Developing an Analytic Road Map for Incomplete Longitudinal Clinical Trial Data

Adam L. Meyers, MS

Lilly

Answers That Matter.

Outline

- Missing data mechanisms
  - Background
  - Missingness in Clinical Trial Data
- Analysis of missing data
  - Last Observation Carried Forward (LOCF) via analysis of variance (ANOVA)
  - Mixed Model Repeated Measures (MMRM)
  - Missing Not at Random (MNAR) methods
- Recommendations and Conclusions

Missing Data Mechanisms[1]

Missing Completely at Random (MCAR):

Conditional on the independent variables in the model, neither the observed or unobserved outcomes of the dependent variable explain dropout

Missing at Random (MAR):

Conditional on the independent variables in the model, the observed outcomes of the dependent variable explain dropout, but the unobserved outcomes do not

Missing Data Mechanisms[1]

Missing Completely at Random (MNAR):

Conditional on the independent variables in the model and the observed outcomes of the dependent variable, the unobserved outcomes of the dependent variable explain dropout

Missing Data Mechanisms[2]

- The selection of the appropriate missing data mechanism depends not only on the data but also the analysis model
- Example:
  - Differential dropout by treatment group → missingness in data not random
  - Include treatment term in analytical model → dropout MCAR
- Pure MCAR
  - Missingness cannot be attributed to anything
- Covariate-dependent MCAR
  - Missingness depends on covariates

Missing Data Mechanisms

- Other terms such as *Ignorable Missingness* and *Informative Censoring* must also consider the analytical method
  - Ignorable missingness is defined as missingness that can be ignored because the observed data provides unbiased parameter estimates
  - What may be ignorable in a likelihood-based analysis may be non-ignorable in a frequentist-based analysis

Missingness in Clinical Trial Data

- Efficacy data in clinical trials are rarely MCAR because the observed outcomes influence dropout (i.e. discontinuation due to lack of efficacy)
- Clinical trials attempt to collect information to explain patient dropout
  - This may minimize MNAR data
- In the scenario of a highly-controlled clinical trial, data may be mostly MAR
- MNAR data can never be ruled out

Implications of Missingness

- All analyses rely on assumptions regarding missing data
- Clinical trial design features to minimize patient dropout should be strongly considered
- Analytical models can influence the missing data mechanism and should be considered when creating an analysis plan

Outline

- Missing data mechanisms
  - Background
  - Missingness in Clinical Trial Data
- Analysis of missing data
  - LOCF via ANOVA
  - MMRM
  - MNAR methods
- Recommendations and Conclusions

Analysis of Missing Data – LOCF (BOCF)

- Last observation carried forward is a single imputation method that assumes for patients with missing observations at endpoint, their responses at endpoint would have been the same as their last observed value

- Baseline observation carried forward (BOCF) is similar, assumes that responses at endpoint would have been the same as the baseline observed values for patients with no post-baseline observations. For patients with at least one post-baseline observation, operates the same way as LOCF

LOCF via ANOVA

- LOCF does not distinguish between observed and imputed data
- ANOVA, as a frequentist method, assumes a MCAR missing data mechanism
- Use of last observed data point yields a constant patient profile at all other unobserved later data points

Mixed Model Repeated Measures (MMRM)[3-5]

- MMRM, as a likelihood-based method, assumes a MAR missing data mechanism but holds under the assumption of MCAR as well

- Models fixed and random effects
  - In the clinical trial setting, treatment is an example of a fixed effect and patient is an example of a random effect

- MMRM includes the random effect of patient within the marginal covariance matrix (combination of within patient and between patient errors)

Mixed Model Repeated Measures (MMRM)[6]

- Controlling for random effects allows for better inference on fixed effects
  - Other types of mixed models handle random effects differently
- Model assumes patients who were improving at the time of dropout would continue to improve, and that, vice versa, patients who were worsening at the time of dropout would continue to worsen
- The trajectory of improvement or worsening after dropout is adjusted mathematically based on observed data from the patient and other patients

What's your primary? LOCF vs. MMRM[1,4,7]

- For clinical trial applications, let's first consider the missing data mechanism:
  - MAR could be considered reasonable given that missingness can often be explained by the observed data and the choice of statistical model
  - MAR holds under MCAR conditions, the converse is not true
  - As previously noted, clinical trials inherently may minimize MNAR data by being highly-controlled

What's your primary? LOCF vs. MMRM

- LOCF continues to be widely used as the primary analysis of mean change
- Why?
  - Perceived as a conservative approach
  - Concern over the performance of MAR methods such as MMRM in a MNAR setting
- Is there statistical evidence to address these issues?

