Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials

Craig Mallinckrodt (cmallinc@lilly.com)
Peter Lane (peter.w.lane@gsk.com)

DIA Virtual Journal Club
27 January 2008

Published in 2008 in the Drug Information Journal 42:303–319.

Outline

- Introduction
- Theory and concepts
- Regulatory concerns
- Handling MNAR data
- Recommendations

Overview

- This position paper summarizes relevant theory and current practice regarding the analysis of longitudinal clinical trial data intended to support regulatory approval of medicinal products
- Consensus of expert working team from PhRMA Efficiency in Clinical Trials initiative

Acknowledgements

PhRMA Expert Team on Missing Data

Peter Lane	**GSK**
Craig Mallinckrodt	**Lilly**
James Mancuso	**Pfizer**
Yahong Peng	**Merck**
Dan Schnell	**P&G**

Acknowledgements

We benefited from review of the draft paper by three prominent academics in the field of missing data and longitudinal analyses, and we thank them for their thoughtful comments and suggestions:

Rod Little (University of Michigan School of Public Health), Geert Molenberghs (Hasselt University Centre for Statistics), and Daniel Scharfstein (Johns Hopkins Bloomberg School of Public Health).

We are also grateful for review and advice from several colleagues in the pharmaceutical industry: Bruce Binkowitz (Merck), Argyha Chattopadhyay (J&J), David Keller (Pfizer), Frank Liu (Merck), Kaifeng Lu (Merck), Edmund Luo (Merck), Akiko Okamoto (J&J), and James Roger (GSK).

Although we incorporated many of the suggestions made by these reviewers, the individuals mentioned above were not asked specifically to endorse the recommendations made in this paper.

Outline

- Introduction
- Theory and concepts
- Regulatory concerns
- Handling MNAR data
- Recommendations

Medical Needs

- **Every hour we expect (approximately)**

15 new diagnoses of schizophrenia
200 deaths due to cancer
30 osteoporosis related hip fractures
2000 new diagnoses of anxiety disorders
1500 surgeries requiring pain treatment
70 deaths due to cardiovascular disease

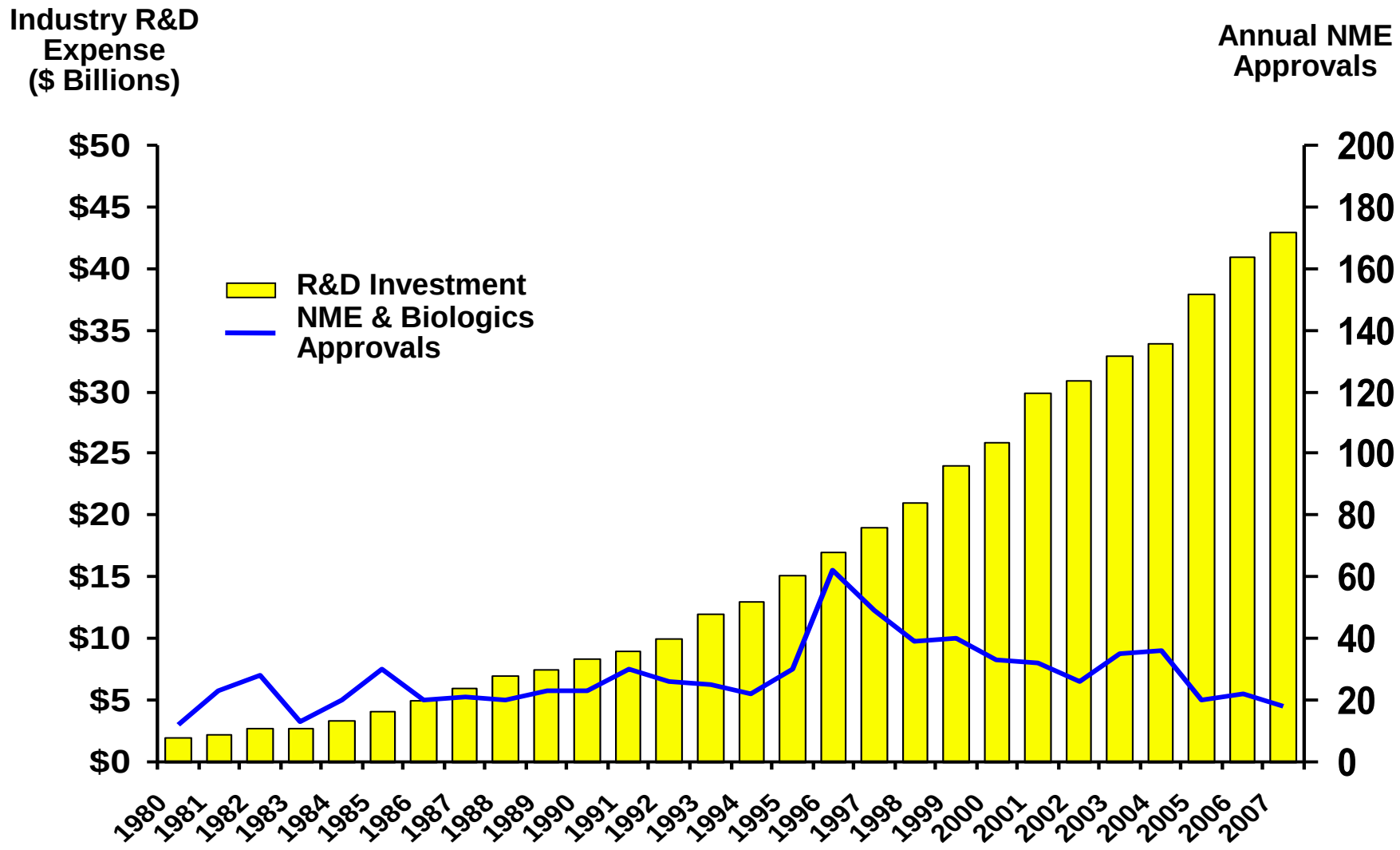
Alan Breier (Chief Medical Officer, Lilly) – Nov 2006

