Sensitivity Analysis of Missing Longitudinal Patient Reported Outcomes in Asthma

Tulin Shekar

Merck, 126 E. Lincoln Avenue, Rahway, N.J. 07065

Abstract

Patient reported outcomes are increasingly used in health research, including randomized controlled trials and observational studies. However, the validity of results in longitudinal studies can be problem on the handling of missing data. The sensitivity analyses on missing data will be provided using data from a randomized controlled clinical trial on asthma patients with Asthma Quality of Life Questionnaire endpoint.

Key Words: Missing data, quality of life, patient reported outcomes, asthma.

1. Introduction

Patient reported outcomes (PROs) are common endpoints for the clinical trials and epidemiologic research. Questionnaires are the main components of the PROs. Questionnaires are made up of items (questions) grouped into one or more subscales, or domains. For example, a widely used PRO for quality of life (AQLQ) assessment in asthma is Asthma Quality of Life Questionnaire (AQLQ) is a disease-specific health-related quality of life instrument that includes both physical and emotional impact of disease. Questionnaire has 32 items with 2-week recall, with 4 domains/categories. Categories include; Symptoms (11 items), Activity Limitation (12 items, 5 of which are individualized), Emotional Function (5 items), and Environmental Exposure (4 items). The items scaled as 7-point (7 = not impaired at all - 1 = severely impaired), higher scores indicate better quality of life. Minimally important difference in score of 0.5 for overall quality of life and for each of the individual domains. AQLQ can be interviewer or self- administered. Many studies which use PROs are longitudinal and are therefore likely to have missing data. For PROs this can be individual questions, termed missing items, or missing questionnaires, as might occur if a patient did not attend an assessment. As a result of missing data we will observe wider confidence intervals and decrease in power caused by the reduction in data. Another more serious impact will be bias in treatment effect.

In general, sensitivity analyses for addressing the missing data are rarely performed for PRO outcomes in clinical trials. The aim of this paper is to address the issue of missing data in PROs using sensitivity analyses.

1.1 Motivating Example

For the purposes of sensitivity analysis the data from a 26 week, multicenter randomized phase III trial in subjects ages 12 yrs and above and with persistent asthma previously treated with low-dose inhaled glucocorticsteroids were used. The treatments were Mometasone Furoate/Formoterol Fumarate Combination Formulation (MF/F), Mometasone Furoate (MF), Formoterol Monotherapy (F), and Placebo. The AQLQ was a key secondary endpoint for these trials. For this example concentration will be on MF/F and Placebo arms as they were the primary comparisons for the trial.

Table 1: Number of observations for AQLQ endpoint at each time-point

	MFF				Placebo			
Weeks	0	4	12	26	0	4	12	26
Time	1	2	3	4	1	2	3	4
Study1								
#Obs	179	174	159	144	188	171	139	117
Study2								
#Obs	190	183	165	154	195	182	139	121

Patients were assessed at 4 time-points: baseline, 4, 12, and 26 weeks. By the fourth assessment (26 weeks), 71% (MF/F=79%, Placebo=62%) of subjects had complete AQLQ data, similar proportions were observed for the second study, see Table 1.
We consider the endpoint Total AQLQ Score: average of the Symptoms, Activity, Emotion and Environment. It has been scaled to a range of 0–7 points, with higher values indicating better AQLQ. The minimum important difference for the AQLQ has been estimated to be 0.5 points in literature. Additional data that were collected were FEV1, Asthma Symptom Diaries and other Pulmonary Function Tests and baseline characteristics such as age, sex, race and weight.

To focus the discussion, the mean difference in Total AQLQ score at the fourth time-point (week 26) estimated using various approaches. Treatment estimates are summarized in Table 2, all mixed models used the Kenward-Roger degrees of freedom, which performs well for both small and large samples. To make comparisons between methods easier, imputation was done with a single EM imputation to make the incomplete data monotone. All analyses were performed in SAS® v9.3 on PC and SAS macro provided at missingdata.org.uk was utilized for the effect sizes provided in Table 2.

1.2 Missing Data Definitions and Descriptions

1.2.1 Missing items

One source of missing data is missing responses to individual questions. Many questionnaires have instructions regarding scoring when some items are missing, and the tool may not be valid or reliable unless these are followed. In general, missing items are infrequent and missing subscale scores due to missing items are even less frequent. In the event of a high proportion of missing items, the appropriateness of the particular questionnaire in the population being studied should be examined.