Conservatism of LOCF

- LOCF underestimates within group changes whenever change increases over time
- LOCF overestimates within group change when change is greatest at intermediate time points
- While underestimating within group change is conservative in terms of improvement over time, it is anticonservative for worsening over time

Conservatism of LOCF

- Consider the following examples:
  - Alzheimer's disease – treatment administered to prevent symptom worsening (mental deterioration)
  - Depression relapse trials – treatment administered to prevent depression relapse
  - Vital signs
- LOCF would underestimate within group changes in each of these scenarios

Conservatism of LOCF[4,8-25]

- There is a significant body of literature that demonstrates LOCF leads to:
  - Biased estimates of treatment effects
  - Biased conclusions of no treatment effect in hypothesis testing
  - Underestimates of standard errors
  - Inflated Type I error
  - Varying observed coverage probabilities for CIs

Conservatism of LOCF[4,6,13-16,26,27]

- While LOCF may yield conservative estimates of within-group change, the primary goal in clinical trials is generally to compare between treatment groups

- Studies have demonstrated that many times LOCF does not act conservatively for between-group comparisons

- In a recent NDA:[28]
  - MMRM yielded a lower p-value than LOCF in 54.5% (110/202) of tests
  - LOCF yielded a lower p-value than MMRM in 34.2% (69/202) of tests
  - MMRM and LOCF yielded equal p-values in 11.4% (23/202) of tests
    - Due primarily to p<.001 outcomes

Performance of MMRM with MNAR data

- Several simulation studies demonstrate that MAR methods provide superior Type I and Type II error control versus LOCF in a setting with MNAR data
- MMRM and LOCF yield identical results when data sets are complete
- Differences exist when data is eliminated via a MNAR mechanism

Performance of MMRM with MNAR data[13]

- Study I:
  - MMRM compared with LOCF via ANOVA in scenarios where there was a true difference in mean change from baseline to endpoint between treatments
  - MMRM estimates of mean change closer to true values than LOCF estimates in every simulated scenario
  - LOCF underestimated standard errors, MMRM estimates were accurate
  - LOCF overestimated treatment differences when there was substantial placebo dropout
  - Expected CI coverage rate (percent of CIs that contain the true value) was 95%, MMRM yielded 94%, LOCF 87%

Performance of MMRM with MNAR data[14]

- Study II:
  - MMRM compared with LOCF via ANOVA in scenarios where there was no true difference in mean change from baseline to endpoint between treatments
  - Expected Type I error rate of 5%, MMRM yield 5.9% and LOCF 10.4%
  - Type I error rates across all scenarios ranged from 5.0% to 7.2% for MMRM and from 4.4% to 36.0% for LOCF

Performance of MMRM with MNAR data[15]

- Study III:
  - MMRM compared with LOCF via ANOVA in two scenarios where there was a true difference in mean change from baseline to endpoint between treatments and in two scenarios where there was no true difference
  - Autoregressive, Compound Symmetry, and Unstructured correlation structures were tested for MMRM in each scenario
  - Type I error rate from LOCF at least as great as from MMRM even when selecting the least appropriate covariance structure
  - MMRM with unstructured covariance matrix provided better Type I error control than LOCF in all scenarios (6.2% vs. 9.8%)

Performance of MMRM with MNAR data[15]

- Study III:
  - With a large true difference between treatments and a higher dropout rate in the superior treatment arm:
    - MMRM yielded an estimate of treatment difference of 12.6 vs. 9.1 for LOCF (true value of 12)
    - Power to detect difference between treatments was 75% for MMRM vs. 59% for LOCF
  - With a small true difference between treatments and a higher dropout rate in the inferior treatment arm:
    - MMRM yielded an estimate of treatment difference of 2.9 vs. 5.2 for LOCF (true value 4)
    - Power to detect difference between treatments was 10% for MMRM vs. 17% for LOCF

MNAR Methodology

- Classes exist that differ via the factorization of the likelihood functions for the joint distribution of the outcome variable and the missingness indicator variable
- Commonly referred to as the measurement process (observed data) and the missingness process (unobserved data)
- MNAR analyses are only as good as the assumed model

MNAR Methodology[1,29-31]

- Selection Models
  - Likelihood function product of the marginal density of the measurement process and the density of the missingness process conditional on the outcomes
  - Can be parametric on non-parametric
  - Consider as a multivariate analysis modeling the main outcome (i.e. mean change analysis) and dropout (categorical analysis)

MNAR Methodology[32,33]

- Pattern-Mixture Models
  - Likelihood function product of the marginal density of the measurement process conditional on the drop-out pattern and the density of the missingness process
  - Model the outcome variable separately for different patterns often based on time of dropout
  - Combine patterns for inference

MNAR Methodology[2,34-38]

- Shared-Parameter Models
  - Likelihood function product of the marginal density of the measurement process and the density of the missingness process, both conditional on a parameter that influences both the outcome and dropout
  - Conditional on the parameter, generally a random effect, the measurement and missingness processes are independent

MNAR Methodology as the Primary?

