Describing Patient Reported Outcomes from Oncology Trials with Informative Dropouts due to Adverse Events

Chia-Wen Ko[1], Paul Kluetz[2], Rajeshwari Sridhara[1]
[1]FDA/CDER/OTS/OB/DBV, 10903 New Hampshire Ave., Silver Spring, MD 20903
[2]FDA/CDER/OND/OHOP, 10903 New Hampshire Ave., Silver Spring, MD 20903

Abstract
Patient-reported outcome (PRO) data are rarely used as primary or secondary endpoints in oncology clinical trials. Reasons include the availability of alternative objective endpoints such as radiographic tumor measurement and overall survival as well as trial design characteristics unique to oncology that subject PRO to potential bias, including open label studies and high degrees of missing data. Nonetheless, there is increasing effort to include high quality clinical outcome data such as PRO in the benefit-risk determination for cancer drugs, and it is critical that we find ways to describe and mitigate the challenges associated with PRO data obtained in cancer clinical trials. One important issue when analyzing PRO data is how to take into account informative missing data as a result of treatment discontinuations due to drug-related toxicities. The goal of this work is to compare statistical methods for describing PRO data obtained from oncology patients in the presence of imbalanced treatment discontinuation between study arms. In this work, we apply imputation strategies and modeling to PRO data from randomized oncology trials submitted to the FDA with different degrees of imbalance between study arms in the incidence of adverse events leading to treatment discontinuation.

Key Words: patient-reported outcome, treatment discontinuation, imputation

1. Introduction

It is increasingly evident that an understanding of how a product will affect patient-reported symptoms and/or "quality of life" is highly relevant for patients and physicians when it comes to choosing a cancer treatment. Although it is unusual to have patient reported outcomes (PROs) as major endpoints in registration trials of cancer treatments, global health related quality of life instruments are typically integrated into these trials as exploratory data to help inform for regulatory and reimbursement decisions.

One important issue with describing PRO data in a cancer trial is the occurrence of missing data arising as a consequence of withdrawal due to adverse side-effects of treatment. Since it is assumed that patients that have an adverse event leading to treatment discontinuation are likely to have a worse quality of life during the treatment, their dropouts are likely to be informative with respect to the analysis of the PRO concept being measured.

The primary analysis proposed in a protocol for a PRO endpoint is usually a mixed effect model repeated measurement (MMRM) analysis, without accounting for dropouts due to adverse events. The basis for not accounting for the reason of dropout, according to

sponsors, is that the MMRM analysis compensates for missing values and has been found to be appropriate under the missing-at-random (MAR) assumption[1]. However, it is not clear whether or not the MAR assumption is always justified for dropouts.

As an initial investigation, we imputed missing longitudinal patient reported health-related quality of life data using the last observation carried forward (LOCF) single imputation approach that operates under the missing-complete-at-random (MCAR) assumption[2], and using the fully conditional specification (FCS) multiple imputation approach that operates under the MAR assumption if a covariate that is predictive of missing outcome data is included in the imputation model. Discrepancy observed between the MMRM analysis based on the original observed data and the MMRM analysis based on the LOCF imputed data would suggest that the patient dropouts did not occur completely at random. Discrepancy observe d between the MMRM analysis based on the observed data and the MMRM analysis based on FCS imputed data would suggest that the dropouts did not occur at random after accounting for only the considered covariates, and additional factors will have to be accounted for in order to provide an un-biased description to the patient reported quality of life data over time.

2. Materials and Methods

2.1 Selected Oncology Trials

We selected 3 oncology registration trials for this investigation. Table 1 below gives a summary to those trials. All the 3 trials had two treatment groups with equal randomization, had at least 10% of patients dropped out from treatment due to adverse events, and had administrated a well-accepted quality of life instrument Quality of Life Questionnaire-Core 36 (QLQ-C30[3]) to all the patients. But more importantly, the 3 trials were selected because of their differences. The trials were different in the studied type of cancer, a double-blind versus an open-label design, and the overall percentage and degree of difference between treatment arms in the percentage of adverse event related dropouts. Another important difference between the trials was that the primary treatment period for Trial 3 was broken into 2 24-week treatment phases. At the end of phase 1, Trial 3 investigators got to assess the patient's response to treatment, and only the patients who were doing well on their treatment could continue with another 24 weeks of treatment.