1.2.2 Patterns of missing data

If a patient drops out of the study, their data from a certain point onwards will be unobserved. This pattern of missingness is termed monotone. Intermittent missingness is defined to be when an outcome is unobserved at one assessment but is observed at a following assessment.

Suppose there are N independent patients with n planned assessments, and ni≥n observed values for patient i. Let Rij=1 if the outcome, Yij, is observed for patient i at time j. Let X represent covariates (such as treatment assignment). The vector of outcome observations can be partitioned into $Y = (Y^O, Y^M)$, where Y^O is the observed data and Y^M is the missing data. Then

MCAR = > P(Rij | Yij, Yij−1...Yil, X) = P(Rij | X) = P(R | X)
MAR = > P(Rij | Yij, Yij−1...Yil, X) = P(Rij | Yij−1...Yil, X) = P(R | Y^O, X)
MNAR = > P(Rij | Yij, Yij−1...Yil, X) = P(Rij | Yij, Yij−1...Yil, X) = P(R | Y^M, Y^O, X)

The likelihood of the complete data (without loss of generality, excluding X) is

$$f(Y^O, Y^M, R) = f(Y^O, Y^M) f(R \mid Y^O, Y^M),$$

and of the observed data is

$$f(Y^O, R) = f(Y^O, Y^M) f(R \mid Y^O, Y^M) dY^M.$$

The integral averages over all possible values of Y^M, weighted by the probability of their occurrence. If data are MAR, then

$$f(R \mid Y^O, Y^M) = f(R \mid Y^O), \text{ so}$$
$$f(Y^O, R) = f(R \mid Y^O)\, f(Y^O, Y^M) d Y^M = f(R \mid Y^O) f(Y^O).$$

$f(R|Y^O)$ gives no information about Y^O, so it can be ignored in the analysis. i.e., if data are MAR, the missingness mechanism does not need to be included in the modelling. A similar argument can be made for MCAR data.

