-facto</i> (effectiveness)	Initially randomized medication	All patients	Undefined	Not included in primary analysis
6	<i>de-facto</i> (effectiveness)	Initially randomized medication	All patients	Planned endpoint	Likely imputed

Choice of Estimators

Estimators for estimand 3 are well-established in the literature. Methods valid if data are missing at random, such as direct likelihood, multiple imputation, and weighted generalized estimating equations provide an unbiased estimate of estimand 3 if missingness arises from a missing at random mechanism (1, 4, 5). The assumption of MAR versus MNAR is untestable because we do not have the missing data about which the assumptions are made, therefore sensitivity analyses must be used to assess robustness of inferences to the assumption of MAR (1, 4, 5).

Conceptually, estimand 6 is a weighted average of two components. The first component is the treatment effect at endpoint in those who adhere to study medication – the same quantity being estimated by estimand 3. The second component is the treatment effect at endpoint in those patients that discontinued study medication. The weights for the two components are the proportions of patients that complete the trial on study medication and the proportion of patients that withdraw from study medication.

In estimating estimand 6, a common approach that circumvents the confounding effects of follow-up data has been to impute the missing data using baseline observation carried forward (BOCF). With BOCF, it is assumed that patients who discontinue drug received no lasting benefit, so the change from baseline after stopping study medication should be zero and thus the change scores after discontinuation should equal the baseline values.

However, the assumption that patients' condition would return to the baseline state after ceasing study medication is questionable in many situations as study effects, placebo effects, and natural

evolution of the disease also influence outcomes. In addition, BOCF is a single imputation approach that does not account for the uncertainty of imputation. Therefore, BOCF is generally not an acceptable approach (1).

Alternatively, the placebo group provides an estimate of no pharmacological benefit of the drug that does reflect the study effect, the placebo effect, and the evolution of the disease. Hence, information from the placebo group may provide a better estimate of estimand θ_6 than BOCF

Carpenter, Roger and Kenward (5) defined and illustrated a family of multiple imputation based approaches for assessing sensitivity in testing efficacy and effectiveness hypotheses. Using the placebo group as the basis for imputing missing values for both the placebo and drug groups is a specific form of their “jump to reference” approach, indicating that after dropout the trajectory of patients is assumed to follow that of a reference group, in this case placebo. A recent simulation study concluded that the so-called placebo multiple imputation (pMI) approach provided unbiased estimates of estimand θ_6 (13).

Analyses of Example Data

The primary focus of the re-analysis, as for the original analyses of these studies, was on efficacy. Specifically the primary objective was to compare drug versus control in mean change from baseline to the planned endpoint visit (estimand 3). The secondary focus of the re-analysis was on effectiveness. That is, to compare drug versus control in mean change from baseline to the planned endpoint assuming that patients who discontinued study medication prior to the endpoint had no benefit from it (estimand 6).

The primary analysis for estimating the primary estimand was a restricted maximum likelihood (REML)-based repeated measures approach. 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. Significance tests were based on least-squares means using a two-sided $\alpha=0.05$ (two-sided 95% confidence intervals). Analyses were implemented using SAS PROC MIXED (14). The primary comparison was the contrast between treatments at the endpoint visit.

The secondary effectiveness estimand was estimated using pMI. Multiple imputation was used to replace missing outcomes for both drug- and placebo-treated patients using an imputation model derived from placebo-treated patients. The imputation model included changes from baseline in the primary efficacy outcome. The endpoint contrast and associated confidence interval and p values were obtained from the multiple completed datasets using the same

approach as for the primary analysis. Rubin's rules [15], as implemented in SAS PROC MIANALYZE (14) were used to combine results and draw inference from the multiple data sets. Additional details on implementing pMI have been reported elsewhere (13)

Results from the direct likelihood (DL) and pMI analyses of each study are summarized in Table 3. All T scores of 2.0 or greater yielded p values less than .05. Using traditional frequentist inference based on 2-tailed $p < .05$, the direct likelihood analyses to assess estimand 3 yielded 10 of 14 significant contrasts versus placebo. The 40mg dose was significant in 1 of 2 contrasts. The AC was significant in 2 of 4 contrasts. Doses of 60mg, 80mg, and 120mg, were significant in 7 of 8 contrasts.

The pMI analyses of estimand 6 yielded 6 of 14 significant contrasts. The 40mg dose was significant in 0 of 2 contrasts. The AC was significant in 1 of 4 contrasts. Doses of 60mg, 80mg, and 120mg, were significant in 5 of 8 contrasts.

Table 3. Results from direct likelihood estimates of estimand 3 and placebo multiple imputation estimates of estimand 6.