Table 1: Summary on Selected Oncology Registration Trials

Trial	Type of cancer	Design Treatment duration (primary period*)	AE-related dropouts	
			Experimental	Control
1	Gastric	Double-blind Continuous (24 weeks)	39/330 (12%)	38/335 (11%)
2	Ovarian	Open-label Continuous (30 weeks)	62/179 (17%)	19/182 (5%)
3	Multiple myeloma	Double-blind 48 weeks (24 + 24 weeks)+	130/381 (34%)	66/377 (17%)

* QLQ-C30 was either assessed more frequently (Trial 1) or only assessed during this period (Trials 2 and 3)

+ Patients with observed clinical benefit in the first 24 weeks according to investigators could continue another 24 weeks of treatment

2.2 PRO Data for Analyses

We used the QLQ-C30 global health status scores for this investigation. QLQ-C30 was the only quality of life questionnaire administrated in all the selected trials. The questionnaire measured five functional dimensions (physical, role, emotional, cognitive, and social), six single-item symptom scales (dyspnea, sleep disturbance, appetite loss, constipation, diarrhea and financial impact) and a global health status/quality-of-life scale. The recall period for QLQ-C30 was the past week. For each domain and scale, a linear transformation was applied to standardize the score between 0 and 100. A higher score for a quality of life scale corresponded with a higher health-related quality of life as reported by the patient.

For all the 3 trials, QLQ-C30 was assessed every 2 treatment cycles (6 weeks for Trials 1 and 3, and 8 weeks for Trial 2) during the primary treatment period.

2.3 MMRM Model for the PRO Data

We used a longitudinal mixed-effect model for repeated measures to estimate the mean change in QLQ-C30 global health status score since baseline. This model included visit time point (categorical) for repeated measures, treatment and treatment by time interaction as fixed effects, as well as baseline score and randomization factors as covariates. The covariance structure between repeated measures was assumed to be unknown in the analyses**.**

2.2 Imputation Methods

As mentioned in the Introduction section, we used the LOCF single imputation and the FCS multiple imputation statistical methods to impute the missing QLQ-C30 values. MMRM model estimated mean changes since baseline in QLQ-C30 global health scores based on observed data were compared to the ones based on imputed data, as a way to evaluate the sensitivity of the MMRM analysis to the missing at random assumption.

The following provides some description to the LOCF and the FCS imputation methods:

- LOCF single imputation[4]: The LOCF method replaced every missing score after treatment withdrawal with the last score on treatment. This imputation method did not consider uncertainty from conducting data imputation, and could not adjust for factors related to the dropouts.
- FCS multiple imputations[5]: The FCS method is also known as the sequential regressions. Each imputation involves a two-step process: the first step is to set initial values for missing QLQ-C30 scores at scheduled visits by random drawing from a distribution with the same mean and standard deviation as the non-missing values; the second step is to impute original missing values by sequential linear regressions, one scheduled visit at a time according to the ordering of visits, based on covariates and the most recently available values at all other visits. This imputation process was repeated 10 times to get a random sample of imputations.

FCS provides the flexibility to account for factors that may be related to the occurrence of missing data and the missing data patterns, while a LOCF approach relies only on the last available value for imputation. For this investigation, two factors were included as

covariates in the FCS sequential regression imputation models: (1) reason for dropout (AE-related versus not AE-related); (2) extent of treatment exposure as measured by the number of treatment cycles completed.