- Rubin (1994): "…even inferences for the data parameters generally depend on the posited missingness mechanism, a fact that typically implies greatly increased sensitivity of inference."[39]

- Laird (1994): "…estimating the unestimable can be accomplished only by making modeling assumptions…The consequences of model misspecification will be more severe in the non-random case."[40]

- Molenberghs, Kenward, and Lesaffre (1997): "…conclusions are conditional on the appropriateness of the assumed model, which in a fundamental sense is not testable."[41]

Outline

- Missing data mechanisms
  - Background
  - Missingness in Clinical Trial Data
- Analysis of missing data
  - LOCF via ANOVA
  - MMRM
  - MNAR methods
- Recommendations and Conclusions

Recommendations for Moving Forward

- Use MAR methods such as MMRM for primary analysis purposes
- Practice inclusive modeling – add ancillary variables that may help explain missingness to make data MAR rather than MNAR
- Including such ancillary variables in a MAR analysis such as MMRM may improve estimates, Type I error control, and power[42]

Recommendations for Moving Forward

- Implement MNAR methods as sensitivity analyses and to test for local influence
- Local influence identifies potentially influential data points and examines the effect of such points
- In particular, several newer approaches exist for selection models[43-49]

References

1. Little R, Rubin D. *Statistical Analysis with Missing Data.* New York: John Wiley & Sons; 1987.
2. Little RJA. Modeling the drop-out mechanism in repeated measures studies. *J. Am. Stat. Assoc.* 1995;90(431):1112–1121.
3. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat. Med.* 1997;16(20):2349–2380.
4. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* New York: Springer; 2000.
5. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38(4):963–974.
6. Little R, Yau L. Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics* 1996;52(4):1324–1333.
7. Rubin DB, Stern HS, Vehovar V. Handling "don't know" survey responses: the case of the Slovenian plebiscite. *J. Am. Stat. Assoc.* 1995;90(431):822–828.
8. Molenberghs G, Kenward MG. *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons; 2007.
9. Gadbury GL, Coffey CS, Allison DB. Modern statistical methods for handling missing repeated measurements in obesity trials: beyond LOCF. *Obes. Rev.* 2003;4(3):175-184.
10. Leon AC, Mallinckrodt CH, Chuang-Stein C, Archibald DG, Archer GE, Chartier K. Attrition in randomized controlled clinical trials: methodological issues in psychopharmacology. *Biol. Psychiatry* 2006;59(11):1001-1005.
11. Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, Carroll RJ. Analyzing incomplete longitudinal clinical trial data. *Biostatistics* 2004;5(3):445-464.
12. Liu G, Gould AL. Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *J. Biopharm. Stat.* 2002;12(2):207-226.
13. Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. *J. Biopharm. Stat.* 2001;11(1-2):9–21.
14. Mallinckrodt CH, Clark WS, David SR. Type I error rates from mixed effects model repeated measures versus fixed effects ANOVA with missing values imputed via last observation carried forward. *Drug Inf. J.* 2001;35:1215–1225.

References

15. Mallinckrodt CH, Kaiser CJ, Watkin JG, Molenberghs G, Carroll RJ. The effect of correlation structure on treatment contrasts estimated from incomplete clinical trial data with likelihood-based repeated measures compared with last observation carried forward ANOVA. *Clin. Trials* 2004;1(6):477–489.

16. Mallinckrodt CH, Kaiser CJ, Watkin JG, Detke MJ, Molenberghs G, Carroll RJ. Type I error rates from likelihood-based repeated measures analyses of incomplete longitudinal data. *Pharm. Stat.* 2004;3(3):171-186.

17. Gibbons RD, Hedeker D, Elkin I, Waternaux C, Kraemer HC, Greenhouse JB, Shea MT, Imber SD, Sotsky SM, Watkins JT. Some conceptual and statistical issues in analysis of longitudinal psychiatric data. Application to the NIMH treatment of Depression Collaborative Research Program dataset. *Arch. Gen. Psychiatry* 1993;50(9):739–750.