Need for More Effective Medicines

Therapeutic Area	Efficacy rate(%)
Alzheimer's	30
Analgesics (Cox-2)	80
Asthma	60
Cardiac Arrhythmias	60
Depression (SSRI)	62
Diabetes	57
HCV	47
Incontinence	40
Migraine (acute)	52
Migraine (prophylaxis)	50
Oncology	25
Osteoporosis	48
Rheumatoid arthritis	50
Schizophrenia	60

Trends in Molecular Medicine
7(5):201-204, 2001

R&D Productivity Decreasing



Source: PhRMA, FDA, Lehman Brothers; *[Dr. Robert Ruffolo]*

Current State of Clinical Research

- **Half of all drugs in Phase III fail**
- **Half the trials of known effective active drugs failed to differentiate from placebo**
- **Implies an unduly high rate of false positive and false negative results**

Current State of Longitudinal Analyses

- **Treatment effects often evaluated by comparing change over time**
- **Valid analyses can be problematic**
- **(B)LOCF common for handling missing data**
  - **Simple and perceived to be "conservative"**
- **Recent advances in theory and implementation make methods with far less restrictive assumptions than (B)LOCF readily accessible**

Current State of Longitudinal Analyses

- **Likelihood-based repeated measures has desirable theoretical and practical attributes**
- **MMRM (Mixed Model Repeated Measures) has been studied extensively in the context of longitudinal clinical trials**
- **Strong theoretical and empirical evidence favoring MMRM over LOCF**
- **LOCF still used extensively**

Outline

- Introduction
- Theory and concepts
- Regulatory concerns
- Handling MNAR data
- Recommendations

Starting Point

- No universally best method for analyzing longitudinal data
- This implies analysis must be tailored to the specific situation at hand
- Consider the desired attributes of the analysis and the characteristics of the data
  - Mechanism giving rise to the missing data

Missing Data Mechanisms

MCAR - missing completely at random

- **Conditional on the independent variables in the model, neither observed nor unobserved outcomes of the dependent variable are associated with drop-out**

MAR - missing at random

- **Conditional on the independent variables in the model, observed outcomes of the dependent variable are associated with drop-out, but unobserved outcomes aren't**