3. Results

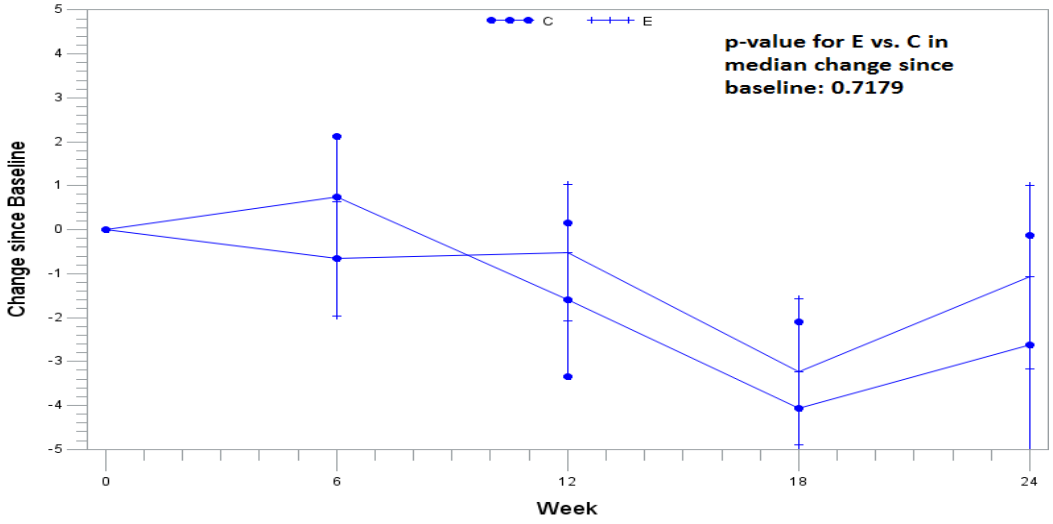
3.1 Results from Trial 1

Trial 1 was a double-blind trial in gastric cancer patients, with 12% of patients in the experimental arm versus 11% of patients in the control arm that had discontinued from treatment due to adverse events. Both arms had 97% of patients completed the questionnaire at baseline, but the completion rate quickly decreased, especially in the control arm. However, the missing PRO data was not mainly due to adverse event related dropouts. The estimated mean change since baseline was comparable between the two treatment groups, with a p-value=0.7179 from the mixed model comparing between the experimental group and the control group in average QLQ-C30 global health score change over time.

Table 2: Trial 1 QLQ-C30 Global Health Score Missing Data Pattern

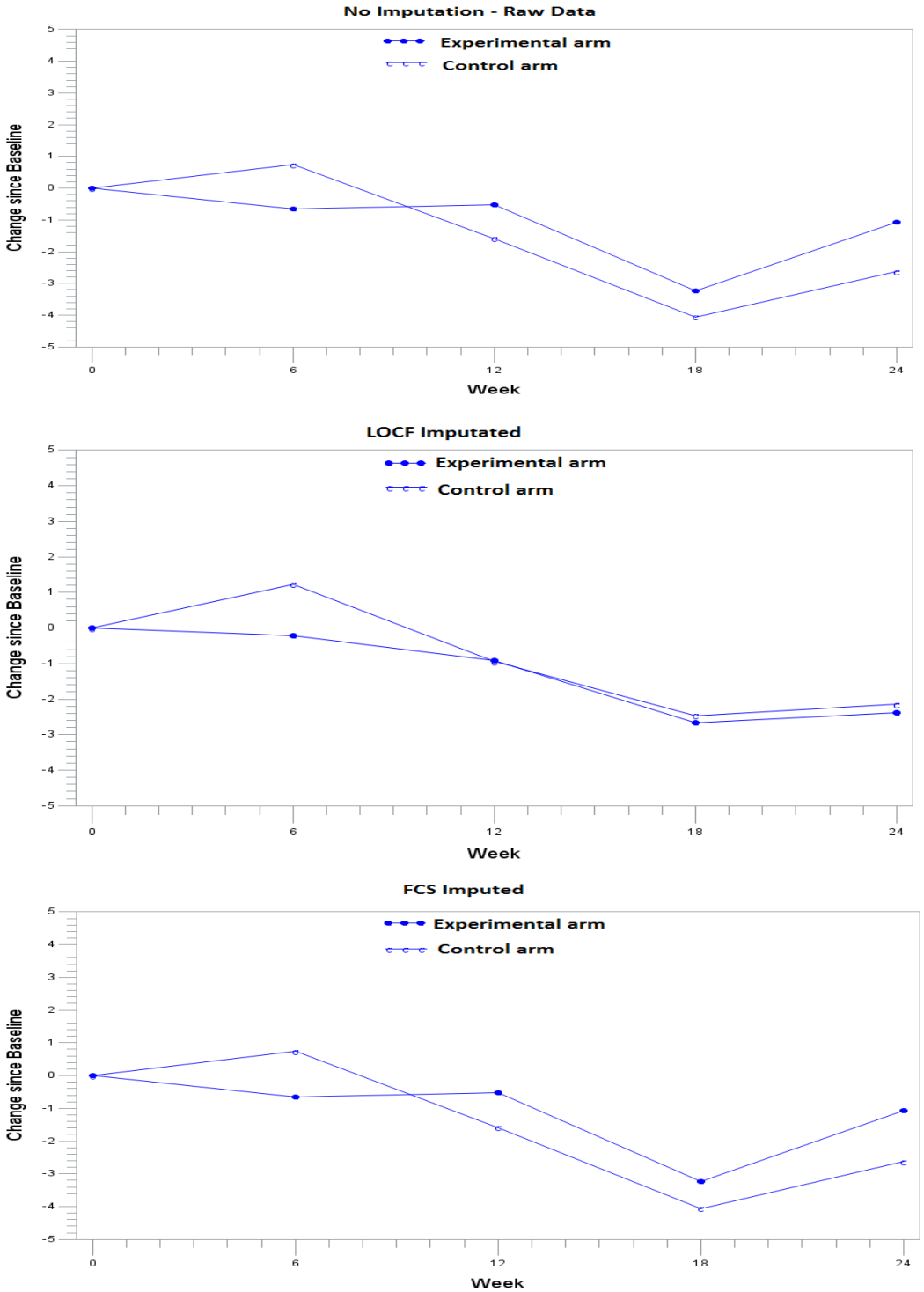
E (N=330)	**Week 0**	**Week 6**	**Week 12**	**Week 18**	**Week 24**
Available	321 (97%)	248 (75%)	178 (54%)	119 (36%)	58 (18%)
Missing		73	70	59	61
AE-related dropout		13	1	9	8
C (N=335)	**Week 0**	**Week 6**	**Week 12**	**Week 18**	**Week 24**
Available	326 (97%)	223 (67%)	130 (39%)	78 (23%)	40 (12%)
Missing		103	93	52	38
AE-related dropout		17	5	5	1

