Methodological Considerations in Comparative Effectiveness Research

Biopharmaceutical Applied Statistics Symposium (BASS) XVII

November 9, 2010 – Hilton Head, SC

Carlos Alatorre, David Nelson, Anthony Zagar

Eli Lilly and Company

Outline

Focus: approaches to handling missing data in pragmatic CER studies

1. Comparative Effectiveness Research (CER)
2. Pragmatic CER studies
3. Methodological challenges
4. Approaches to missing data, simulations, and results
5. Conclusions

Lilly

Research motivation

- Current environmental expectations of clinical research reflect the need for greater understanding of how products are used and how they work in routine clinical practice
- Randomized Controlled Trials (RCTs) are the gold standard in clinical research, but are costly and their tightly controlled conditions limit their ability to generalize findings to routine patient care
- Some stakeholders more interested in the therapeutic application and sustainability of outcomes in real-world medical practice with intended post-launch populations
  - Intent is to inform patient, provider, payer and policy decision-making: Does it work? vs. Can it work?
- Pragmatic studies are a type of comparative effectiveness research that intend to reflect real-world evidence
- Missing data in the presence of treatment switching is a common issue in pragmatic studies and can lead to invalid results

Lilly

Part I – Comparative Effectiveness Research

The evidence predicament

- Approximately 18,000 RCTs published each year[1]
  - But available evidence for clinical and payer decision-making is inadequate with limited or poor quality
- Evidence producers
  - National Institutes of health (NIH): discovery and proof of concept focus
  - Industry: Food and Drug Administration (FDA) and market focus
  - Agency for Healthcare Research and Quality (AHRQ): modest budget, broad portfolio*
  - Health Technology Assessment (HTA) bodies: reviews
- Decision makers have had no significant influence in what evidence is created
- Comparative Effectiveness Research can help

* Recently, the American Recovery and Reinvestment Act gave AHRQ approximately $400 M as well as the authority to determine where the entire $1.1 B would be used -- so more recently they have had a more reasonable budget for CER

Lilly

Comparative Effectiveness Research

- CER is the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care.[2]
  - *Generation*: includes **new** original research where evidence is generated (e.g., prospective and retrospective observational studies, pragmatic trials)
  - *Synthesis*: includes systematic reviews of **existing** evidence
- The purpose of CER is to assist consumers, clinicians, payers, and policy makers to make informed decisions that will improve health care at both the individual and population levels.[2]
- Focus is on populations and conditions representative of real-world clinical practice to assess outcomes that are directly relevant to clinical and policy decisions.

Lilly

Comparative Effectiveness Research

Other definitions offer considerable overlap

Organization	Definition
American College of Physicians	Evaluation of the relative (clinical) effectiveness, safety and cost of 2 or more medical services, drugs, devices, therapies, or procedures used to treat the same condition.[3]
Institute of Medicine (IOM) – Roundtable on Evidence-Based Medicine	Comparison of one diagnostic or treatment option to ≥1 others. **Primary** CER involves the direct generation of clinical info on the relative merits or outcomes of one intervention in comparison to ≥1 others. **Secondary** CER involves the synthesis of primary studies to allow conclusions to be drawn.[4]
Agency for Healthcare Research and Quality (AHRQ)	A type of health care research that compares results of one approach for managing a disease to results of other approaches. CER usually compares ≥2 types of treatment, such as different drugs, for the same disease but it can also compare medical procedures and tests. The results can be summarized in a systematic review.[5]
Medicare Payment Advisory Commission (MedPAC)	Evaluation of the relative value of drugs, devices, diagnostic and surgical procedures, diagnostic tests, and medical services. By value, it is meant the clinical effectiveness of a service compared with its alternatives.[6]
Congressional Budget Office (CBO)	A rigorous evaluation of the impact of different options that are available for treating a given medical condition for a particular set of patients. Such research may compare similar treatments, such as competing drugs, or it may analyze very different approaches, such as surgery and drug therapy.[7]
Center for Medical Technology Policy (CMTP)	The direct comparison of existing health care interventions to determine which work best for which patients and which pose the greatest benefits and harms. The core question of comparative effectiveness research is which treatment works best, for whom, and under what circumstances.[8]