Study	Arm	Direct Likelihood			Placebo MI			T	Fraction ²
		Delta ¹	SE	T	Delta ¹	SE	T		
1A	40	1.31	0.95	1.38	1.22	1.10	1.11	0.93	
1A	80	1.52	0.99	1.54	1.29	1.12	1.15	0.85	
1A	AC	2.44	0.96	2.54	1.99	1.11	1.79	0.82	
1B	40	2.79	1.07	2.61	2.1	1.09	1.92	0.75	
1B	80	3.53	0.98	3.60	2.91	1.05	2.78	0.82	
1B	AC	1.37	1.23	1.11	0.99	1.13	0.88	0.72	
2A	80	2.18	0.66	3.30	1.94	0.73	2.67	0.89	
2A	120	3.34	0.7	4.77	3.03	0.78	3.88	0.91	
2A	AC	2.9	0.73	3.97	2.71	0.79	3.45	0.93	
2B	180	1.34	0.66	2.03	1.3	0.71	1.84	0.97	
2B	120	1.69	0.66	2.56	1.7	0.71	2.41	1.00	
2B	AC	1.16	0.68	1.71	1	0.71	1.41	0.86	
3A	60	4.86	0.96	5.06	4.22	1.09	3.86	0.87	
3B	60	2.16	0.94	2.30	1.7	1.00	1.7	0.79	

1. Delta = the endpoint contrast, the difference in LSMENAS between drug and placebo and the predefined study endpoint.

2. Fraction = the delta obtained from pMI divided by the delta from the primary analysis.

Given the lack of experience with pMI and that the trials were not powered for secondary effectiveness objectives, it is useful to look beyond statistical significance versus placebo and to benchmark results of the experimental drug versus the active comparator. Mean changes to endpoint for AC ranged from 0.99 to 2.71, with an average of 1.55. Results for the various doses of the experimental drug ranged from 1.22 to 4.22, with an average of approximately 2.1.

For the four AC arms the estimate of effectiveness ranged from 72% to 93% of the magnitude of estimand 3, with the average being approximately 83%. For the experimental drug estimates of effectiveness ranged from 75% to 100% of the magnitude of estimand 3, with an average of approximately 88%.

The average SE from pMI was approximately 10% larger than from direct likelihood. This difference, along with the smaller mean changes, explains why four contrasts that were significant for estimand 3 only approached significance for estimand 6.

Protocol 2 had higher completion percentages than other protocols, and not surprisingly, also had effectiveness estimates that were closer to the corresponding efficacy estimates. Protocol 2 included an extension phase after the acute treatment period and was conducted in Eastern Europe, whereas the other studies did not have extension phases and were conducted in the US. It is not clear whether these design differences caused the higher completion percentages.

Discussion

An important evolution in the debate on missing data has been recognition of need for clarity in objectives. The objectives of primary focus can often be categorized generally as either assessing efficacy or effectiveness. Recent guidance emphasizes collecting follow-up data on patients who withdraw from randomized study medication prior to the planned endpoint of the trial (1, 12). These recommendations further state that the follow up data should be included in analyses. Although the follow up data can mask or enhance the risks and benefits caused by the initially randomized medications, these confounds are not problematic when inference is on the treatment regimens rather than the initially randomized medications. However, in many situations the causal effects of the initially randomized medications are the primary focus. As O'Neill and Temple (16) noted, including follow-up data as part of the primary estimand is more customary in outcomes trials, whereas in symptomatic treatment trials follow-up data are usually not included in the primary estimand, for the previously noted reasons.

For the example data, estimand 3 from the NRC guidance (1) was chosen as the primary analysis. This estimand assessed the difference in outcome improvement at the planned endpoint assuming that all patients adhered to the study medications. An effectiveness estimand not discussed in the NRC guidance (1) was proposed for secondary analyses. This estimand assessed the difference in outcome improvement at the planned endpoint assuming that all patients who did not adhere had no benefit from the drug.

Both estimands focused on causal effects of the initially randomized medications in all patients. Estimand 3 assessed what would have happened if patients took the medications as directed. Estimand 6 assessed what was actually observed. Given the confounding effects of follow up

medications and the ethical need to allow them, estimand 6 was estimated using pMI, which imputed the data after discontinuation of the initially randomized study medication assuming that the experimental drug had no benefit after it was discontinued, that is, to assume the statistical behavior of drug treated patients became that of placebo treated patients after discontinuation of study drug (or addition of rescue medication).

Estimand 6 is actually not new, but its usefulness has been limited in that it has typically been coupled with suboptimal estimators. The most common means of estimating estimand 6 has been BOCF, which is well-known to entail assumptions that are unlikely to hold and to underestimate the uncertainty of imputation (1). Therefore, the alternative pMI analytic approach has been proposed, tested, and proven useful in situations where it can be assumed that the drug has no benefit after patients stop taking it (13).