Missing Data Mechanisms

MNAR - missing not at random

- **Conditional on the independent variables in the model and the observed outcomes of the dependent variable, the unobserved outcomes of the dependent variable are associated with drop-out**

Consequences

- **Missing data mechanism is a characteristic of the data AND the model**
- **Differential drop-out by treatment indicates covariate dependence, not mechanism**
- **Mechanism can vary from one outcome to another in the same dataset**

Consideration

- **Ignorability of missingness depends on both the mechanism and the method**
- **E.g., for missingness to be ignorable in a likelihood-based analysis the mechanism must be MCAR or MAR, whereas in GEE the mechanism must be MCAR for the missingness to be ignorable**

Missing Data in Clinical Trials

- **Efficacy data in clinical trials are seldom MCAR because the observed outcomes typically influence drop-out (DC for lack of efficacy)**
- **Trials are designed to observe all the relevant information, which minimizes MNAR behavior**
- **Hence in the highly controlled scenario of clinical trials missing data may be mostly MAR**
- **MNAR can never be ruled out**

Implications

- **All analyses rely on missing data assumptions**
- **Any options in the trial design to minimize drop-out should be strongly considered**
- **Analytic models influence missingness mechanism and the analysis plan should capitalize on this fact**

Assumptions

- **ANOVA with LOCF assumes**
  - **MCAR**
  - **Constant profile**
- **Likelihood-based analyses assume**
  - **MAR (observed data are valid predictors of unobserved data)**
- **MAR always more plausible than MCAR**
  - **MCAR is a subset of MAR**
  - **MAR valid in every case where MCAR is valid but MCAR not always valid when MAR is valid**

MMRM Defined

- **One of the members of a broader class of direct likelihood-based analyses**
- **Makes use of fully and partially observed sequences via within-patient covariances**
- **Distinguishing feature is that the random effects are modeled as part of the within-patient error correlation structure**

The Mixed-Effects Model

$$Y = X\beta + Zb + \varepsilon$$

$$b_i \sim N(0,D)$$

$$\varepsilon_i \sim N(0,\Sigma)$$

$b_1 \ldots b_n$ and $\varepsilon_1 \ldots \varepsilon_n$ independent

$$Y \sim N(X\beta, V)$$

$$V = ZDZ' + \Sigma$$

Random Effects in MMRM

recall $V = ZDZ' + \sum$

- **Given an unstructured V, the dispersion of Y is invariant to the modeling choice of the random effects**
- **In other words we don't explicitly model the random effects. We let them go to error and model them as part of the within-subject correlations**

In SAS Terms

Consider the following two models in SAS PROC MIXED:

1. **class therapy time patient;**
   model Y = therapy time therapy*time;
   repeated /subject=patient type=un;

2. **class therapy time patient;**
   model Y = therapy time therapy*time;
   repeated /subject=patient type=un;
   random int /subject=patient;

Model 1 and Model 2 yield identical fixed-effects results

Data Characteristics Influencing MMRM

- **Confirmatory trials have a comparatively small number of fixed measurement times compared with the number of subjects, hence a full multivariate approach (multivariate normal approach) is possible**
  - **Unstructured modeling of time and correlation**
  - **More parsimonious models for time and correlation are easily implemented in the MMRM framework**

MMRM: Full Multivariate Approach

```
proc mixed;
    class subject treatment time site;
    model Y  = baseline treatment time site
                  treatment*time baseline*time/ddfm=kr;
    repeated time /  sub = subject type = un;
    lsmeans treatment*time / cl diff;
    run;
```

MMRM: Structured Modeling of Correlation

```
proc mixed;
    class subject treatment time site;
    model Y  = baseline treatment time site
                  treatment*time baseline*time/ddfm=kr;
    repeated time /  sub = subject type = arh(1);
    lsmeans treatment*time / cl diff;
    run;
```

MMRM: Structured Modeling of Time

In a previous data step set time2 = time

```
proc mixed;
    class subject treatment time2 site;
    model Y  = baseline treatment time site
                  treatment*time baseline*time/ddfm=kr;
    repeated time2 /  sub = subject type = un;
    lsmeans treatment / cl diff at time = t1;
    lsmeans treatment / cl diff at time = t2;
     lsmeans treatment / cl diff at time = tx….;
    run;
```

