A multiple-imputation based approach to sensitivity analysis and effectiveness assessment in longitudinal clinical trials

TESHOME AYELE, Birhanu; Lipkovich, Ilya; MOLENBERGHS, Geert & Mallinckrodt, Craig H. (2014) A multiple-imputation based approach to sensitivity analysis and effectiveness assessment in longitudinal clinical trials. In: Journal of Biopharmaceutical Statistics, 24 (2), p. 211-228.

DOI: 10.1080/10543406.2013.859148
Handle: http://hdl.handle.net/1942/18606

A Multiple Imputation Based Approach to Sensitivity Analyses and Effectiveness Assessments in Longitudinal Clinical Trials

Teshome, Birhanu [1], Ilya Lipkovich Ph.D.[2] , Geert Molenberghs, Ph.D.[1,3], Craig H. Mallinckrodt, Ph.D.[2]

1. I-BioStat, Universiteit Hasselt B-3590 Diepenbeek, Belgium
2. Eli Lilly & Co., Lilly Corporate Center, Indianapolis, IN 46285.
3. I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

Running Head:
Multiple Imputation to Assess Sensitivity and Effectiveness

Abstract

It is important to understand the effects of a drug as actually taken (effectiveness) and when taken as directed (efficacy). The primary objective of this investigation was to assess the statistical performance of a method referred to as placebo multiple imputation (pMI) as an estimator of effectiveness and as a worst reasonable case sensitivity analysis in assessing efficacy. The pMI method assumes the statistical behavior of placebo- and drug-treated patients after drop out is the statistical behavior of placebo-treated patients. Thus, in the effectiveness context pMI assumes no pharmacological benefit of the drug after dropout. In the efficacy context pMI is a specific form of a missing not at random analysis expected to yield a conservative estimate of efficacy. In a simulation study with 18 scenarios the pMI approach generally provided unbiased estimates of effectiveness and conservative estimates of efficacy. However, the confidence interval coverage was consistently greater than the nominal coverage rate. In contrast, LOCF and BOCF were conservative in some scenarios and anti-conservative in others with respect to efficacy and effectiveness. As expected, direct likelihood (DL) and standard multiple imputation (MI) yielded unbiased estimates of efficacy and tended to over-estimate effectiveness in those scenarios where a drug effect existed. However, in scenarios with no drug effect, and therefore the true values for both efficacy and effectiveness were zero, DL and MI yielded unbiased estimates of efficacy and effectiveness.

Key words: Missing data, longitudinal analyses, multiple imputation

Introduction

It has been debated whether the primary analysis in longitudinal clinical trials should focus on efficacy or effectiveness. An important aspect of this debate is the impact of missing data arising from patient discontinuation. Missing data in clinical studies remains an active area of investigation, with entire issues of journals [1] and entire text books [2] devoted to the topic. The meaning and consequences of missing data depends on the situation [3]. The setting addressed here is that of phase II or phase III clinical trials for investigational drugs to treat the symptoms of chronic illnesses, such as depression, pain, or diabetes.

In such settings efficacy may be viewed 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, counter to the fact that some dropped out. Effectiveness in these same settings may be viewed as the effects of the drug as actually taken, recognizing that patients who discontinue the drug, particularly because of safety or tolerability issues, are unlikely to have lasting benefit from it. Carpenter, Roger, and Kenward refer to hypotheses about efficacy and effectiveness as the *de-jure* and *de-facto* hypotheses, respectively [4].

It is important to understand both what happens when a drug is evaluated as actually taken and when taken as directed, especially when including safety assessments in the scope of inference. And while it is important to consider when to put the greatest emphasis on which research question [5], the present discussion focuses on what endpoints and analyses are most appropriate for each. The following table, which borrows heavily from introductory chapters in the recent National Academy of Science guidance on the prevention and treatment of missing data [6] summarizes the estimands and estimators that may be associated with efficacy and effectiveness hypotheses.

Effectiveness of the initial randomized medication at the planned endpoint of the trial is essentially the maintained benefit at the planned endpoint attributable to the randomized medication for the period of time in which it was taken. For testing this hypothesis, it is not adequate to assess patients only until they drop out of the trial, follow up data from the time of dropout until the planned endpoint of the trial are needed. However, ethical considerations often mandate that alternative medication be allowed after patients discontinue randomized study medication.

In the intention-to-treat (ITT) framework where inference is drawn based on the originally assigned treatment, including follow-up data when alternative medications are allowed can mask or exaggerate both the efficacy and safety effects of the initially assigned treatments, thereby invalidating causal inferences for the effectiveness of the originally assigned medication [5]. Therefore, it has been proposed in the NAS guidance [6[and elsewhere [7], that the hypothesis of interest is that of a treatment regiment, that is, initiating treatment with a particular intervention. However, the treatment regiment hypothesis is not useful in the situations of interest here as it is unlikely an investigational medication can be approved for use as part of a regiment unless it has first been proven safe and effective on its own.

A number of techniques have been used to impute the missing (follow up) data to circumvent problems from the confounded follow-up data. Last and baseline observation carried forward (LOCF and BOCF) are perhaps the two most commonly used methods. Although the acronyms imply truly carrying observations forward in time, an LOCF result can be interpreted as either the change observed while actually taking drug, or as the change to the designed endpoint of the trial assuming the patients' condition would not have changed after discontinuing the drug. 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 values 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 time evolution also influence outcomes. Therefore, if patients receive no pharmacological benefit from a drug, either because it has no effect or because they discontinue taking the medication, their outcomes would be equal to their baseline values only if the study effect and the placebo effect were zero.

Alternatively, the placebo group provides an estimate of no pharmacological benefit of the drug that reflects the study effect and placebo effect. Hence, information from the placebo group may provide a better estimate of effectiveness for patients who discontinue drug than using patients' last or baseline observation.

Carpenter, Roger and Kenward [4] define and illustrate a family of multiple imputation based approaches for assessing sensitivity in testing *de-jure* (efficacy) and *de-facto* (effectiveness) hypotheses. Using the placebo group to impute missing values for both the placebo and drug groups is a specific form of their "jump to reference" approach. Although the principles and assumptions underlying the jump to reference approach are clear and easy to understand [4], the performance of the method has not been rigorously evaluated.