The pMI approach is a specific member of the more general family of controlled multiple imputation approaches (5). The assumptions of pMI are transparent and open to debate and assessment, with imputations conducted in a principled manner using commercially available software.

With a suitable estimator now available, the 6th estimand (effectiveness at endpoint due to the initial medication) may be considered in those situations where inferences regarding effectiveness of the initially randomized medications are sought. It was particularly useful to couple estimand 3 as the primary objective with estimand 6 being secondary to more fully understand the causal effects of the investigational drug in the example data.

Results from the confirmatory trials in the example NDA provided clear evidence of efficacy for the experimental drug (estimand 3). In addition, effectiveness results were easy to interpret in that the average advantage of drug over placebo for effectiveness (estimand 6) was approximately $7/8$ the magnitude of the average advantage in efficacy, and at least on par with the active comparator included in 4 of the trials. Although data after discontinuation of study medication until the planned trial endpoint may have secondary uses, such follow-up were not required for the primary analyses of either estimand 3 or 6.

The findings of this investigation need to be interpreted in light of several caveats and considerations. Given the diversity in clinical settings, no one estimand or pair of estimands is universally appropriate. Hence, the example presented here was not intended as an overall recommendation. Rather, it was an illustration specific to clinical trials for investigational drugs to treat the symptoms of chronic illnesses, such as depression, pain, or diabetes.

In addition, using pMI to estimate estimand 6 will not always be appropriate. For example, drugs that have disease modifying properties or long half-lives (e.g., antibodies) may have effects that persist after discontinuation of medication. Therefore, the assumption of pMI that the statistical behavior of drug-treated patients becomes that of placebo-treated patients immediately after discontinuation would not be valid. However, the family of controlled imputation approaches in which pMI sits is easily adaptable to these situations. Moreover, the transparency of assumptions makes it easy to understand whether pMI or other implementations from the controlled imputation family are appropriate for each specific scenario.

Lastly, pMI and the family of approaches wherein it sits are fairly new. Therefore, more experience in actual clinical trial settings would be useful, along with additional study to better understand and/or correct the standard errors and confidence intervals from pMI that tend to overestimate uncertainty.

References

1. 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 Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
2. Mallinckrodt, CH. and Kenward, MG. Conceptual Considerations Regarding Choice of Endpoints, Hypotheses, and Analyses In Longitudinal Clinical Trials. *Drug Information Journal*. **43(4)**: 449-458.
3. Permutt and Piniero. Dealing with the missing data challenge in clinical trials. *Drug Information Journal*. **43(4)**: 449-458.
4. Mallinckrodt CH, Lane PW, Schnell D, Peng U, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal*. 2008; 42:305-319.
5. Carpenter Roger Kenward. Analysis of Longitudinal Trials with Missing Data:| A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation. Submitted.
6. Nemeroff CB, Schatzberg AF, Goldstein DJ, [Detke MJ, Mallinckrodt C, Lu Y, Tran PV](#): Duloxetine for the treatment of major depressive disorder. *Psychopharmacol Bull* 2002;36: 106-132.

7. 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.
8. 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. *Eur Neuropsychopharmacol* 2004;14: 457-470.
9. Perahia DG, Wang F, Mallinckrodt CH, Walker DJ, Detke MJ: Duloxetine in the treatment of major depressive disorder: a placebo- and paroxetine-controlled trial. *Eur Psychiatry* 2006;21: 367-378.
10. Detke MJ, Lu Y, Goldstein DJ, [Hayes JR, Demitrack MA](#): Duloxetine, 60 mg once daily, for major depressive disorder: a randomized double-blind placebo-controlled trial. *J Clin Psychiatry* 2002;63: 308-315.
11. Detke MJ, Lu Y, Goldstein DJ, [McNamara RK, Demitrack MA](#): Duloxetine 60 mg once daily dosing versus placebo in the acute treatment of major depression. *J Psychiatr Res* 2002;36: 383-390.
12. Fleming, TR. Addressing Missing Data in Clinical Trials. *Ann Intern Med.* 2011;154:113-117.
13. Teshome B, Lipkovich I, Molenberghs, G, Mallinckrodt C. A Multiple Imputation Based Approach to Sensitivity Analyses and Effectiveness Assessments in Longitudinal Clinical Trials. Accepted. *J Biopharm. Stats.*

14. SAS Institute Inc. SAS/STAT® User's Guide, Version 9.1, SAS Institute Inc.: Cary, NC, 2003.
- 15/ Rubin DB. Multiple Imputation for Nonresponse in Surveys. Wiley: New York, 1987 p. 75.
16. 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