2. First steps: Describing the extent and patterns of missingness:

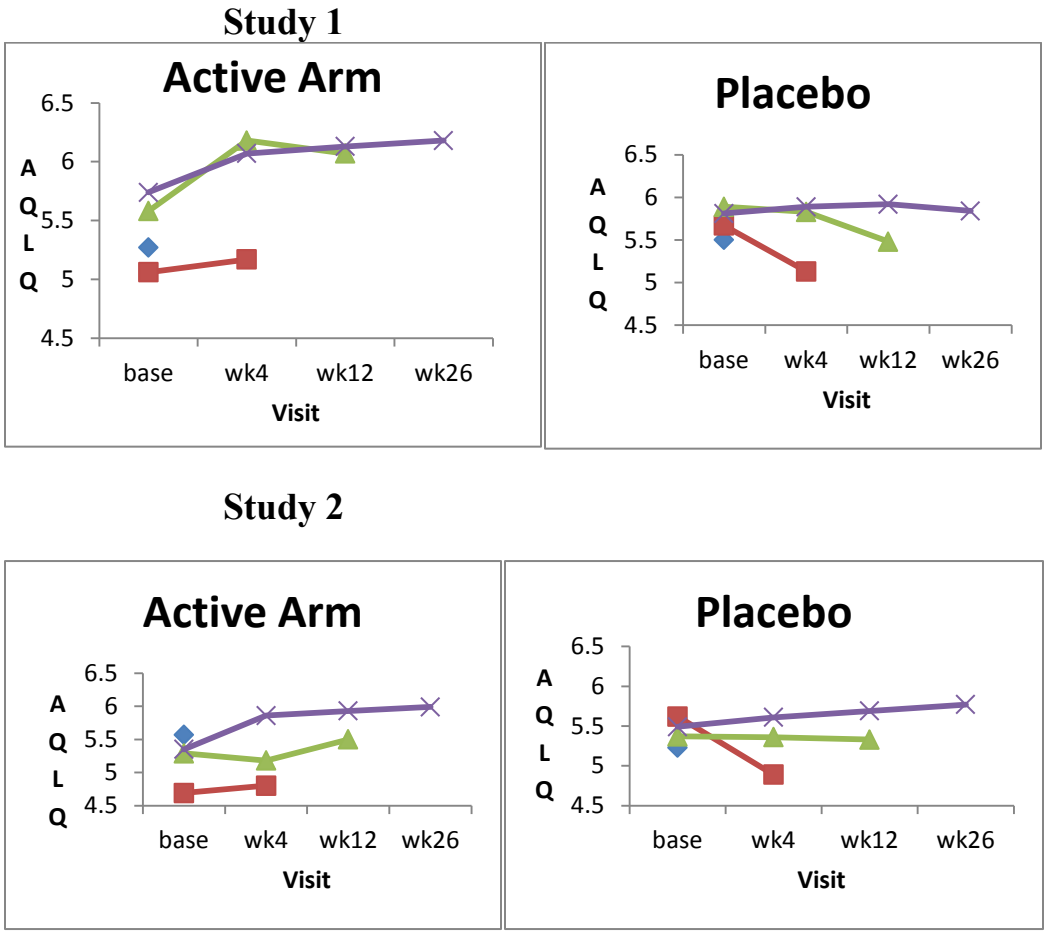

Figure 1: Missingness patterns: AQLQ stratified by treatment group and dropout time. The possible range of response is 0-7 points, with higher values better quality of life (AQLQ) response.

To explore mechanisms of missingness, a good place to start is a graph of the AQLQ versus time, stratified by dropout time, as shown in Figure 1. Missingness patterns that contrast the scores across the different times of dropout. For example, are patients who have lower baseline values more likely to drop out, or are steeper rates of increase or decrease over time associated with dropout? The separation between the missingness patterns is more dramatic in the placebo arm than active arm, thus we would expect more sensitivity across analyses in this arm.

2.1 Simple imputation and LOCF

Simple imputation refers to filling in missing observations with a single value. Here we report the individual's last observation carried forward (LOCF), the difference in AQLQ Total score between the MF/F and Placebo at the fourth time-point was 0.62 (ANCOVA with baseline AQLQ Total score as a covariate; $p<0.001$) as compared to observed cases difference of 0.38 ($p<0.001$) for Study 1 and for Study 2 , 0.50 (ANCOVA with baseline AQLQ Total score as a covariate; $p<0.001$) as compared to observed cases difference of 0.25 ($p=0.004$) (Table 2). For AQLQ measures, under LOCF, patients who drop out

would get the same AQLQ scores as their last assessment's AQLQ, which is improbable in many situations, as we are observing inflated differences compared to observed cases.

2.2 Complete case analysis

A complete case analysis occurs when only those participants with complete data for all assessments are analyzed. Data is thrown away, which is unethical and statistically inefficient. The difference in AQLQ between the MF/F and Placebo at the fourth time-point is 0.41($p<0.001$) for Study1 and 0.21($p=0.0182$) for Study2 and (ANCOVA, conditioning on baseline AQLQ). The smaller differences between the two arms are not unexpected as we are comparing only the selected patients who remain on the trial until the Week 26 assessment.

2.3 Maximum likelihood methods

Maximum likelihood refers to a parameter estimation method which is used in various longitudinal models including mixed models, latent variable modelling, item response theory and structural equation models (where it is often referred to as full information maximum likelihood). In randomized clinical trials, measurements are often collected on each subject at a baseline visit and several post-randomization time points. The longitudinal analysis of covariance in which the post-baseline values form the response vector and the baseline value is treated as a covariate can be used to evaluate the treatment differences at the post baseline time points. In a constrained longitudinal data analysis in which the baseline value is included in the response vector together with the post baseline values and a constraint of a common baseline mean across treatment groups is imposed on the model as a result of randomization. If the baseline value is subject to missingness, the constrained longitudinal data analysis is shown to be more efficient for estimating the treatment differences at post baseline time points than the longitudinal analysis of covariance. The efficiency gain increases with the number of subjects missing baseline and the number of subjects missing all post baseline values. Using terms with treatment, time, and the interaction of treatment with time and baseline spirometry as a covariate. The estimate of the difference in AQLQ at the fourth time-point between treatment groups was 0.52, $p<0.001$, and 0.36, $p<0.001$ for Study 1 and Study 2 respectively, indicating lower change from baseline in AQLQ for the placebo group.