MMRM: Structured Modeling of Time and Correlation

In a previous data step, set time2 = time

```
proc mixed;
    class subject treatment time2 site;
    model Y  = baseline treatment time site
                  treatment*time baseline*time/ddfm=kr;
    repeated time2 /  sub = subject type = arh(1);
    lsmeans treatment / cl diff at time = t1;
    lsmeans treatment / cl diff at time = t2;
    lsmeans treatment / cl diff at time = tx….;
run;
```

Last Observation Carried Forward (LOCF)

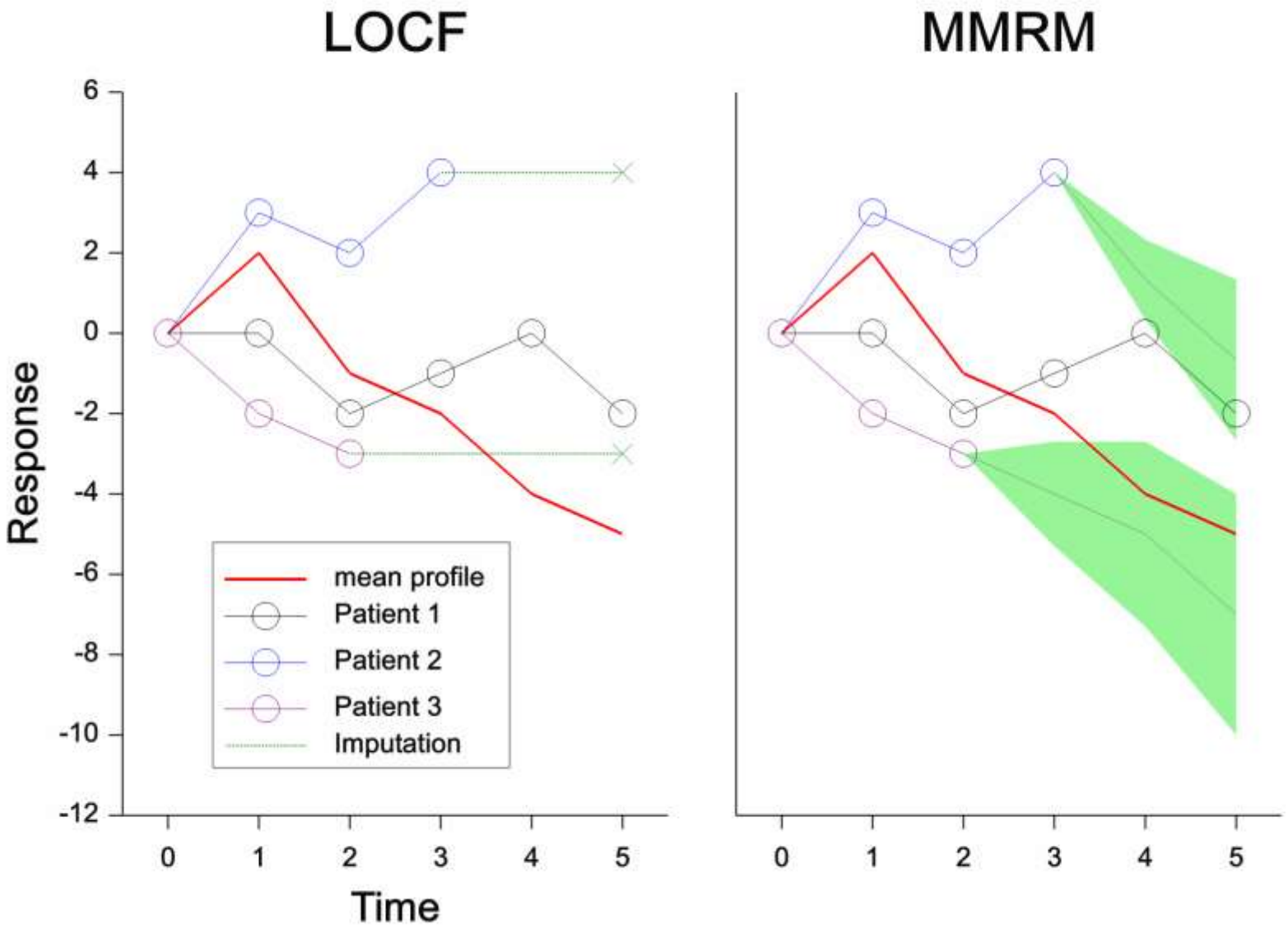
			Visit						
ID	Trt	BL	1	2	3	4	5	6	LOCF
1	1	22	20	18	16	14	12	10	-12
2	1	22	21	18	15	12	9	6	-16
3	1	22	22	21	20	*	*	*	-2
4	2	20	20	20	20	21	21	22	2
5	2	21	22	22	23	24	25	25	4
6	2	18	19	21	*	*	*	*	3