Figure 1: Trial 1 Primary MMRM Analysis for QLQ-C30 Global Health Sore



The estimates for mean change in the global health score were almost identical for the original analysis and the analysis with missing data imputed by the multiple imputation approach. The LOCF analysis, on the other hand, almost diminished the difference in score change between the two treatment groups.

Figure 2: Trial 1 Raw Data versus Missing Data Imputed MMRM Estimates



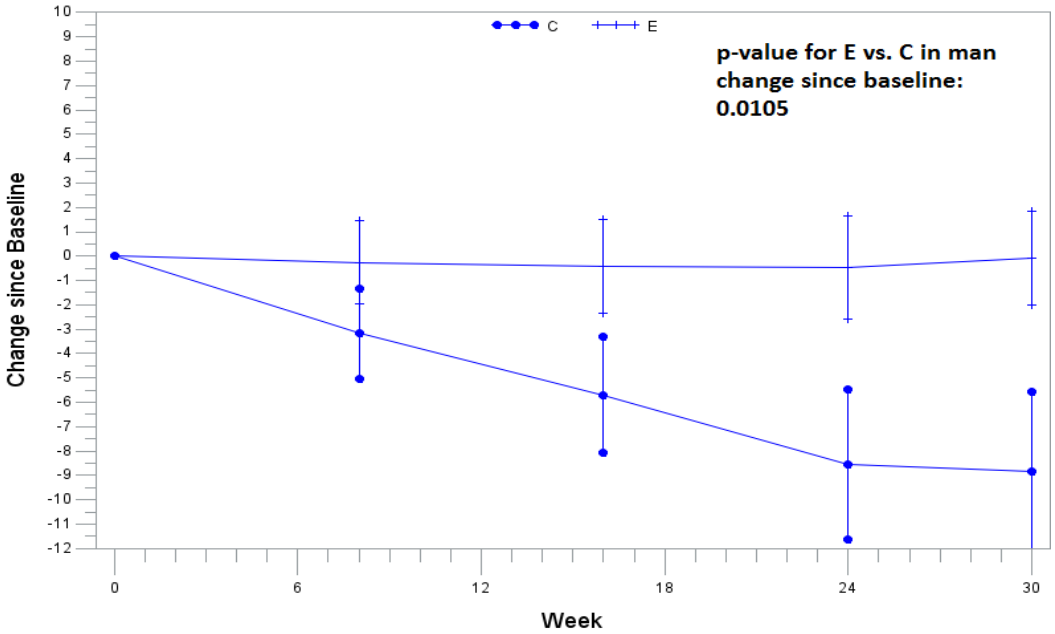
3.2 Results from Trial 2

Trial 2 was an open-label trial in ovarian cancer patients, with 17% AE-related dropout rate in the experimental group versus 5% AE-related dropout rate in the control group. Like Trial 1, Trial 2 had a good completion rate for the questionnaire at baseline, but the completion rate quickly decreased, especially in the control arm. However, unlike Trial 1, the percentage of missing PRO data as a result of AE-related dropout was quite different between the groups; the AE-related dropout contributed to only a small portion of missing data for the control group, but contributed to about one-third of the missing data for the experimental group. The estimated mean change since baseline was not comparable between the two treatment groups, and the p-value from the mixed model comparing average change over time between the treatment groups was 0.01.