Lilly

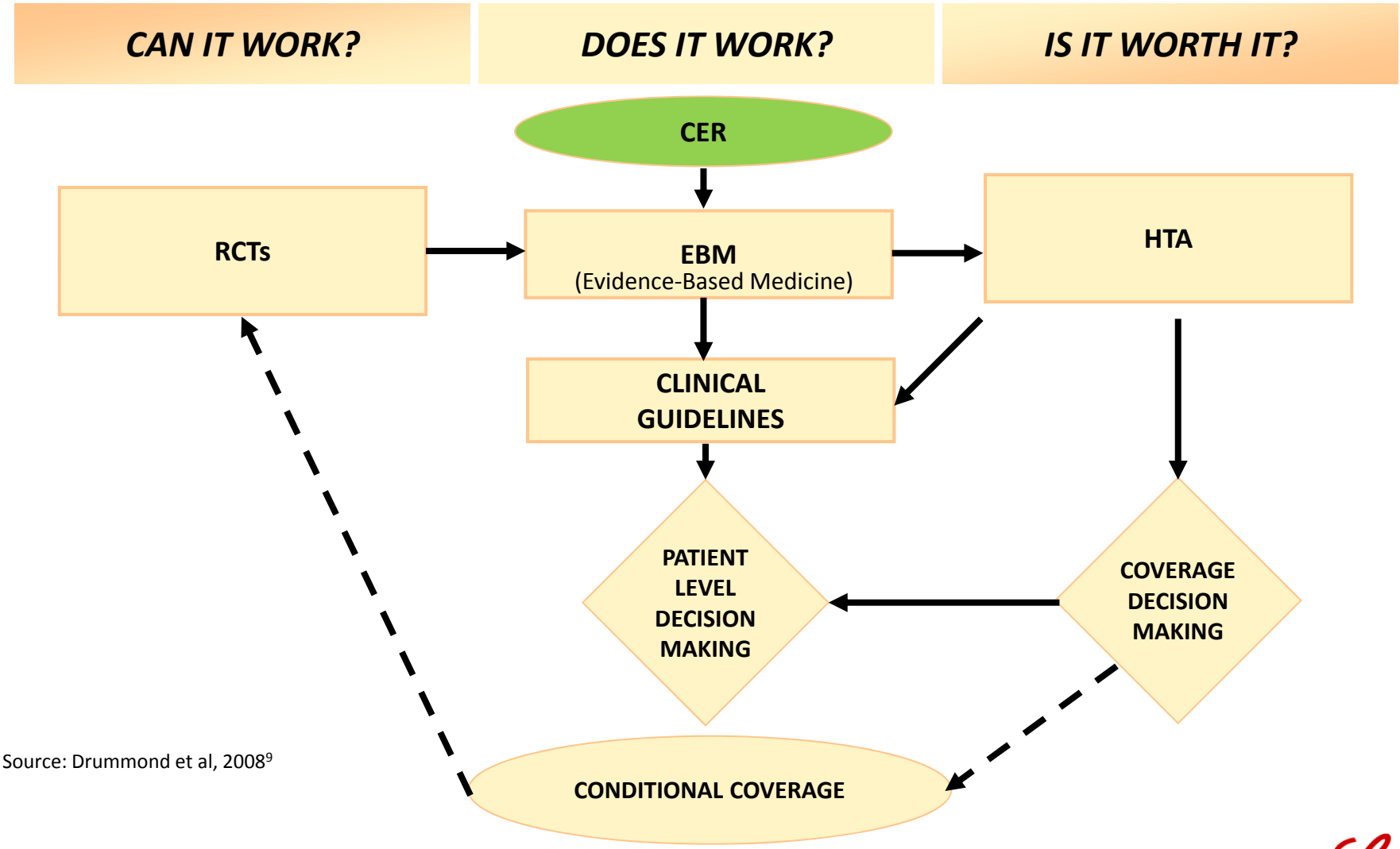
Comparative Effectiveness Research

Distinguishing characteristics[2]

1. Informs a specific clinical decision from the individual patient perspective or a health policy decision from the population perspective
2. Compares at least two alternative interventions, each with the potential to be "best practice"
3. Describes results at the population and subgroup levels
4. Measures outcomes—both benefits and harms—that are important to patients and decision makers
5. Employs methods and data sources appropriate for the decision of interest (e.g., observational studies, pragmatic trials, systematic reviews)
6. Conducted in settings that are similar to those in which the intervention will be used in practice

Lilly

CER plays a key role in generating evidence for clinical and policy decision-making



Source: Drummond et al, 2008[9]