MMRM

			Visit					
ID	Trt	BL	1	2	3	4	5	6
1	1	22	20	18	16	14	12	10
2	1	22	21	18	15	12	9	6
3	1	22	22	21	20	*	*	*
4	2	20	20	20	20	21	21	22
5	2	21	22	22	23	24	25	25
6	2	18	19	21	*	*	*	*

MMRM sees that subject 3 was improving, but more slowly than subjects 1 and 2. Means at visits 4-6 are based on projected slow improvement of subject 3.

Graphical comparison of methods



Research Showing MAR is Useful and / or Better than LOCF

1. Arch. Gen. Psych. 50: 739–750.
2. Arch. Gen. Psych. 61: 310–317.
3. Biol. Psychiatry. 53: 754–760.
4. Biol. Psychiatry. 59: 1001–1005.
5. Biometrics. 52: 1324–1333.
6. Biometrics. 57: 43–50.
7. Biostatistics. 5:445–464.
8. BMC Psychiatry. 4: 26–31.
9. Clinical Trials. 1: 477–489.
10. Computational Statistics and Data Analysis. 37: 93–113.
11. Drug Information J. 35: 1215–1225.
12. J. Biopharm. Stat. 8: 545–563.
13. J. BioPharm. Stat. 11: 9–21.

Research Showing MAR is Useful and / or Better than LOCF

14. J. Biopharm. Stat. 12: 207–212.
15. J. Biopharm. Stat. 13:179–190.
16. J. Biopharm. Stat. 16: 365–384.
17. Neuropsychopharmacol. 6: 39–48.
18. Obesity Reviews. 4:175–184.
19. Pharmaceutical Statistics. 3:161–170.
20. Pharmaceutical Statistics. 3:171–186.
21. Pharmaceutical Statistics. 4: 267–285.
22. Pharmaceutical Statistics. 7:93–106.
23. Statist. Med. 11: 2043–2061.
24. Statist. Med. 14: 1913–1925.
25. Statist. Med. 22: 2429–2441.

A Note on Multiple Imputation

- **MI and MMRM yield asymptotically similar results when implemented with similar models**
- **We focused on MMRM because it has been studied more extensively in the context of the primary analysis in confirmatory trials**

Outline

- Introduction
- Theory and concepts
- Regulatory concerns
- Handling MNAR data
- Recommendations

Reasons why LOCF is Used as the Primary Analysis

- **LOCF perceived to be conservative**
- **Concern over how MAR methods perform under MNAR**
- **Simplicity: more explicit modeling choices needed in MAR methods**
- **LOCF thought to measure something more valuable**

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**
- **Underestimating within group change is conservative for progressive improvement but anti conservative for progressive impairment (Alzheimer's, maintenance studies)**

Conservatism of LOCF

- **Primary analyses are based on contrasts BETWEEN groups**
- **Bias in LOCF has been shown analytically and empirically to be influenced by many factors**
- **Direction and magnitude of bias highly situation-dependent and difficult to anticipate**

Biostatistics. 5:445-464.