Table 3: Trial 2 QLQ-C30 Global Health Score Missing Data Pattern

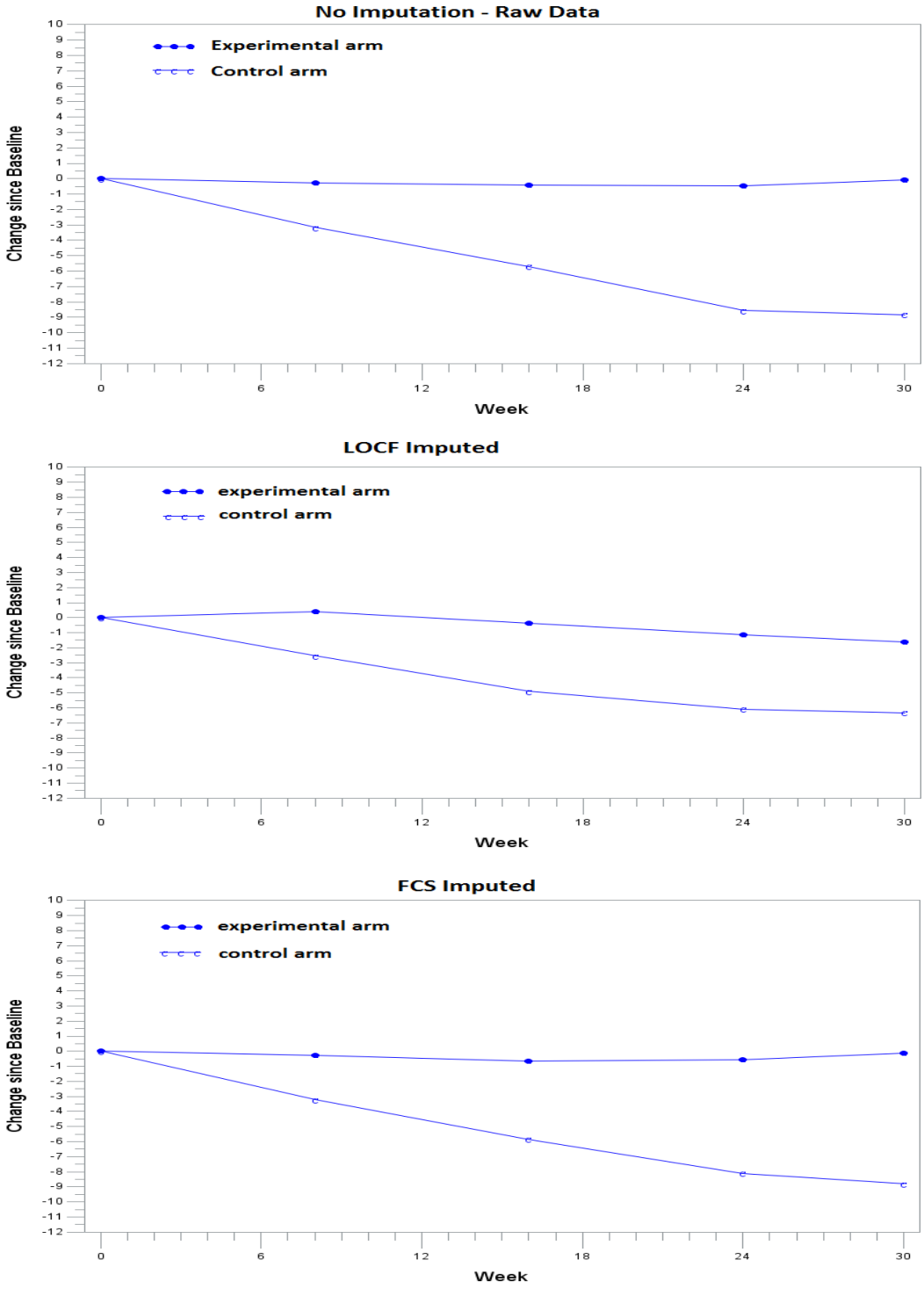
E (N=179)	**Week 0**	**Week 8**	**Week 16**	**Week 24**	**Week 30**
Available	168 (94%)	147 (82%)	107 (60%)	68 (38%)	51 (28%)
Missing		21	40	39	17
AE-related dropout		9	15	13	6
C (N=182)	**Week 0**	**Week 8**	**Week 16**	**Week 24**	**Week 30**
Available	174 (96%)	118 (65%)	65 (36%)	27 (15%)	14 (8%)
Missing		56	53	38	13
AE-related dropout		3	5	4	4