Therefore, the primary objective of this investigation was to assess the statistical performance of a method we refer to as placebo multiple imputation (pMI) (a specific form of jump to reference), as an estimator of effectiveness (as actually taken hypothesis). The behavior of pMI was also considered in the context of a sensitivity analysis in testing the efficacy (taken as directed) hypothesis. In this context, pMI assumes the statistical behavior of drug-treated patients after drop out is the statistical behavior of placebo-treated patients. Thus, pMI can also be interpreted as a specific form of a missing not at random analysis expected to yield a conservative estimate of efficacy.

2. Motivating example

A clinical trial in major depressive disorder, originally reported by Detke et al [8], is used as a motivating example for the present research. The primary objective in that trial was to compare the efficacy of an experimental antidepressant with placebo to support a New Drug Application. As such, this was a phase III (confirmatory) trial. Patients were

randomly assigned (1:1 ratio) to placebo (n=139) or the experimental drug (n=128), with the double-blind treatment period lasting 9 weeks. Study visits were scheduled once a week for the first 3 weeks after randomization, and every two weeks thereafter. The experimental drug was found to be significantly superior to placebo on the a priori declared primary efficacy analysis (direct likelihood-based repeated measures) of mean change to endpoint on the $HAMD_{17}$ total score ($p = 0.025$). The completion rates were 64.7% for placebo compared with 60.9% for the experimental drug. The rates of dropout due to adverse events were 4.3% for placebo and 12.5% for the experimental drug, while the corresponding rates of dropout due to lack of efficacy were 13.7% and 5.5%, respectively.

Although results were significant based on the primary analysis, it was reasonable to wonder how effective the medication would be in actual practice given the rates of dropout in the trial, and to what degree missing data might have biased the estimate of treatment efficacy.

3. Methods

Description of pMI

Placebo multiple imputation estimates visitwise means or mean changes assuming that the statistical behavior of drug treated patients who discontinue becomes that of placebo-treated patients after the time of dropout. Two views may be taken of this estimand: 1) as an assessment of effectiveness, assuming patients who discontinue before the endpoint receive no pharmacological benefit after dropout; and 2) as a worst reasonable case assessment of efficacy – the outcome that would have been observed had the patient stayed on drug.

To implement this approach, multiple imputation was used to replace missing outcomes for drug-treated subjects who discontinued using multiple draws from the posterior predictive distribution estimated from subjects who were randomized to the placebo arm in that same trial.

To set up the imputation model, define observed subject-specific covariates ($\mathbf{X}$) and partially observed outcomes ($\mathbf{Y}_{obs}$) whose joint distribution drive the imputation mechanism for missing outcomes. Let $\mathbf{y}_i=\{\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}\}$, the 1xT outcome vector containing for the i-th subject k_i observed outcomes and $T-k_i$ unobserved outcomes;

And, $\mathbf{x}_i$ is a 1xP vector of fully observed covariates.

Most missing values in the clinical trial settings addressed here are caused by dropouts resulting in a monotone pattern of missingness. Therefore, Bayesian regression employing factorization of the multivariate normal density for the data with monotone misingness pattern [10, pp. 167-167], such as is available is SAS PROC MI [9], provides an easy and fast way to impute the missing values. The basic idea is to estimate the parameters for the imputation model using only data from the placebo arm and then use

those parameters to impute missing values for both the drug-treated and placebo-treated patients. Partially observed outcomes from treated subjects are used when imputing their missing outcomes as follows:

Data are processed sequentially by repeatedly calling SAS PROC MI to impute missing outcomes at visits t=1,..,T.

1. *Initialization*. Set t=0 (baseline visit)
2. *Iteration*. Set t=t+1. Create a data set combining records from placebo and treated subjects with columns for covariates **X** and outcomes at visits 1,..,t with outcomes for all treated subjects set to missing at visit t and set to observed or imputed values at visits 1,..,t-1.
3. *Imputation*. Run Bayesian regression in SAS proc MI on this data to impute missing values for visit t using previous outcomes for visits 1 to t-1 and baseline covariates. Note that only placebo data will be used to estimate the imputation model since no outcome is available for treated subjects at visit t.
4. Replace imputed data for all treated subjects at visit t with their observed values, whenever available. If t < T then go to Step 2, otherwise proceed to Step 5.
5. Repeat steps 1-4, m=10 times with different seed values to create m imputed data sets.
6. *Analysis*. For each completed data set, evaluate treatment difference at the last scheduled visit, T, using the same repeated measures model as would have been applied had the data been complete,
7. *Combined Inference*. Compute pMI-based estimate and associated confidence interval CI for the treatment contrasts at last scheduled visit using Rubin's combining rules [10, p. 75], as implemented in SAS PROC MIANALYZE.

Simulation study

A simulation study was conducted to assess the properties of pMI. Although the simulations were not intended to mimic any particular clinical setting, many input parameters for the simulation study were taken from the depression trial used as motivating data. Key details of the simulation are summarized below. .

Results from pMI were compared with results from LOCF, BOCF, direct likelihood, (DL) and standard multiple imputation (MI) in 18 scenarios that were arranged as a 2 x 3 x 3 factorial. Focus was on comparing pMI vs. BOCF and LOCF in regards to the effectiveness estimand whereas focus was on comparing pMI vs. DL and MI in regards to the efficacy estimand.

Scenarios included two trajectories of patient response: 1) Improvement (IMP), where the mean trends were for patients to improve over time, such as would often be the case for symptomatic treatments of chronic illnesses; 2) Worsening (WOR), where the mean trends were for patient to worsen over time, such as would often be the case for disease modification treatments in progressive illnesses such as Alzheimer's disease. Scenarios also included 3 dropout patterns, all with an overall dropout rate of 30%: 1) equal rates (30%) in the drug and placebo groups (=); 2) higher dropout in the drug group (HD), 40% dropout in the drug group vs. 20% in the placebo group; 3) higher dropout in the placebo

group (HP), 20% dropout in the drug group vs. 40% in the placebo group. Lastly, scenarios included three levels of treatment effects.

Dropout was simulated by deleting values according to a logistic model relating probability of dropout with changes from baseline in the simulated efficacy outcomes. Specific values for the logistic model were chosen so as to yield the desired dropout rates in the various scenarios. Of particular note, however, is that the dropout mechanism was missing at random

For efficacy, the difference between drug and placebo in mean change to endpoint was a standardized effect size (ES) of 0.5, 0.3, or 0.0. For effectiveness, the mean difference at endpoint resulted from a mixture distribution where the effect size of completes was 0.5, 0.3, or 0.0, as described for efficacy and the effect size for patients who dropped out was 0.0. Therefore, the true value for the endpoint contrasts was the weighted mean of the two groups. For example, with ES = 0.5 the true advantage of drug over placebo for efficacy = 2.75. For effectiveness, the corresponding true values with 20%,.30%, and 40% dropout in the drug group were 2.24, 2.05, and 1.68. The true values for the placebo group are summarized in Table 2. Given these trajectories, the assumptions for BOCF and LOCF were not valid.