Results From a Recent NDA

- **MMRM yielded a lower P value than LOCF in 110/202 tests (54.5%)**
- **LOCF yielded a lower P value than MMRM in 69/202 tests (34.2%)**
- **Methods yielded equal p values in 23/202 tests (11.4%)** **(usually as a result of both being < .001)**

BMC Psychiatry. 4: 26-31.

Performance of MMRM in MNAR Data

- **Obviously MAR methods can be biased by MNAR data**
- **Real question is how does MAR perform relative to MCAR when data are MNAR**
  - **MAR performs better than MCAR when data are MNAR**

References Comparing MMRM and LOCF in MNAR Data

Journal of Biopharm. Statistics (2001) 11:9–21

Drug Information Journal (2001) 35:1215–25

Pharmaceutical Statistics (2004) 3:171–186

Archives of General Psychiatry (2004) 61:310–317

Clinical Trials (2004) 1:477–489

Pharmaceutical Statistics (2007) 7:215–225

Pharmaceutical Statistics (2008) 7:93–106

Specific Simulation Study Results

- **Identical endpoint contrasts in complete data**
- **Type I error rates (true null, MNAR)**
  - **MMRM 5.79%, range 4.65% – 7.17%**
  - **LOCF 10.79%, range 4.43% – 36.30%**

DIJ (2001) 35:1215–25

- **CI coverage (false null, MNAR)**
  - **MMRM 94.24%, range 92% – 95%**
  - **LOCF 86.88%, range 51% – 95%**

J Biopharm Stat. (2001) 11:9–21

Specific Simulation Study Results

- **Bias in treatment difference**
  - **63 MNAR scenarios based on 7 real trials**
  - **MMRM had less bias than LOCF for 73%**

Pharmaceutical Statistics (2008) 7:93–106

More Explicit Modeling Choices Needed

- **Convergence in MMRM is not a problem**
  - **Prepare data properly**
  - **Use software features such as input starting values for parameter estimates and Fisher-scoring for initial rounds of iteration**
- **Even egregiously misfitting the correlation structure provided better control of Type I and Type II error than LOCF**

Clinical Trials. 1: 477–489.

Specifying An MMRM Analysis

Mean changes from baseline will be analyzed using a restricted maximum likelihood (REML)-based repeated measures approach. Analyses will include the fixed, categorical effects of treatment, investigative site, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariates of baseline score and baseline score-by-visit-interaction. An unstructured (co)variance structure will be used to model the within-patient errors. If this analysis fails to converge, the following structures will be tested: (insert a list of structures appropriate for the specific application). The (co)variance structure converging to the best fit, as determined by Akaike's information criterion, will be used as the primary analysis. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. Significance tests will be based on least-squares means using a two-sided α=0.05 (two-sided 95% confidence intervals). Analyses will be implemented using (insert software name). The primary comparison will be the contrast between treatments at (insert time point).

LOCF Thought to Measure Something More Valuable

- **LOCF is “factual”, MAR is “counterfactual”**
  - **LOCF is what is actually observed**
  - **MAR is what is estimated to happen if patients stayed on study**
- **LOCF said to assess effectiveness, MAR assesses efficacy**

Interpreting an LOCF Result

- **Abbreviation implies explicit imputation of missing values with inference drawn at specific time points**
- **Non-longitudinal interpretation of LOCF is common**
  - **LO, LAV**
  - **Drop-out is an outcome**

Non-longitudinal Interpretation of LOCF

- **An LOCF result can be interpreted as an index of rate of change and duration on study drug**
  - **A composite of efficacy, safety, and tolerability**
  - **Intuitively appealing**
  - **Simple**

Non-longitudinal Interpretation of LOCF

- **An LOCF result can be interpreted as an index of rate of change and duration on study drug**
    - **A composite of efficacy, safety, tolerability**
- **An index with unknown weightings**
- **The same estimate of mean change via LOCF can imply different clinical profiles**
- **The LOCF penalty is not necessarily proportional to the risk**
- **Result can be manipulated by design and behavior – So what parameter or hypothesis is being tested?**

Outline

- Introduction
- Theory and concepts
- Regulatory concerns
- Handling MNAR data
- Recommendations

Modeling Philosophies

- **Restrictive modeling**
  - **Simple models with few independent variables**
  - **Often include only the design factors of the experiment**

Psychological Methods, 6, 330-351.