Figure 3: Trial 2 Primary MMRM Analysis for QLQ-C30 Global Health Score



The estimates for mean change in the global health score again were very similar between the ones from the original mixed model analysis without further data imputation and the ones from the multiple imputations. The LOCF analysis, like in Trial 1, had reduced the estimated difference in score change between the two treatment groups.

Figure 4: Trial 2 Raw Data versus Missing Data Imputed MMRM Estimates



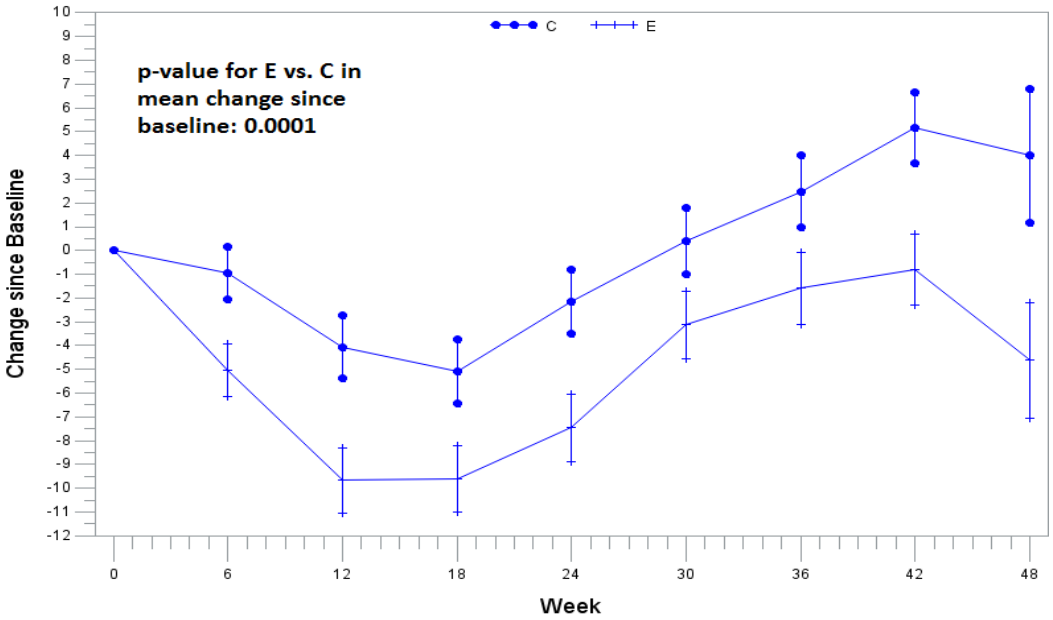
3.3 Results from Trial 3

Trial 3 was a double-blind trial in patients with multiple myeloma. The trial had a good proportion of patients dropped out due to adverse events in either treatment group. The AE-related dropout rate was 34% in the experimental group and was 17% in the control group. Unlike either Trial 1 or Trial 2, Trial 3 had a less PRO data completion rate in the experimental group, and the general quality of life was rated significantly lower by patients on the experimental group. Also important to notice was that about 50% of the missing data in the experimental group during the first 24-week treatment phase was adverse event related, and the self-rated general health score was increasing in patients who were allowed to continue to the second phase of the treatment.