4. Results

Simulation results were summarized in terms of Bias, Relative Bias, Variance, Mean Square Error (MSE), Confidence Interval (CI) coverage and Rejection rates. Bias was defined as the difference between the mean estimate and the true value of the parameter. Positive bias indicated average estimate of treatment contrasts smaller than true value when ES > 0 and average contrast favoring placebo when ES = 0. Negative bias indicated average estimate of treatment contrasts larger than true value when ES > 0 and average contrasts favoring drug when ES = 0.

4.1. Efficacy

Tables 3 through 8 summarize results from tests of the efficacy hypothesis. As expected, DL and MI provided unbiased estimates of efficacy with confidence interval (CI) essentially equal to the nominal rates. In contrast BOCF and LOCF were biased, with the direction of bias varying by scenario, leading to poor CI coverage and large MSE.

The bias in estimates from pMI was generally smaller than the bias in BOCF and LOCF. In 17/18 scenarios the bias from pMI was conservative as the mean estimate of efficacy was smaller than the corresponding true value; in the 18^{th} scenario pMI was essentially unbiased.

In addition, the variance in estimates from LOCF and BOCF was lower than from DL or MI. The variance in estimates from pMI was generally intermediate to those from LOCF / BOCF vs. DL and MI. The variance in estimates from pMI varied according to how much drug group data was replaced by placebo data; as the proportion of drug treated

data being replaced by placebo increases the sample becomes more homogonous and variance in treatment contrasts decreases.

Power for pMI was close to the power from DL and MI when dropout rate was equal in the drug and control groups (DO =) or when dropout was higher on placebo (HP); however, when dropout was higher on drug (DO = HD) power from pMI was appreciable lower than from DL or MI.

When ES =0 BOCF and LOCF provided the desired control of false positive (FP) results in only 2 of 6 scenarios, with at least triple the desired rate of FP (2.5%) in 4 of 6 scenarios, with maximum rates of 64% for BOCF and 34% for LOCF. DL and MI always provided the desired control. For pMI, the FP rate was always lower than the desire rate of 2.5%.

Effectiveness

The variance in estimates and rejection rates apply both efficacy and effectiveness as the same analysis is interpreted in two contexts that only vary by what is considered the true value for the treatment difference. Therefore, Table 5 and Table 8 summarize the variance in estimates and the rejection rates for efficacy and effectiveness estimands. Bias, relative bias, MSE and CI coverage for the effectiveness estimand are summarized in Tables 9 – 12, respectively.

Regarding bias, pMI had minimal to no bias in all scenarios; BOCF and LOCF had large biases in most all scenarios, with the bias in BOCF favoring drug effectiveness in 5/18 scenarios. Although DL and MI were biased in favor of drug whenever ES > 0 (because effectiveness < efficacy). DL and MI unbiased when ES = 0 because in these scenarios the true value for effectiveness = the true value for efficacy = 0.

In addition, MSE for pMI was fairly consistent across scenarios and often smaller then corresponding MSE for other methods. In contrast MSEs from the other methods varied across scenarios and were often greater than the MSE from pMI.

The range in CI coverage (Table 12) for the effectiveness estimand ranged from 2% to 85% for BOCF, from 24% to 81% for LOCF, from 76% – 95% for DL, from 77% to 95% for MI, and from 98% - 99% for pMI.

Additional results

Results from estimates and tests of the main effect of treatment are shown in Appendix A (Table A1-A10). These results generally agreed with results from the endpoint contrast.

5. Results from example data

Results from analyses of the actual clinical trial data are summarized in Table 13. The mean change to the endpoint visit on placebo was approximately 8 points compared with approximately 10 points on drug. Therefore, as is typically the case in depression clinical trials, an appreciable placebo response was observed, thereby invalidating the assumptions for BOCF and LOCF.

The endpoint contrast from pMI was 1.54 compared with 1.08 from BOCF, 2.13 from DL, 2.09 from MI, and 1.75 from LOCF. Therefore, in the effectiveness context, the pMI result suggested that the effectiveness of the drug was approximately 73% the magnitude of the efficacy, as estimated by DL and MI.

In the efficacy context, the pMI result can form the lower bound, or worst reasonable estimate of efficacy, to be combined with other sensitivity analyses to define a "region of ignorance". That is, a region wherein the true value almost certainly lies, but exactly where is not certain.

The SE from pMI was slightly less than the SE from MI and DL and greater than the SE from BOCF and LOCF. Given that the CI coverage in the simulation results for effectiveness was greater than the nominal coverage, the marginally significant p value from pMI is of less interest. However, it is relevant to note that the p value from pMI was considerably smaller than the p value from BOCF.

6. Discussion

To our knowledge, the simulation study in the present research is the first rigorous evaluation of pMI, a specific form of the jump to reference imputation approach detailed by Carpenter, Roger, and Kenward [4]. The pMI approach generally provided unbiased estimates of effectiveness in these simulations where there was no benefit from drug after discontinuation. However, the confidence interval coverage was consistently greater than the nominal coverage rate. In addition, pMI yielded conservative estimates of efficacy in all scenarios, whereas LOCF and BOCF were conservative in some scenarios and anti-conservative in others. As expected, DL an MI yielded unbiased estimates of efficacy and tended to over-estimate effectiveness in those scenarios where a drug effect existed. However, in those scenarios where there was no drug effect, and therefore the true values for both efficacy and effectiveness were zero, DL and MI yielded unbiased estimates of efficacy and effectiveness.

These results should be viewed in light of the strengths and limitations of the present investigation. With 18 scenarios, the simulation study was, on the one hand, comprehensive, but still narrow in scope relative to the vast array of clinical situations. Moreover, several implementations of pMI may be worth considering. For example, we also considered an imputation approach similar to what Carpenter, Roger, and Kenward refer to as "copy to reference". In our implementation of copy to reference only baseline severity was used in the imputation model and all postbaseline data were replaced by

imputed values for those patients that dropped out. Detailed results are not reported as this method did not provide unbiased or nearly unbiased effectiveness estimates in all scenarios. Independent replication of the simulation results and more experience with pMI in actual settings would be useful.