The new PCORI institute will further CER efforts

- In 2009, a $1.1 Billion in funding for CER was part of the American Recovery and Reinvestment Act (ARRA), more commonly known as the economic stimulus package.[10] This one-time funding was distributed among AHRQ, NIH, and HHS, and included the establishment of a Federal Coordinating Council for CER (FCCCER) and funding for a new Institute of Medicine (IOM) effort to develop a prioritized listing of CER research topics.
- A non-profit Patient-Centered Outcomes Research Institute (PCORI) was included in the health care reform law enacted in March 2010, the Patient Protection and Affordable Care Act (PPACA)[11], with the goal of promoting and funding CER.
- PCORI creates an enormous opportunity for the transformation of health care to a system that is based on the systematic creation and use of CER evidence in the delivery of care.[12]

Lilly

Part II – Pragmatic CER studies

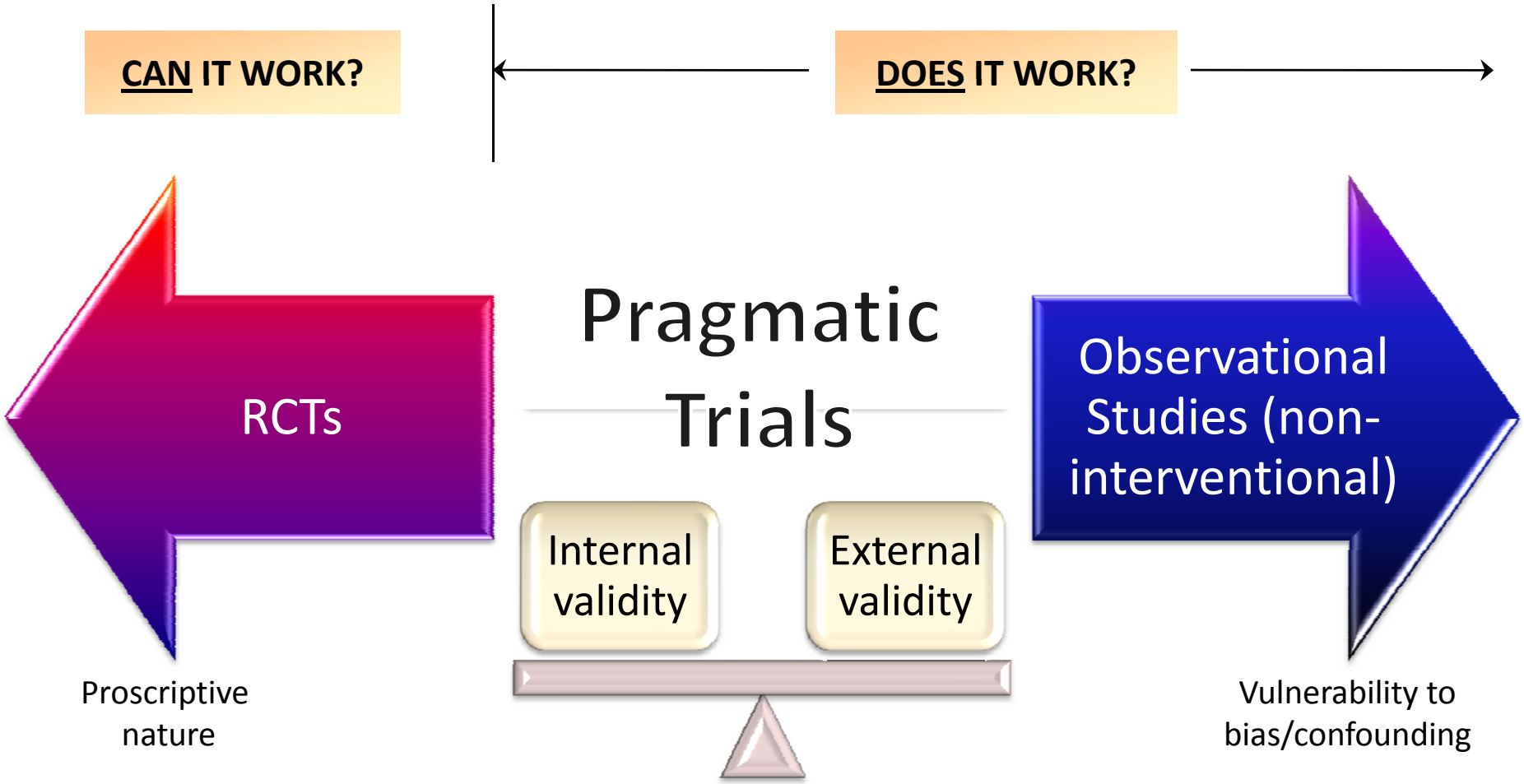
Pragmatic CER Studies

- A type of original CER (i.e., generation of new evidence)
- Goal: to answer relevant questions faced by a broad range of decision-makers (patients, providers, payers) as they make healthcare decisions
  - e.g., demonstrate effectiveness among broad patients who will receive treatment post launch
- Definition: prospective research with minimized protocol-driven care
  - Randomized
  - Reflective of routine care (naturalistic) – relevant comparators, conditions, and outcomes
  - Inclusion of diverse populations from heterogeneous practice settings
- Can take place in phase 3b* (see relevant CMTP methodological guidance)[13] or phase 4

* Based on consultation and agreement with regulatory agencies.