Table 4: Trial 3 QLQ-C30 Global Health Score Missing Data Pattern

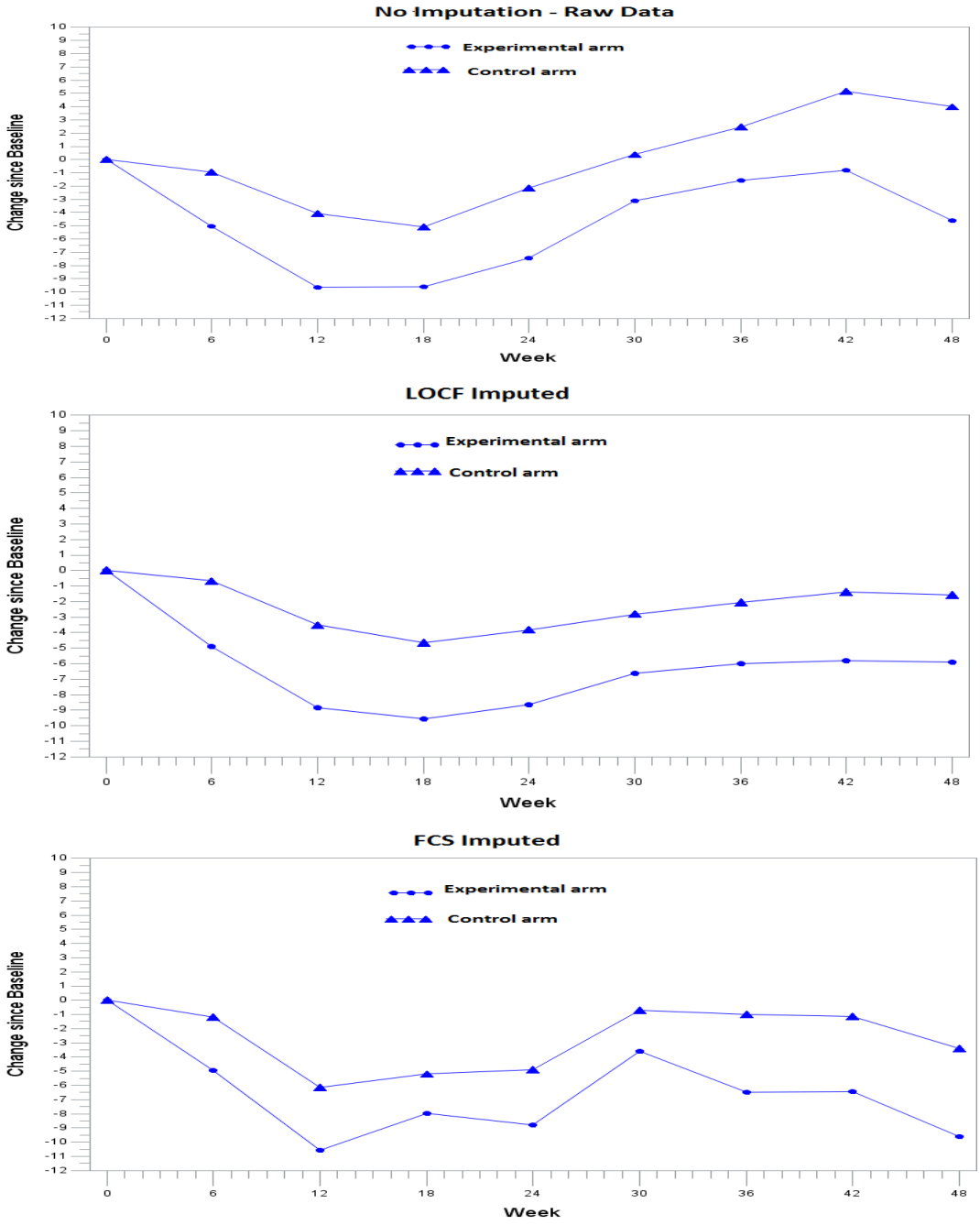
E (N=387)	**WK 0**	**WK 6**	**WK 12**	**WK 18**	**WK 24**	**WK 30**	**WK 36**	**WK 42**	**WK 48**
Available	382 (99%)	291 (75%)	217 (56%)	192 (50%)	151 (39%)	145 (37%)	121 (31%)	112 (29%)	38 (10%)
Missing		91	74	25	41	6	24	9	74
AE-related dropout		45	37	21	15	6	6	2	3
C (N=381)	**WK 0**	**WK 6**	**WK 12**	**WK 18**	**WK 24**	**WK 30**	**WK 36**	**WK 42**	**WK 48**
Available	367 (96%)	306 (80%)	246 (65%)	215 (56%)	181 (48%)	157 (41%)	124 (33%)	114 (30%)	27 (7%)
Missing		61	60	31	34	24	33	10	87
AE-related dropout		21	23	12	6	8	1	2	1

Figure 5: Trial 3 Primary MMRM Analysis for QLQ-C30 Global Health Score



The estimates for mean change in the global health score were not similar between the original and the imputed analyses. The LOCF analysis had brought down estimated improvement in quality of life during the second 24 weeks of treatment. The multiple imputation approach not only brought down the originally estimated improvement in the second phase of treatment, the degree of adjustment was different between treatment groups. And since the multiple imputation analysis also accounted for extent of treatment exposure at dropout, the estimates for the 1st phase of treatment were different as well.