In the present work, reasons for dropout were not differentiated as all dropouts resulted from the same model. In practice, it would be useful to separately consider dropouts and their impact by reasons for dropout. For example, it has been suggested that dropouts due to adverse events were the main area of concern and that methods like DL or MI provided reasonable estimates of effectiveness for other reasons of discontinuation [11]. Therefore, rather than applying pMI to all dropouts, it may be useful to impute missing values from the placebo group only for drug treated patients that drop out due to adverse events. However, for initial assessments, the approach in the present work of applying pMI to all drug treated dropouts was useful in that it tested the method with high and differential rates of dropout, thereby allowing assessment of performance under extreme conditions.

Given these results, further investigation of pMI in scenarios not covered in the present work is warranted and use of pMI as an a priori specified sensitivity analysis in situations similar to those investigated in this study is justified.

References

1. Drug Information Journal (Volume 43, Number 4, July, 2009)

2. Molenberghs G, Kenward MG. Missing Data in Clinical Studies. Chichester: John Wiley & Sons; 2007.

3 Permutt T and Piniero J. Dealing with the missing data challenge in clinical trials. *Drug Information Journal*. **43(4)**: 449-458.

4. Carpenter J, Roger J, Kenward M. Analysis of Longitudinal Trials with Missing Data:|
A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation. Submitted for publicatoin.

5. 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.

6. National Research Council. 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: 2010; The National Academies Press.

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

8. 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.

9. SAS Institute Inc. SAS/STAT® User's Guide, Version 9.1, SAS Institute Inc.: Cary, NC, 2003.

10. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, 1987.

11. Kim, Yongman. Missing Data Handling in Chronic Pain Trials. Journal of Biopharmaceutical Statistics. 2011; 21: 2, 311 — 325

Table 1. Estimands and estimators commonly used to assess efficacy and effectiveness in clinical trials

Hypothesis	Estimands	Estimators[1]	Data included in analysis
Efficacy	Mean change to planned endpoint	DL	Observed data while on drug
		MI	Observed data while on drug
		ANOVA	Observed data while on drug + LOCF imputation
Effectiveness	Mean change to last observation	ANOVA	Observed data while on drug
	Mean change to planned endpoint	ANOVA	Observed data while on drug + BOCF imputation
	Mean change to planed endpoint	ANOVA	Observed data while on drug + follow up data (rescue meds not allowed)
Treatment regiment	Mean change to planned endpoint	Various	Observed data while on drug + follow up data (rescue meds allowed)

1. DL = direct likelihood. MI = multiple imputation

Table 2. Visit wise means in the placebo group (Placebo population means)

Visit	Improvement	Worsening
1	18.8	18.8
2	16.8	20.8
3	14.8	22.8
4	12.8	24.8
5	10.8	26.8

Table 3. Bias in estimates of the efficacy estimand

			Bias in mean estimates of change from baseline					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	1.4311	1.3290	0.0429	0.0445	0.7641	0.0493
		HD	2.3137	1.9949	0.0395	0.0392	0.9743	0.0493
		HP	0.7669	0.8765	0.0396	0.0408	0.4768	0.0493
IMP	0.3	=	0.8764	0.8148	0.0441	0.0468	0.5242	0.0493
		HD	1.6901	1.4217	0.0381	0.0405	0.6499	0.0493
		HP	0.2282	0.3616	0.0367	0.0419	0.3228	0.0493
IMP	0	=	0.0123	0.0187	0.0386	0.0437	0.1048	0.0493
		HD	0.6806	0.5121	0.0408	0.0387	0.1324	0.0493
		HP	-0.6653	-0.4697	0.0372	0.0365	0.0667	0.0493
Worse	0.5	=	1.6940	1.4418	0.0339	0.0353	0.7058	0.0493
		HD	0.8134	1.1669	0.0427	0.0416	0.8993	0.0493
		HP	2.1932	1.6053	0.0445	0.0506	0.4961	0.0493
Worse	0.3	=	0.9985	0.8653	0.0274	0.0368	0.4632	0.0493
		HD	0.1219	0.5792	0.0405	0.0434	0.6042	0.0493
		HP	1.6321	1.0841	0.0441	0.0520	0.2075	0.0493
Worse	0	=	0.0079	0.0220	0.0325	0.0339	0.1152	0.0493
		HD	-0.8361	-0.2693	0.0422	0.0372	0.1231	0.0493
		HP	0.8393	0.3022	0.0459	0.0447	-0.0286	0.0493

Table 4. Relative Bias in estimates of the efficacy estimand

			Relative bias in mean estimates of change from baseline					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	0.5195	0.4825	0.0156	0.0162	0.2774	0.0179
		HD	0.8400	0.7242	0.0144	0.0142	0.3537	0.0179
		HP	0.2784	0.3182	0.0144	0.0148	0.1731	0.0179
IMP	0.3	=	0.5303	0.4930	0.0267	0.0283	0.3172	0.0298
		HD	1.0227	0.8603	0.0231	0.0245	0.3933	0.0298
		HP	0.1381	0.2188	0.0222	0.0254	0.1953	0.0298
Worse	0.5	=	0.6150	0.5234	0.0123	0.0128	0.2562	0.0179
		HD	0.2953	0.4236	0.0155	0.0151	0.3265	0.0179
		HP	0.7962	0.5828	0.0161	0.0184	0.1801	0.0179
Worse	0.3	=	0.6042	0.5236	0.0166	0.0223	0.2803	0.0298
		HD	0.0738	0.3505	0.0245	0.0263	0.3656	0.0298
		HP	0.9876	0.6560	0.0267	0.0315	0.1256	0.0298

Table 5. Variance in estimates of the efficacy estimand

			Variance in estimates of change from baseline					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	0.3583	0.4655	0.6995	0.7493	0.4760	0.5214
		HD	0.2924	0.3339	0.6913	0.7184	0.3909	0.5214
		HP	0.2701	0.3271	0.6821	0.7360	0.5419	0.5214
IMP	0.3	=	0.3459	0.4537	0.7048	0.7626	0.4442	0.5214
		HD	0.2825	0.3286	0.6991	0.7479	0.3716	0.5214
		HP	0.2660	0.3239	0.6799	0.7133	0.5257	0.5214
IMP	0	=	0.3127	0.4235	0.6896	0.7156	0.4222	0.5214
		HD	0.2679	0.3202	0.7005	0.7339	0.3579	0.5214
		HP	0.2632	0.3225	0.6852	0.7222	0.5237	0.5214
Worse	0.5	=	0.2073	0.2749	0.6841	0.7312	0.4702	0.5214
		HD	0.2049	0.2741	0.6779	0.7154	0.3984	0.5214
		HP	0.1965	0.2705	0.6791	0.7044	0.5359	0.5214
Worse	0.3	=	0.2193	0.2760	0.6807	0.6900	0.4504	0.5214
		HD	0.2019	0.2719	0.6901	0.7219	0.3800	0.5214
		HP	0.1938	0.2682	0.6802	0.7020	0.5233	0.5214
Worse	0	=	0.2285	0.2743	0.6742	0.6911	0.4281	0.5214
		HD	0.2012	0.2688	0.6846	0.7259	0.3650	0.5214
		HP	0.1956	0.2670	0.6803	0.7008	0.5207	0.5214