Modeling Philosophies

Inclusive modeling

- **Auxiliary variables included to improve performance of the missing data procedure – expand the scope of MAR**
  - **Baseline covariates**
  - **Time-varying post-baseline covariates: Must be careful to not dilute treatment effect. Can be dangerous to include time-varying post-baseline covariates in analysis model; may be better to use via imputation (or propensity scoring or weighted analyses)**

Psychological Methods, 6, 330-351.

Rationale for Inclusive Modeling

- **MAR: conditional on the dependent and independent variables in the analysis, unobserved values of the dependent variable are independent of drop-out**
- **Hence adding more variables that explain drop-out can make missingness MAR that would otherwise be MNAR**

Common MNAR methods

- **General classes of MNAR methods based on different factorizations of the likelihood functions for the joint distribution of the outcome variable and the indicator variable for whether or not a data point is observed**
- **Factorization: hypothetical "full" data split into two parts: the actually observed part and the missing part, which are often described as the measurement process and the missingness process, respectively**

Selection Models

- **Full data likelihood is the product of the marginal density of the measurement process and the density of the missingness process conditional on the outcomes**
- **Conceptually, a multivariate analysis. One variable is the same outcome as in an MAR analysis. The second outcome is the indicator variable for drop-out (analyzed via logistic regression) or time to drop-out (analyzed via proportional hazards)**

Pattern-Mixture Models

- **Full data likelihood is the product of the measurement process conditional on the drop-out pattern and the marginal density of the missingness process**
- **Conceptually, pattern-mixture models (typically) assess the outcome variable separately for different groups (patterns), often defined by time of drop-out, and then combine results across groups for final inference**

Shared-Parameter Models

- **Similar to selection models in that they jointly model the measurement and drop-out processes**
- **Assumes that a parameter, typically a random effect, influences both the outcome variable and drop-out, such that conditional upon this parameter, the measurement and drop-out processes are independent**

Considerations for MNAR as Primary

Can do better than MAR only by making assumptions

- **Can't test assumptions as don't have missing data about which assumptions are made**
- **Also true we can't test validity of MAR, but consequences of violating assumptions and model misspecification more severe in MNAR**
- **No single MNAR analysis is definitive**
- **Standard software lacking, complex, can be difficult to converge**

Outline

- Introduction
- Theory and concepts
- Regulatory concerns
- Handling MNAR data
- Recommendations