Pragmatic CER Studies

*A balancing act… in shades of gray**



*Tradeoff between internal validity and feasibility, generalizability, cost, time

Pragmatic CER Studies

Key strengths and limitations

» Typical Strengths

- Randomization (minimizes bias)
- Generalizability (reflects heterogeneity of patients, providers, and conditions of care)
- Reduced patient and investigator burden

» Typical Limitations

- Bias (unblinded), time-varying confounding (BP in CV outcomes, adherence), missing data
- Safety data collection intensity and unknown practice pattern in phase 3b
- A wider range of severity and co-morbidities among the diverse set of patients can result in more statistical noise in the data.[14]

Lilly

Pragmatic CER Studies

Types of questions

- Policy: real-world assessment of consequences of treatment initiation
  - Question: What are the real-world consequences (benefits, risks, economic impact, etc.) of the decision to start patients on Tx A vs. Tx B?
  - Method: Intent-to-treat (assign all outcomes to the Tx to which the patient was first randomized)

- Effectiveness: real-world assessment of individual therapy
  - Question: What is the comparative effectiveness of Tx A vs. Tx B?
  - Method: Marginal Structural Model to account for switching (tease out the role of individual treatments in producing the outcomes observed irrespective of initial Tx assignment)

Lilly

Part III – Methodological challenges

Lilly

Methodological challenges

Without randomization – due to selection bias, standard methods produce only 'associations' and not 'causal effects' …… unless selection bias is controlled

With randomization – standard methods can produce estimation of causal treatment effects

HOWEVER:

even with randomization (e.g., pragmatic study), time-varying confounders can influence non-compliance and medication switching, which standard methods cannot address

Lilly

Methodological challenges (continued)

- Bias (avoid)
  - Systematic error that results in incorrect estimates; affects accuracy and validity
  - Methods: Randomization, propensity score can help
- Confounding (control)
  - Third factor associated with both exposure and outcome; affects causality
  - Methods: Randomization (baseline confounding) and Marginal Structural Models (time-varying confounding) can help
- Missing Data (address)
  - Missingness due to drop-outs (monotone) or non-response (non-monotone); creates bias
  - Methods: Limited literature; Marginal Structural Models (MSM), Mixed Models Repeated Measures (MMRM), and Multiple Imputation (MI) could help under certain conditions

Lilly

Our research focus

- Pragmatic CER studies can be susceptible to issues of confounding, bias, and missing data. Even with randomization, time-varying confounders can affect compliance and switching which standard stat methods cannot address.
- Under some assumptions, Marginal Structural Models (MSM) can provide an approach for assessing causal effects of treatments in an unbiased fashion when patients switch treatments and time-dependent confounders exist.[15,16, 17, 18] MSMs can also adjust for bias due to dropouts (monotone missingness).
- Given that it is not uncommon to encounter pragmatic CER data with monotone and non-monotone missingness patterns when treatment switching is also present, **our research** intended to compare various methodological approaches to handling missing data using simulated sets of pragmatic CER data with varying conditions of missingness.

Lilly

Our research focus (continued)

- Missing data in the exposure of interest or covariates are common in longitudinal studies. Standard stat analyses of observational data often exclude information from individuals with incomplete data. This leads to biased estimates of treatment effect and loss of precision.
- There has been little work assessing the impact of missing data in MSM models and on comparing results from various missing data techniques. Moodie et al (2008) proposed a method for dealing with missingness in MSMs by weighting subjects by the inverse probability of missingness and found that Multiple Imputation (MI) performed better than their method in MSM models.[19] Shortreed and Forbes (2010) investigated the impact of missing data in a complex, realistic setting using a standard application of MSMs and found that MSM under monotone missingness performed reasonably well except when the data were incomplete according to the Missing Not At Random (MNAR) mechanism.[20]
- **Our research** compared various missing-data methods using data with different combinations of sample sizes, missing mechanisms, and levels of missingness.