Figure 6: Trial 3 Raw Data versus Missing Data Imputed MMRM Estimates



4. Summary and Discussion

One can agree that quality of life is just as important as quantity of life in cancer patients. FDA is encouraging oncology drug sponsors to monitor patient's symptoms and functions during treatment using patient reported health outcomes also known as PROs. However, because cancer treatments can cause significant toxicity, patient dropouts due to adverse events could be substantial.

The usually proposed primary analysis for a PRO is a MMRM analysis, without accounting for dropouts due to adverse events. This preliminary investigation evaluates whether or not imputations for missing values could serve as sensitivity analyses to the proposed primary MMRM analysis to see if the primary analysis is robust against the MAR assumption.

Using real data from 3 oncology registration trials, we compared between the observed data only and missing data imputed estimates from MMRM for mean change in PRO during expected treatment period. We imputed missing QLQ-C30 global health scores using the LOCF single imputation approach that operates under the MCAR assumption, as well as using the FCS multiple imputation approach that operates under the MAR assumption if a covariate that is predictive of missing outcome data is included in the imputation model. The covariates of interest that were included in the FCS multiple imputation models were the reason for dropout and the extent of treatment exposure. Both the LOCF and FCS imputation methods were commonly used, software-ready (e.g. SAS version 9) , and did not require all variables in the imputation model to follow a specific joint distribution[6,7].

The results we have found so far suggest that the FCS multiple imputations accounting for dropouts due to adverse events may be useful as a sensitivity analysis to justify the MAR assumption in the MMRM analysis for monitoring PRO data during treatment. The estimates from multiple imputations were similar to the ones from the original MMRM analysis in Trial 1 when the two treatment groups were comparable in AE-related dropout rates, and in Trial 2 when one treatment group had a much higher AE-related dropout rate but the occurrence of dropouts due to adverse events was similarly distributed from one visit to another during the treatment period. In Trial 3, where the cause for dropout appeared to have changed during the course of treatment, the FCS multiple imputations generated different estimates from the ones from the initial MMRM analysis. The LOCF single imputation approach had generated estimates that were different from the others, suggesting a MCAR assumption for missing PRO data was not likely to be valid in the studied trials.

Although the LOCF analysis estimated a smaller difference between treatment groups in QLQ-C30 global health score change, we did not intend to imply that the LOCF analysis was more conservative as the analysis also had under-estimated the uncertainty of variation. Another consideration was any occurrence of death would prevent assessments of quality of life in patients. We had repeated the analyses with imputations truncated to the last scheduled visit prior to death in patients who had died during treatment. The results were very similar (not shown here); possibly because not many patients died during the primary treatment period in the studied trials.

Finally, we wish to emphasize that this is an initial evaluation as a part of efforts to look into various issues with reporting patient-reported outcomes in product labels and further research is necessary to evaluate the impact of missing data in determining treatment effect based on a patient reported outcome.

Acknowledgements

We wish to acknowledge the Office of Hematology and Oncology Products (OHOP) for encouraging a better data collection and analysis of patient-reported outcomes, and Dr. Laura Lee Johnson for providing useful comments on the JSM draft presentation.

References

1. Siddiqui O, Hung HM, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *J Biopharm Stat* 2009; 19(2): 227-246.
2. Rubin D. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons, Inc.; 1987.
3. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JCJM, Kaasa S, Klee MC, Osoba D, Razavi D, Rofe PB, Schraub S, Sneeuw KCA, Sullivan M, Takeda F. The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute* 1993; 85: 365-376.
4. Mallincrrodt CH, Clark SWK, Carroll RJ, Molenberghs G. Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations. *J Biopharm Stat* 2003; 13(2): 179-190.
5. Raghunathan TE, Lepkowski JM, Van Hoewyk J, et al. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv Methodol* 2001; 27(1): 85-95.
6. SAS Institute Inc. The MI procedure. In: *SAS/STAT 9.3 User's Guide*. Cary, NC: SAS Institute Inc; 2011: 4552-4665.
7. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol* 2010; 171: 624-632.