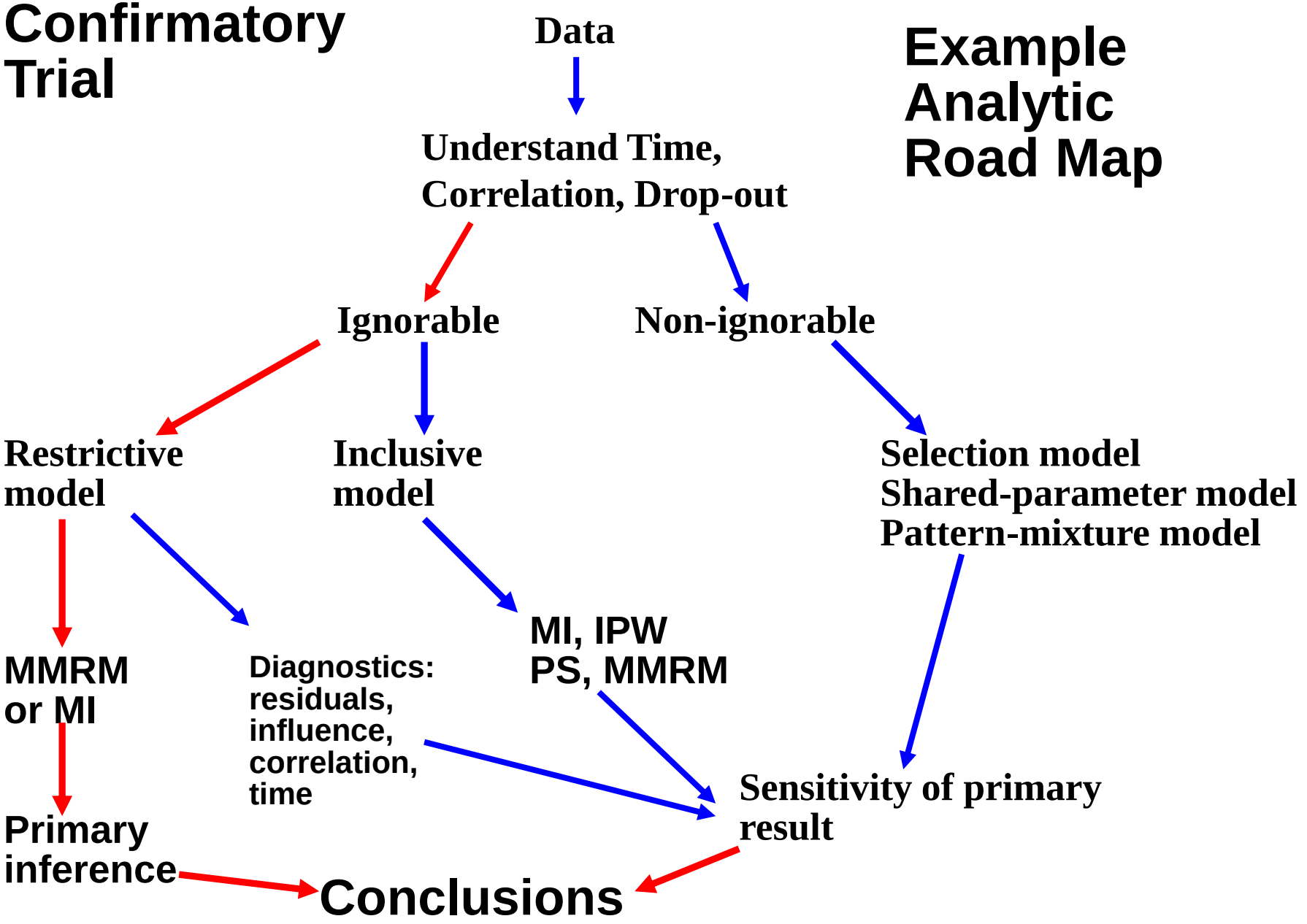
Recommended Analytic Approach

- **MAR (MMRM) with restrictive modeling as primary**
- **Use MAR with inclusive modeling and MNAR methods as sensitivity analyses**
- **Use local influence to investigate impact of influential patients**

Pharmaceutical Statistics. 4: 267–285.
J. Biopharm. Stat. 16: 365-384.

Confirmatory Trial

Example Analytic Road Map



Example Trial in Depression

	Drug	Placebo
Protocol complete	60.9%	64.7%
Adverse event	12.5%	4.3%
Lack of efficacy	5.5%	13.7%

Differential rates, timing, and/or reasons for drop-out do not distinguish between MCAR, MAR, MNAR

Results from Example Trial

- **MAR with restrictive modeling as primary**
  - **MMRM: Δ = 2.17, p = .024**
- **MAR with inclusive modeling**
  - **MI including a variety of approaches for using AE data in imputation: Δ = 2.1 – 2.3**

Results from Example Trial

- **MNAR Selection model**
  - **Δ = 2.2, p = .018**
  - **No evidence of MNAR**
- **Local influence identified 5 influential patients**
  - **Excluding influential drug-treated or placebo-treated patients did not alter significance of treatment contrast or evidence for MNAR**

Implications of Sensitivity Results

- **Comforting that no subjects had a huge influence on results. (Impact bigger if it were a smaller trial)**
- **Similar to other depression trials we have investigated, results not influenced by MNAR data**
- **We can be confident in the primary result**

Discussion

No universally best method for analyzing longitudinal data

- **Analysis must be tailored to the specific situation at hand**
- **MMRM well-suited for use as the primary analysis in confirmatory trials**
- **MNAR can never be ruled out. Sensitivity analyses and efforts to lower rates of drop-out are essential**
- **LOCF (and BOCF) are not suitable choices**