18. Heyting A, Tolboom JT, Essers JG. Statistical handling of drop-outs in longitudinal clinical trials. *Stat. Med.* 1992;11(16):2043–2061.

19. Lavori PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. *Stat. Med.* 1995;14(17):1913–1925.

20. Siddiqui O, Ali MW. A comparison of the random-effects pattern mixture model with last-observation-carried-forward (LOCF) analysis in longitudinal clinical trials with dropouts. *J. Biopharm. Stat.* 1998;8(4):545-563.

21. Shao J, Zhong B. Last observation carry-forward and last observation analysis. *Stat. Med.* 2003;22(15):2429-2441.

22. Carpenter J, Kenward M, Evans S, White I. Last observation carry-forward and last observation analysis. Letter to the Editor. *Stat. Med.* 2004;23(20):3241–3244.

23. Cook RJ, Zeng L, Yi GY. Marginal analysis of incomplete longitudinal binary data: a cautionary note on LOCF imputation. *Biometrics* 2004;60(3):820–828.

24. Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. New York: Springer; 2005.

25. Beunckens C, Molenberghs G, Kenward MG. Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clin. Trials* 2005;2(5):379–386.

26. Lane PW. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharm. Stat. (early view)* 2007;DOI: 10.1002/pst.267.

References

27. Lavori PW. Clinical trials in psychiatry: should protocol deviation censor patient data? *Neuropsychopharmacology* 1992;6(1):39–48.

28. Mallinckrodt CH, Raskin J, Wohlreich MM, Watkin JG, Detke MJ. The efficacy of duloxetine: a comprehensive summary of results from MMRM and LOCF_ANOVA in eight clinical trials. *BMC Psychiatry* 2004; 4:26.

29. Rubin DB. Inference and missing data. *Biometrika* 1976;63(3),581–592.

30. Diggle PD, Kenward MG. Informative dropout in longitudinal data analysis (with discussion). *Appl. Stat.* 1994;43,49–93.

31. Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *J. Am. Stat. Assoc.* 1998;93:1321–1339.

32. Little RJA. Pattern-mixture models for multivariate incomplete data. *J. Am. Stat. Assoc.* 1993;88(421):125–134.

33. Little RJA. A class of pattern-mixture models for normal incomplete data. *Biometrika* 1994;81(3):471–483.

34. Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* 1988;44:175–188.

35. Ten Have TR, Kunselman AR, Pulkstenis EP, Landis JR. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* 1998;54(1):367–383.

36. Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* 1989;45(3):939–955.

37. Mori M, Woodworth GG, Woolson RF. Application of empirical Bayes inference to estimation of rate of change in the presence of informative right censoring. *Stat. Med.* 1992;11(5):621–631.

38. Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. *Biometrics* 1995;51(1):151–168.

39. Rubin DB. Discussion to Diggle PJ, Kenward MG. Informative dropout in longitudinal data analysis. *Appl. Stat.* 1994;43:80:82.

References

40. Laird NM. Discussion to Diggle PJ, Kenward MG. Informative dropout in longitudinal data analysis. *Appl. Stat.* 1994;43:84.

41. Kenward MG. Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Stat. Med.* 1998;17(23):2723-2732.

42. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol. Methods* 2001;6(4):330-351.

43. Shen S, Beunckens C, Mallinckrodt C, Molenberghs G. A local influence sensitivity analysis for incomplete longitudinal depression data. *J. Biopharm. Stat.* 2006;16(3):365-384.

44. Zhu HT, Lee SY. Local influence for incomplete-data models. *Journal of the Royal Statistical Society B* 2001;63:111-126.

45. Verbeke G, Molenberghs G, Thijs H, Lesaffre E, Kenward MG. Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics* 2001;57(1):7-14.

46. Thijs H, Molenberghs G, Verbeke G. The milk protein trial: influence analysis of the dropout process. *Biom. J.* 2000;42(5):617-646.

47. Molenberghs G, Verbeke G, Thijs H, Lesaffre E, Kenward M. Mastitis in dairy cattle: local influence to assess sensitivity of the dropout process. *Comput. Stat. Data Anal.* 2001;37(1):93-113.

48. Troxel AB, Ma G, Heitjan DF. An index of local sensitivity to nonignorability. *Statistica Sinica* 2004;14:1221-1237.

49. Ma G, Troxel AB, Heitjan DF. An index of local sensitivity to nonignorable drop-out in longitudinal modeling. *Stat. Med.* 2005;24(14):2129-2150.