Lilly

Part IV – Comparing approaches to missing data

Simulation Study of Various Methods to Handle Missing Values and Treatment Switching

- Assume there is a pragmatic study in which patients are initially randomized to one of two treatments
- At Visit 1, patients are assigned to treatment and baseline (BL) information is recorded, including the BL value used to calculate the study's primary endpoint, Δ AVAR (change from baseline)
- At Visits 2 through 4, AVAR and the covariates (X1-X4) are re-measured, and whether the patient is on the same treatment or switched to the other treatment is recorded

Lilly

Information Recorded, Schedule By Visit:

	Severity Measure, AVAR	Variable of Interest, Treatment	Covariates, X1-X4
Visit 1 (Baseline)	X	X	X
Visit 2	X	X	X
Visit 3	X	X	X
Visit 4	X	X	X

Information for Analysis :

	Dependent Variable, Δ AVAR (from BL)	Effect of Interest, Previous Visit TRT	Previous Visit X1-X4	Baseline Treatment	Baseline AVAR	Baseline X1-X4
Visit 2	X	X	X	X	X	X
Visit 3	X	X	X	X	X	X
Visit 4	X	X	X	X	X	X

Reality: Patients are Missing Visits

	Dependent Variable, Δ AVAR (from BL)	Effect of Interest, Previous Visit TRT	Previous Visit X1-X4	Baseline Treatment	Baseline AVAR	Baseline X1-X4
Visit 2	X	X	X	X	X	X
Visit 3	X	X	X	X	X	X
Missing Visit 4	?	X	X	X	X	X

	Dependent Variable, Δ AVAR (from BL)	Effect of Interest, Previous Visit TRT	Previous Visit X1-X4	Baseline Treatment	Baseline AVAR	Baseline X1-X4
Visit 2	X	X	X	X	X	X
Missing Visit 3	?	X	X	X	X	X
Visit 4	X	?	?	X	X	X

Our Approach To Compare Analytical Methods in Pragmatic Trials

- Use a simulated dataset, designed to mimic the features of a pragmatic trial of neuroscience compounds
- The dataset has similar features as one used to illustrate MSM in the recently published *Analysis of Observational Health Care Data Using SAS* (Chapter 9)[18]
- Our approach was to use this dataset in a resampling study, while varying both the type of missingness and the amount of missing visits
- Compare Bias and Empirical Mean Squared Error (to assess both bias and precision of the estimate) of several analytical methods

Lilly

Visits Can Be Missing in Three Ways

- There may be no particular reason why some patients miss their visits and others didn't. That is, the probability that *Y* is missing may has no relationship to *X* or *Y,* whether or not they are observed. Such data said to be *missing completely at random* (MCAR)
- Perhaps patients with a certain characteristic, like being *"yes"* for covariate X1, is more likely to miss a visit. That is, the probability that *Y* is missing depends only on the value of *X*. Such data are *missing at random* (MAR)
- Perhaps missing visits are related to the severity measure. That is, the probability that *Y* is missing depend on the unobserved value of *Y* itself. Such data are *not missing at random* (NMAR)
- We examined all three patterns of missingness

Lilly

Other Characteristics of the Analyses

- We also varied whether missing visits occurred at rates of 10%, 30% or 50%
- Missing visits that were generated were a combination of monotonic (missing the remainder of visits once one is missed) and non-monotonic (missing a visit and returning) patterns
- We performed analyses on studies of 100 patients and 1000 patients

Lilly

Dealing with the Missing Values

- Several methods were applied (Baseline Observation Carried Forward [BOCF], Last Observation Carried Forward [LOCF], Completer's Analysis, Multiple Imputation)
- Multiple Imputation was performed using Markov Chain Monte Carlo (MCMC) methods
- All missing variables necessary for the analysis were imputed: treatment, dependent variables, and covariates
- Imputations were performed five times
- The results combined into a single estimate of treatment effect

Lilly

Analytical Methods - MMRM

- Mixed Model Repeated Measures (MMRM)
- MMRM with Intent to Treat (ITT) for the randomized classes (so ignore switching)
- MMRM with time dependent covariates, including therapy switching
- MMRM was performed with and without MI
- Analysis of subgroups: Study completers (no missing visits) and those on their original treatment (visits after switching were excluded)
- We utilized the autoregressive (AR1) correlation structure for the repeated measures

Lilly

Marginal Structural Models

- MSM uses Inverse Probability of Treatment Weighting (IPTW)
- IPTW calculates probability of an individual receiving their *treatment* conditional on their observed covariates
- IPTW also calculated for *missing* their next visit
- The “treatment selection” and “censoring/missing” weights are then multiplied to obtain a general weight
- Each subject is weighted by the inverse of that probability
- Designed for monotonic missing patterns