Table 6. Mean Square Error in estimates of the efficacy estimand

			Mean square error in estimates of change from baseline					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	2.4063	2.2317	0.7013	0.7513	1.0598	0.5238
		HD	5.6455	4.3137	0.6929	0.7199	1.3401	0.5238
		HP	0.8583	1.0953	0.6837	0.7377	0.7693	0.5238
IMP	0.3	=	1.1139	1.1176	0.7068	0.7648	0.7189	0.5238
		HD	3.1390	2.3498	0.7006	0.7495	0.7940	0.5238
		HP	0.3181	0.4547	0.6812	0.7150	0.6299	0.5238
IMP	0	=	0.3129	0.4239	0.6911	0.7175	0.4332	0.5238
		HD	0.7311	0.5825	0.7022	0.7354	0.3755	0.5238
		HP	0.7058	0.5431	0.6866	0.7236	0.5282	0.5238
Worse	0.5	=	3.0769	2.3538	0.6853	0.7324	0.9683	0.5238
		HD	0.8665	1.6358	0.6797	0.7171	1.2072	0.5238
		HP	5.0068	2.8474	0.6811	0.7070	0.7820	0.5238
Worse	0.3	=	1.2163	1.0247	0.6814	0.6914	0.6649	0.5238
		HD	0.2167	0.6073	0.6917	0.7238	0.7451	0.5238
		HP	2.8575	1.4434	0.6821	0.7047	0.5664	0.5238
Worse	0	=	0.2286	0.2748	0.6752	0.6923	0.4413	0.5238
		HD	0.9003	0.3413	0.6864	0.7273	0.3802	0.5238
		HP	0.9001	0.3583	0.6824	0.7028	0.5215	0.5238

Table 7. Confidence interval coverage in estimates of the efficacy estimand

			Percentage of confidence intervals containing the true value					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	12.2	22.7	94.2	94.9	94.5	94.0
		HD	0.2	1.3	95.2	95.3	93.4	94.0
		HP	49.3	42.7	95.4	95.2	94.0	94.0
IMP	0.3	=	40.1	47.0	94.8	94.4	97.4	94.0
		HD	3.9	10.7	95.4	94.7	98.5	94.0
		HP	79.2	71.8	96.0	94.6	96.1	94.0
IMP	0	=	79.7	76.6	94.4	94.9	99.4	94.0
		HD	52.5	63.3	95.5	94.9	99.5	94.0
		HP	50.5	63.6	95.5	94.9	97.5	94.0
Worse	0.5	=	0.6	6.1	94.6	94.3	93.7	94.0
		HD	36.3	17.1	95.2	95.2	93.3	94.0
		HP	0.1	2.8	95.3	95.5	93.9	94.0
Worse	0.3	=	23.2	36.8	95.0	94.4	96.6	94.0
		HD	84.7	57.6	95.2	95.2	97.7	94.0
		HP	1.9	21.5	95.3	95.6	96.6	94.0
Worse	0	=	84.8	80.9	94.9	94.3	99.0	94.0
		HD	35.8	75.4	95.3	95.0	99.7	94.0
		HP	36.5	75.1	95.2	94.9	97.5	94.0

Table 8. Rejection rates in estimates of the efficacy estimand

			Percentage of datasets where the null hypothesis of no treatment difference was rejected					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	83.5	82.4	88.6	85.5	55.7	96.0
		HD	28.7	52.6	90.1	86.4	41.1	96.0
		HP	99.2	97.8	90.6	86.3	79.2	96.0
IMP	0.3	=	53.5	53.5	46.4	41.5	14.3	59.5
		HD	6.6	18.3	49.4	45.8	7.9	59.5
		HP	91.7	82.5	47.8	44.4	31.3	59.5
IMP	0	=	9.5	11.0	2.3	2.1	0.1	2.6
		HD	0.3	1.5	1.9	2.2	0.2	2.6
		HP	49.0	34.4	1.8	2.2	0.8	2.6
Worse	0.5	=	82.0	88.9	90.3	87.8	64.8	96.0
		HD	99.6	95.8	90.5	87.4	48.0	96.0
		HP	42.3	83.2	90.3	87.8	79.9	96.0
Worse	0.3	=	50.3	58.8	50.5	45.1	20.3	59.5
		HD	97.4	76.7	49.7	46.1	11.2	59.5
		HP	6.9	43.2	48.0	45.5	36.0	59.5
Worse	0	=	7.5	7.9	2.3	2.7	0.6	2.6
		HD	64.1	21.0	2.0	2.4	0.2	2.6
		HP	0.1	2.4	1.9	2.1	1.0	2.6

Table 9. Bias in estimates of the effectiveness estimand

			Analyze of change from baseline using				
			BOCF	LOCF	DL	MI	PMI
IMP	0.5	=	0.7267	0.6246	-0.6614	-0.6598	0.0598
		HD	1.2434	0.9247	-1.0307	-1.0311	-0.0960
		HP	0.2529	0.3625	-0.4744	-0.4732	-0.0372
IMP	0.3	=	0.3618	0.3002	-0.4704	-0.4678	0.0097
		HD	1.0253	0.7569	-0.6267	-0.6243	-0.0149
		HP	-0.0926	0.0408	-0.2841	-0.2789	0.0019
IMP	0	=	0.0123	0.0187	0.0386	0.0437	0.1048
		HD	0.6806	0.5121	0.0408	0.0387	0.1324
		HP	-0.6653	-0.4697	0.0372	0.0365	0.0667
Worse	0.5	=	0.8519	0.5998	-0.8082	-0.8068	-0.1363
		HD	-0.2646	0.0889	-1.0352	-1.0364	-0.1787
		HP	1.6215	1.0336	-0.5272	-0.5211	-0.0757
Worse	0.3	=	0.5001	0.3668	-0.4710	-0.4617	-0.0352
		HD	-0.5513	-0.0940	-0.6326	-0.6298	-0.0690
		HP	1.2873	0.7394	-0.3006	-0.2928	-0.1372
Worse	0	=	0.0079	0.0220	0.0325	0.0339	0.1152
		HD	-0.8361	-0.2693	0.0422	0.0372	0.1231
		HP	0.8393	0.3022	0.0459	0.0447	-0.0286