2.4 Multiple Imputations

MI is based on filling in missing data by drawing from a distribution of likely values, and does not suffer from the variance underestimation of simple imputation. Suppose there are two observations on each participant: Y1 (which is completely observed) and Y2 (which is incomplete). The association between Y1 and Y2 from those participants who have complete data is used to fill in the missing values of Y2. Mathematically, we draw values of Y2 from the conditional distribution Y2|Y1. The implicit and un-testable assumption is that the relationship between Y1 and Y2 is the same for those who complete and those who do not. This is the MAR assumption. There are three steps to MI: (1) Impute multiple (M) times, using a regression model called the imputation model, so that there are M complete sets of data; (2) Analyze each of the M data sets; (3) Combine the results using Rubin's rules. Although a common recommendation is that only a few (3–5) imputations are needed, tried 20, 50,100 imputations, results were stable, reported the 100 imputation results. An MMRM was then fit to the multiply

imputed data, and the treatment estimate at the 26 Week was 0.52, p<0.001. This is close in magnitude to the MMRM without multiple imputations (0.52) for Study1, which suggests that the auxiliary variables are not contributing additional information for Study #1. Similar results were observed in the Study #2.

Table 2: Summary of estimates of the difference in quality of life (AQLQ) between experimental and control groups at Week 26

Method Study #1	Estimate (MF/F-Placebo)	SE	P value
ANCOVA	0.38	0.096	<0.001
ANCOVA-CC	0.41	0.110	<0.001
ANCOVA-LOCF	0.62	0.099	<0.001
Mixed Model	0.52	0.095	<0.001
Mixed Model-MI	0.52	0.097	<0.001
Mixed Model-CDC	0.49	0.097	<0.001
Mixed Model-LMCF	0.49	0.098	<0.001
Multivariate*	0.39	0.091	<0.001
Method Study #2	Estimate (MF/F-Placebo)	SE	P value
ANCOVA	0.25	0.086	0.004
ANCOVA-CC	0.21	0.087	0.0182
ANCOVA-LOCF	0.50	0.095	<0.001
Mixed Model	0.36	0.082	<0.001
Mixed Model-MI	0.35	0.085	<0.001
Mixed Model-CDC	0.35	0.084	<0.001
Mixed Model-LMCF	0.39	0.085	<0.001
Multivariate*	0.49	0.092	<0.001

SE: Standard Error,MI: Multiple Imputation,Models include:treatment,time,treatment*time, CDC:Copy Differences from Control, LMCF: Last Mean Carried Forward, *AQLQ and time on treatment

3. Conclusion

The excessive missing data for the PRO outcomes can be minimized with careful planning of design and analyses. In the above example, the wide range of estimates of treatment effect, from 0.38 to 0.62 points and 0.21 to 0.50 points, for Study 1 and Study 2 respectively shows that different modelling approaches rely on different assumptions. A general approach for primary analysis should be on MAR assumption then to perform post sensitivity analysis. Caution must be exercised for interpretation of the results. In addition recommend that the Statistical Analysis Plan should incorporate such an analysis.

Acknowledgements

I wish to thank my family for their support and encouragement throughout my research work.

References

http://www.missingdata.org.uk

Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epidemiol 1994; 47(1): 81-87.

Yuan, Y.(2011),Multiple Imputation Using SAS Software,Journal of SAS software, Vol 45, Iss. 6

Bell, M., Fairclough D. (2013), Practical and statistical issues in missing data for longitudinal patient reported outcomes,Statistical Methods in Medical Research, 1-20.

Barnard J, Rubin DB (1999). Small-Sample Degrees of Freedom with Multiple Imputation,Biometrika, 86, 948-955.

Brand JPL (1999), Development, Implementation and Evaluation of Multiple Imputation Strategies for the Statistical Analysis of Incomplete Data Sets Ph.D. thesis, Erasmus University Rotterdam.

Horton NJ, Lipsitz SR (2001). Multiple Imputation in Practice: Comparison of Software Packages for Regression Models with Missing Variables. Journal of the American Statistical Association, 55, 244-254.

Lavori PW, Dawson R, Shera D (1995). A Multiple Imputation Strategy for Clinical Trials with Truncation of Patient Data. Statistics in Medicine, 14, 1913-1925.
Little RJA, Rubin DB (2002). Statistical Analysis with Missing Data. 2nd edition. John Wiley & Sons

Rubin DB (1976). Inference and Missing Data. Biometrika, 63, 581-592.

SAS Institute Inc (2011a). SAS/STAT Product Documentation. SAS Institute Inc., Cary, NC.URL

SAS Institute Inc (2011b). SAS/STAT User's Guide - Procedures. SAS Institute Inc.,Cary, NC.