Lilly

Traditional Methods

- Last Observation Carried Forward (LOCF) and Baseline Observation Carried Forward (BOCF)
- Difference Between Groups Estimated at last visit using ANCOVA adjusting for baseline AVAR
- All models were compared using differences between the Least Squares Means between the two treatments overall and at the last visit (results presented herein are “last visit” results)

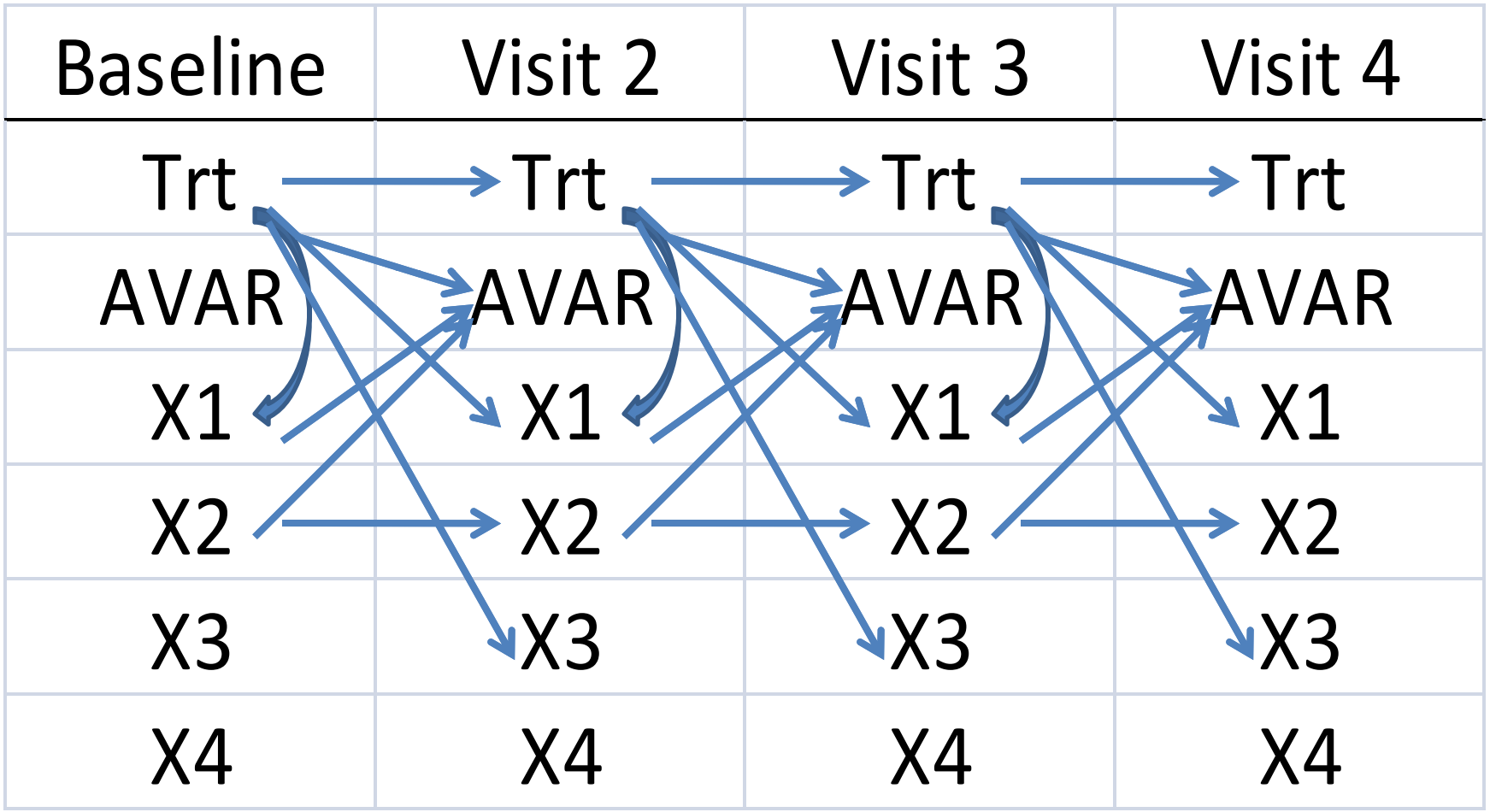
Lilly

Characteristics of the Simulated Dataset

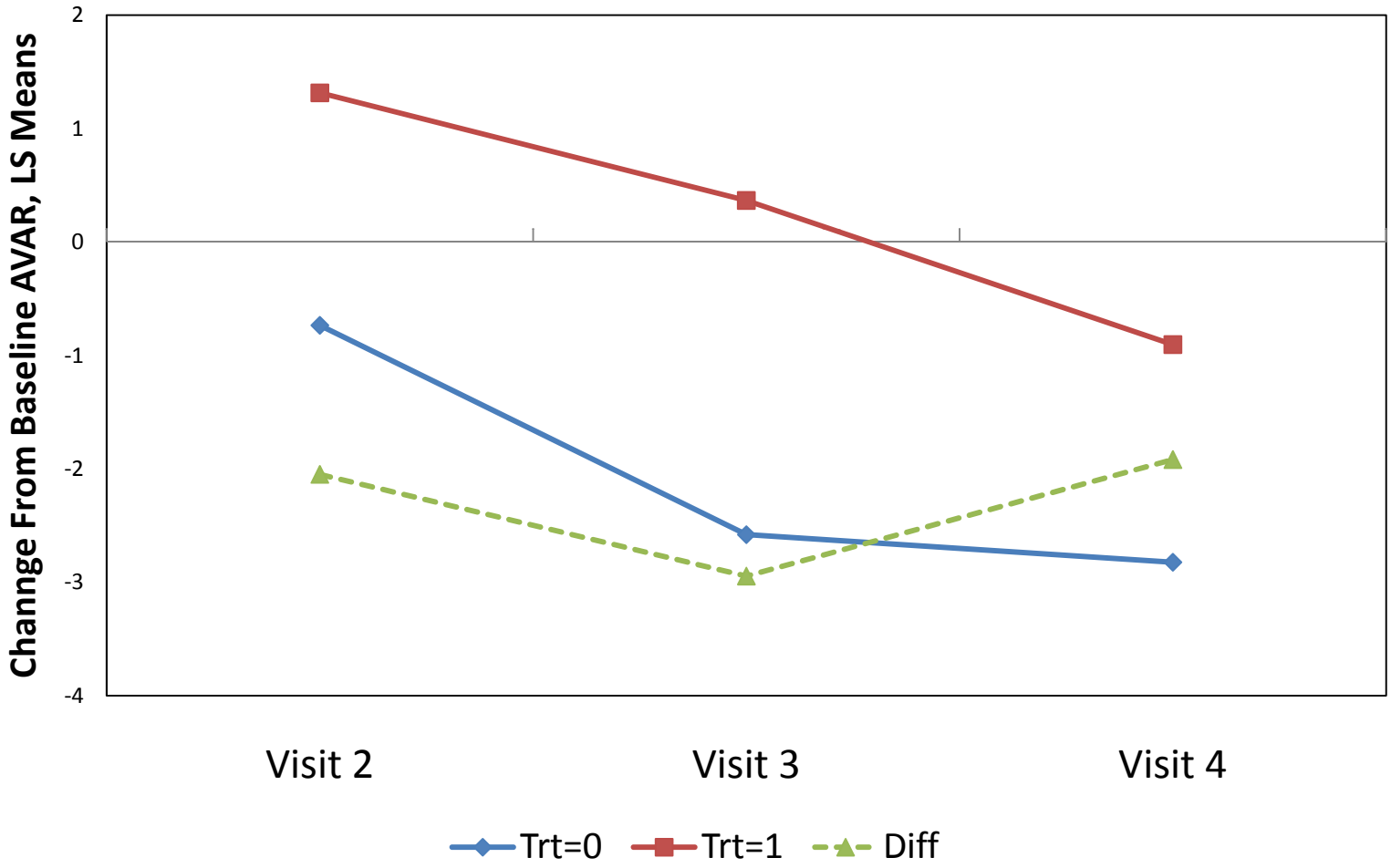
- Baseline treatment affects the probability of receiving the same treatment at visit 2, and also affects binary covariates X1 and X3 at visit 2
- Baseline treatment is not related to continuous covariates X2 and X4 at visit 2
- The Visit 2 AVAR is dependent upon baseline X1 and X3 values
- Same pattern persists for Visits 3 and 4

Lilly

Relationships Among Variables



“Gold Standard” Relationship Between Previous Visit Treatment and Change in AVAR



Comparing Methods

- We will focus on comparisons of LS Means at the last visit
- With **complete** data, time-dependent MMRM model would fit this generated data perfectly, but
  - Under what conditions of missingness will MMRM lose accuracy and precision to estimate the treatment effect?
  - Will Multiple Imputation, using complete cases only, or those that do not switch treatments improve MMRM?
  - Will MSM models improve performance when MMRM fails?