Table 10. Relative bias in estimates of the effectiveness estimand

			Analyze of change from baseline using				
			BOCF	LOCF	DL	MI	pMI
IMP	0.5	=	0.3545	0.3047	0.3226	0.3218	0.0291
		HD	0.7383	0.5490	0.6120	0.6122	0.0570
		HP	0.1129	0.1618	0.2117	0.2112	0.0166
IMP	0.3	=	0.3179	0.2638	0.4134	0.4110	0.0085
		HD	1.0380	0.7662	0.6345	0.6321	0.0151
		HP	0.0696	0.0306	0.2133	0.2094	0.0014
Worse	0.5	=	0.4455	0.3136	0.4226	0.4218	0.0713
		HD	0.1578	0.0530	0.6175	0.6181	0.1066
		HP	0.7429	0.4735	0.2415	0.2387	0.0347
Worse	0.3	=	0.4333	0.3178	0.4081	0.4000	0.0305
		HD	0.5629	0.0960	0.6459	0.6430	0.0705
		HP	0.9843	0.5653	0.2298	0.2238	0.1049

Table 11. Mean square error in estimates of the effectiveness estimand

			Analyze of change from baseline using				
			BOCF	LOCF	DL	MI	pMI
IMP	0.5	=	0.8865	0.8557	1.1370	1.1846	0.4796
		HD	1.8384	1.1889	1.7538	1.7814	0.4001
		HP	0.3341	0.4585	0.9072	0.9600	0.5433
IMP	0.3	=	0.4768	0.5438	0.9261	0.9815	0.4442
		HD	1.3337	0.9014	1.0919	1.1377	0.3718
		HP	0.2746	0.3255	0.7606	0.7911	0.5257
IMP	0	=	0.3129	0.4239	0.6911	0.7175	0.4332
		HD	0.7311	0.5825	0.7022	0.7354	0.3755
		HP	0.7058	0.5431	0.6866	0.7236	0.5282
Worse	0.5	=	0.9330	0.6346	1.3373	1.3820	0.4888
		HD	0.2750	0.2820	1.7497	1.7894	0.4303
		HP	2.8259	1.3388	0.9571	0.9759	0.5416
Worse	0.3	=	0.4694	0.4106	0.9025	0.9031	0.4516
		HD	0.5058	0.2807	1.0903	1.1185	0.3848
		HP	1.8511	0.8148	0.7705	0.7877	0.5421
Worse	0	=	0.2286	0.2748	0.6752	0.6923	0.4413
		HD	0.9003	0.3413	0.6864	0.7273	0.3802
		HP	0.9001	0.3583	0.6824	0.7028	0.5215

Table 12. Confidence interval coverage for the effectiveness estimand

			Analyze of change from baseline using				
			BOCF	LOCF	DL	MI	pMI
IMP	0.5	=	51.0	59.0	88.3	89.3	99.1
		HD	17.4	38.3	76.1	78.7	99.3
		HP	79.1	72.5	91.7	91.8	98.1
IMP	0.3	=	72.2	73.2	92.0	91.8	99.4
		HD	28.2	50.5	87.9	89.2	99.5
		HP	84.1	80.4	93.1	94.3	97.7
IMP	0	=	79.7	76.6	94.4	94.9	99.4
		HD	52.5	63.3	95.5	94.9	99.5
		HP	50.5	63.6	95.5	94.9	97.5
Worse	0.5	=	32.6	54.9	85.0	85.4	98.8
		HD	77.9	80.0	76.0	76.9	99.0
		HP	1.9	24.7	90.6	91.3	98.2
Worse	0.3	=	62.1	69.9	91.3	92.5	98.6
		HD	59.3	81.1	86.6	88.7	99.5
		HP	7.3	45.9	93.4	93.6	97.8
Worse	0	=	84.8	80.9	94.9	94.3	99.0
		HD	35.8	75.4	95.3	95.0	99.7
		HP	36.5	75.1	95.2	94.9	97.5

Table 13. Endpoint treatment contrasts by analytic method from the actual clinical trial dataset.

Method	Endpoint Contrast	Standard error	P value
Direct likelihood	2.13	0.97	0.030
MI	2.09	0.92	0.023
pMI	1.54	0.94	0.102
LOCF	1.75	0.88	0.047
BOCF	1.08	0.89	0.253

Appendix A.

Main effect (Over all)

Table A1. Bias in estimates of the efficacy estimand

			Bias in mean estimates of change from baseline					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	0.2528	0.1507	0.0213	0.0212	0.3725	0.0218
		HD	1.1354	0.8167	0.0163	0.0176	0.4468	0.0218
		HP	-0.4113	-0.3018	0.0170	0.0177	0.2221	0.0218
IMP	0.3	=	0.1694	0.1078	0.0214	0.0235	0.2584	0.0217
		HD	0.9831	0.7147	0.0154	0.0179	0.3021	0.0217
		HP	-0.4788	-0.3454	0.0161	0.0200	0.1553	0.0217
IMP	0	=	0.0123	0.0187	0.0185	0.0228	0.0614	0.0218
		HD	0.6806	0.5121	0.0167	0.0157	0.0733	0.0218
		HP	-0.6653	-0.4697	0.0161	0.0148	0.0457	0.0218
Worse	0.5	=	0.5157	0.2636	0.0178	0.0181	0.3044	0.0218
		HD	-0.3649	-0.0113	0.0171	0.0188	0.3854	0.0218
		HP	1.0150	0.4270	0.0199	0.0196	0.2185	0.0218
Worse	0.3	=	0.2915	0.1583	0.0147	0.0165	0.2061	0.0217
		HD	-0.5851	-0.1278	0.0168	0.0172	0.2635	0.0217
		HP	0.9251	0.3771	0.0199	0.0238	0.0895	0.0217
Worse	0	=	0.0079	0.0220	0.0158	0.0148	0.0662	0.0218
		HD	-0.8361	-0.2693	0.0174	0.0167	0.0660	0.0218
		HP	0.8393	0.3022	0.0204	0.0205	-0.0061	0.0218

Table A2. Relative Bias in estimates of the efficacy estimand

			Relative bias in mean estimates of change from baseline					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	0.1604	0.0956	0.0135	0.0135	0.2363	0.0138
		HD	0.7203	0.5181	0.0103	0.0112	0.2834	0.0138
		HP	0.2610	0.1914	0.0108	0.0112	0.1409	0.0138
IMP	0.3	=	0.1791	0.1140	0.0226	0.0248	0.2732	0.0230
		HD	1.0397	0.7558	0.0162	0.0189	0.3195	0.0230
		HP	0.5063	0.3652	0.0170	0.0212	0.1643	0.0230
Worse	0.5	=	0.3272	0.1672	0.0113	0.0115	0.1931	0.0138
		HD	0.2315	0.0072	0.0108	0.0120	0.2445	0.0138
		HP	0.6439	0.2709	0.0126	0.0124	0.1386	0.0138
Worse	0.3	=	0.3083	0.1674	0.0155	0.0174	0.2179	0.0230
		HD	0.6188	0.1352	0.0177	0.0182	0.2787	0.0230
		HP	0.9783	0.3988	0.0210	0.0251	0.0946	0.0230

Table A3. Variance in estimates of the efficacy estimand

			Variance in estimates of change from baseline					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	0.3583	0.4655	0.3285	0.3425	0.2745	0.2759
		HD	0.2924	0.3339	0.3117	0.3177	0.2423	0.2759
		HP	0.2701	0.3271	0.3090	0.3194	0.2798	0.2759
IMP	0.3	=	0.3459	0.4537	0.3299	0.3442	0.2629	0.2759
		HD	0.2825	0.3286	0.3137	0.3289	0.2366	0.2759
		HP	0.2660	0.3239	0.3085	0.3146	0.2756	0.2759
IMP	0	=	0.3127	0.4235	0.3207	0.3283	0.2501	0.2759
		HD	0.2679	0.3202	0.3134	0.3206	0.2317	0.2759
		HP	0.2632	0.3225	0.3093	0.3177	0.2752	0.2759
Worse	0.5	=	0.2073	0.2749	0.3123	0.3206	0.2727	0.2759
		HD	0.2049	0.2741	0.3062	0.3136	0.2518	0.2759
		HP	0.1965	0.2705	0.3029	0.3060	0.2786	0.2759
Worse	0.3	=	0.2193	0.2760	0.3106	0.3085	0.2664	0.2759
		HD	0.2019	0.2719	0.3077	0.3143	0.2466	0.2759
		HP	0.1938	0.2682	0.3029	0.3088	0.2773	0.2759
Worse	0	=	0.2285	0.2743	0.3074	0.3097	0.2590	0.2759
		HD	0.2012	0.2688	0.3065	0.3190	0.2417	0.2759
		HP	0.1956	0.2670	0.3032	0.3069	0.2765	0.2759

Table A4. Mean Square Error (MSE) in estimates of the efficacy estimand

			MSE in estimates of change from baseline					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	0.4222	0.4882	0.3290	0.3429	0.4133	0.2764
		HD	1.5816	1.0009	0.3120	0.3180	0.4419	0.2764
		HP	0.4393	0.4182	0.3093	0.3197	0.3291	0.2764
IMP	0.3	=	0.3746	0.4653	0.3303	0.3447	0.3297	0.2764
		HD	1.2490	0.8394	0.3140	0.3292	0.3278	0.2764
		HP	0.4952	0.4432	0.3087	0.3150	0.2997	0.2764
IMP	0	=	0.3129	0.4239	0.3211	0.3288	0.2539	0.2764
		HD	0.7311	0.5825	0.3137	0.3209	0.2371	0.2764
		HP	0.7058	0.5431	0.3096	0.3179	0.2773	0.2764
Worse	0.5	=	0.4733	0.3444	0.3127	0.3209	0.3654	0.2764
		HD	0.3381	0.2743	0.3065	0.3140	0.4003	0.2764
		HP	1.2267	0.4529	0.3033	0.3064	0.3263	0.2764
Worse	0.3	=	0.3042	0.3011	0.3108	0.3088	0.3088	0.2764
		HD	0.5443	0.2882	0.3080	0.3146	0.3160	0.2764
		HP	1.0496	0.4104	0.3033	0.3094	0.2853	0.2764
Worse	0	=	0.2286	0.2748	0.3077	0.3099	0.2634	0.2764
		HD	0.9003	0.3413	0.3068	0.3193	0.2460	0.2764
		HP	0.9001	0.3583	0.3036	0.3073	0.2765	0.2764

Table A5. Confidence Interval (CI) Coverage from the efficacy estimand

			Percentage of CI containing the true value					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	76.8	76.0	95.4	94.5	94.1	95.9
		HD	21.9	46.5	95.2	95.1	93.8	95.9
		HP	71.5	75.0	95.5	95.2	95.3	95.9
IMP	0.3	=	77.4	75.0	95.2	94.5	96.8	95.9
		HD	31.6	53.0	95.0	95.0	96.6	95.9
		HP	66.6	71.8	95.4	96.0	96.0	95.9
IMP	0	=	79.7	76.6	95.1	95.2	98	95.9
		HD	52.5	63.3	95.1	95	98.5	95.9
		HP	50.5	63.6	95.4	95.9	96.5	95.9
Worse	0.5	=	60.0	74.2	95.4	95.4	94.1	95.9
		HD	72.3	81.2	95.2	95.1	93.8	95.9
		HP	19.9	65.4	95.1	95.9	95.2	95.9
Worse	0.3	=	76.0	77.4	95.0	95.1	95.9	95.9
		HD	56.8	79.6	95.2	95.4	96.1	95.9
		HP	26.9	69.0	95.0	94.8	96.3	95.9
Worse	0	=	84.8	80.9	95.1	95	97	95.9
		HD	35.8	75.4	95.1	94.9	97.7	95.9
		HP	36.5	75.1	95.2	95.3	96.7	95.9

Table A6. Rejection Rates in estimates of the efficacy estimand

			percentage of datasets where the null hypothesis of no treatment difference was rejected					
			BOCF	LOCF	DL	MI	pMI	FULL
IMP	0.5	=	83.5	82.4	77.0	74.9	48.3	81.9
		HD	28.7	52.6	78.0	77.6	43.7	81.9
		HP	99.2	97.8	78.3	77.0	66.1	81.9
IMP	0.3	=	53.5	53.5	34.4	34.5	17.1	41.7
		HD	6.6	18.3	37.6	35.9	13.1	41.7
		HP	91.7	82.5	38.5	37.3	28.7	41.7
IMP	0	=	9.5	11.0	1.9	1.8	0.8	1.5
		HD	0.3	1.5	2.0	2.3	0.4	1.5
		HP	49.0	34.4	1.9	1.8	1.3	1.5
Worse	0.5	=	82.0	88.9	78.8	77.9	59.5	81.9
		HD	99.6	95.8	78.5	78.4	51.9	81.9
		HP	42.3	83.2	78.3	77.4	68.0	81.9
Worse	0.3	=	50.3	58.8	38.1	37.2	23.9	41.7
		HD	97.4	76.7	39.0	37.3	16.8	41.7
		HP	6.9	43.2	38.1	37.1	31.2	41.7
Worse	0	=	7.5	7.9	2.1	2.1	1.1	1.5
		HD	64.1	21.0	2.0	2.2	0.7	1.5
		HP	0.1	2.4	1.8	2.0	1.4	1.5

Table A7. Bias in estimates of the effectiveness estimand

			Analyze of change from baseline using				
			BOCF	LOCF	DL	MI	PMI-II
IMP	0.5	=	-0.0878	-0.1899	-0.3193	-0.3194	0.0319
		HD	0.6558	0.3370	-0.4634	-0.4620	-0.0328
		HP	-0.6373	-0.5277	-0.2090	-0.2082	-0.0038
IMP	0.3	=	-0.0782	-0.1398	-0.2262	-0.2241	0.0108
		HD	0.6869	0.4185	-0.2809	-0.2783	0.0059
		HP	-0.6190	-0.4856	-0.1241	-0.1202	0.0151
IMP	0	=	0.0123	0.0187	0.0185	0.0228	0.0614
		HD	0.6806	0.5121	0.0167	0.0157	0.0733
		HP	-0.6653	-0.4697	0.0161	0.0148	0.0457
Worse	0.5	=	0.1809	-0.0712	-0.3170	-0.3167	-0.0304
		HD	-0.8051	-0.4516	-0.4231	-0.4214	-0.0549
		HP	0.7862	0.1982	-0.2089	-0.2092	-0.0104
Worse	0.3	=	0.0950	-0.0383	-0.1819	-0.1801	0.0095
		HD	-0.8582	-0.4009	-0.2563	-0.2559	-0.0096
		HP	0.7882	0.2402	-0.1170	-0.1131	-0.0474
Worse	0	=	0.0079	0.0220	0.0158	0.0148	0.0662
		HD	-0.8361	-0.2693	0.0174	0.0167	0.0660
		HP	0.8393	0.3022	0.0204	0.0205	-0.0061

Table A8. Relative bias in estimates of the effectiveness estimand

			Analyze of change from baseline using				
			BOCF	LOCF	DL	MI	pMI
IMP	0.5	=	0.0710	0.1537	0.2584	0.2585	0.0258
		HD	0.5980	0.3073	0.4225	0.4213	0.0299
		HP	0.4719	0.3908	0.1547	0.1542	0.0028
IMP	0.3	=	0.1120	0.2002	0.3240	0.3210	0.0155
		HD	1.0578	0.6444	0.4325	0.4286	0.0091
		HP	0.7686	0.6030	0.1542	0.1493	0.0187
Worse	0.5	=	0.1457	0.0574	0.2553	0.2551	0.0245
		HD	0.7087	0.3975	0.3724	0.3709	0.0483
		HP	0.5835	0.1471	0.1550	0.1552	0.0077
Worse	0.3	=	0.1268	0.0511	0.2428	0.2404	0.0127
		HD	1.2762	0.5962	0.3811	0.3805	0.0143
		HP	0.9746	0.2970	0.1447	0.1399	0.0586

Table A9. MSE in estimates of the effectiveness estimand

			Analyze of change from baseline using				
			BOCF	LOCF	DL	MI	pMI
IMP	0.5	=	0.3660	0.5016	0.4305	0.4445	0.2755
		HD	0.7224	0.4475	0.5264	0.5312	0.2434
		HP	0.6762	0.6056	0.3527	0.3628	0.2798
IMP	0.3	=	0.3520	0.4732	0.3810	0.3944	0.2630
		HD	0.7543	0.5037	0.3926	0.4063	0.2366
		HP	0.6492	0.5597	0.3239	0.3291	0.2758
IMP	0	=	0.3129	0.4239	0.3211	0.3288	0.2539
		HD	0.7311	0.5825	0.3137	0.3209	0.2371
		HP	0.7058	0.5431	0.3096	0.3179	0.2773
Worse	0.5	=	0.2400	0.2800	0.4128	0.4209	0.2737
		HD	0.8532	0.4780	0.4852	0.4912	0.2548
		HP	0.8146	0.3098	0.3465	0.3498	0.2787
Worse	0.3	=	0.2283	0.2775	0.3437	0.3410	0.2665
		HD	0.9384	0.4326	0.3734	0.3798	0.2467
		HP	0.8151	0.3259	0.3166	0.3216	0.2796
Worse	0	=	0.2286	0.2748	0.3077	0.3099	0.2634
		HD	0.9003	0.3413	0.3068	0.3193	0.2460
		HP	0.9001	0.3583	0.3036	0.3073	0.2765

Table A10. CI Coverage from the effectiveness estimand

			Analyze of change from baseline using				
			BOCF	LOCF	DL	MI	PMI-II
IMP	0.5	=	77.8	73.0	91.7	92.3	97.4
		HD	57.2	73.1	87.9	87.7	98.5
		HP	55.5	61.4	94.6	94.6	96.5
IMP	0.3	=	78.3	74.7	93.5	93.0	97.9
		HD	54.1	68.1	92.7	92.5	98.8
		HP	56.1	63.3	94.9	95.8	96.4
IMP	0	=	79.7	76.6	95.1	95.2	98
		HD	52.5	63.3	95.1	95	98.5
		HP	50.5	63.6	95.4	95.9	96.5
Worse	0.5	=	81.7	77.9	91.8	91.3	96.7
		HD	36.3	62.9	88.9	89.1	97.8
		HP	38.7	78.3	93.9	93.9	96.2
Worse	0.3	=	84.3	79.5	94.3	94.4	96.8
		HD	33.2	66.4	93.0	93.0	97.7
		HP	39.4	77.4	94.8	95.1	96.8
Worse	0	=	84.8	80.9	95.1	95	97
		HD	35.8	75.4	95.1	94.9	97.7
		HP	36.5	75.1	95.2	95.3	96.7