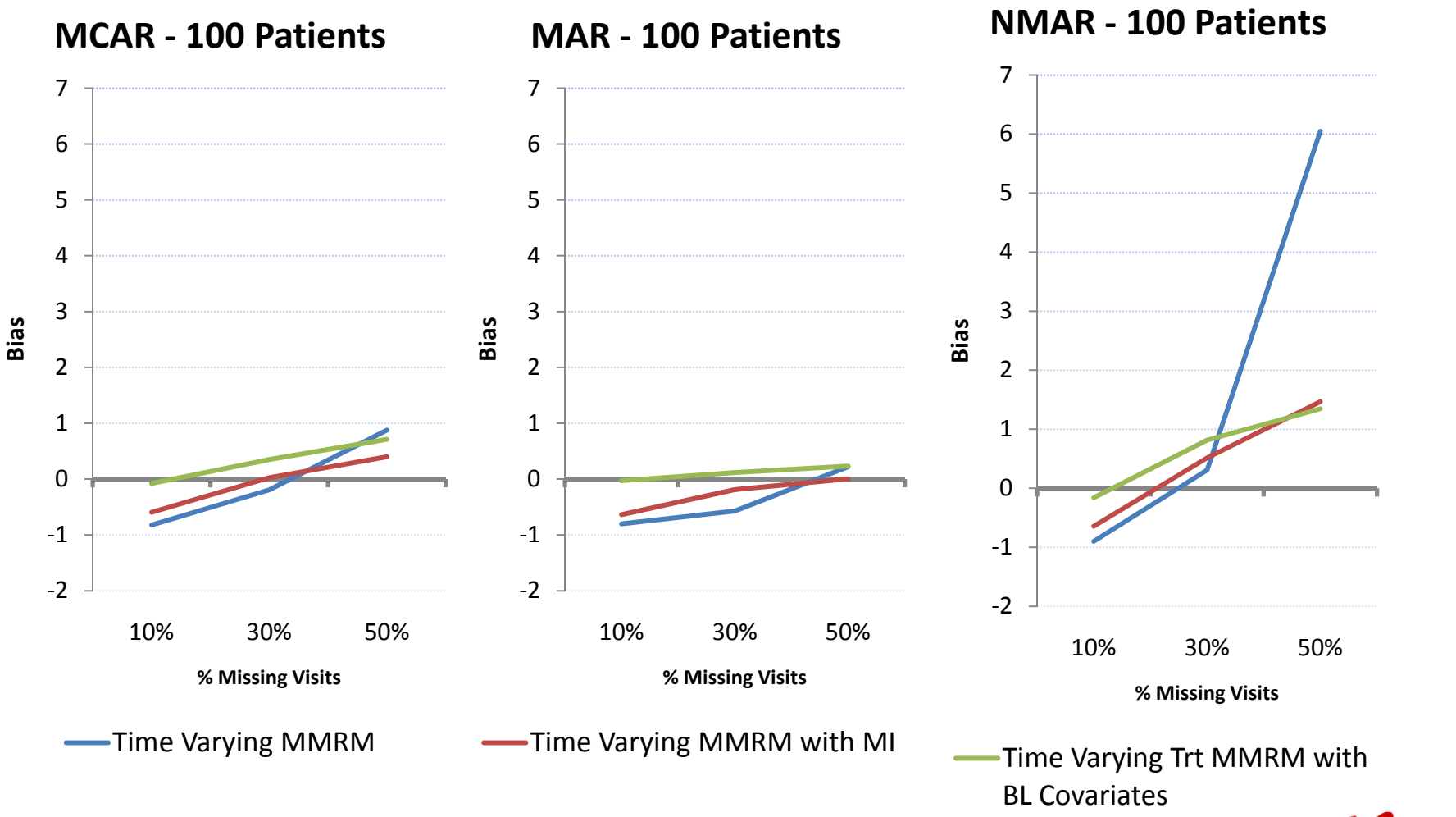
Lilly

Bias and MSE Evaluated After Varying:

Type of Missingness	Percentage of Missing Visits	Missingness Pattern	Sample Size	Methods
MCAR MAR NMAR	10% 30% 50%	Data contains both Monotonic and Non-Monotonic represented	100 1000	Time Dependent MMRM MSM +/- MI TD Treatment or ITT Completers Original Trt BOCF/LOCF

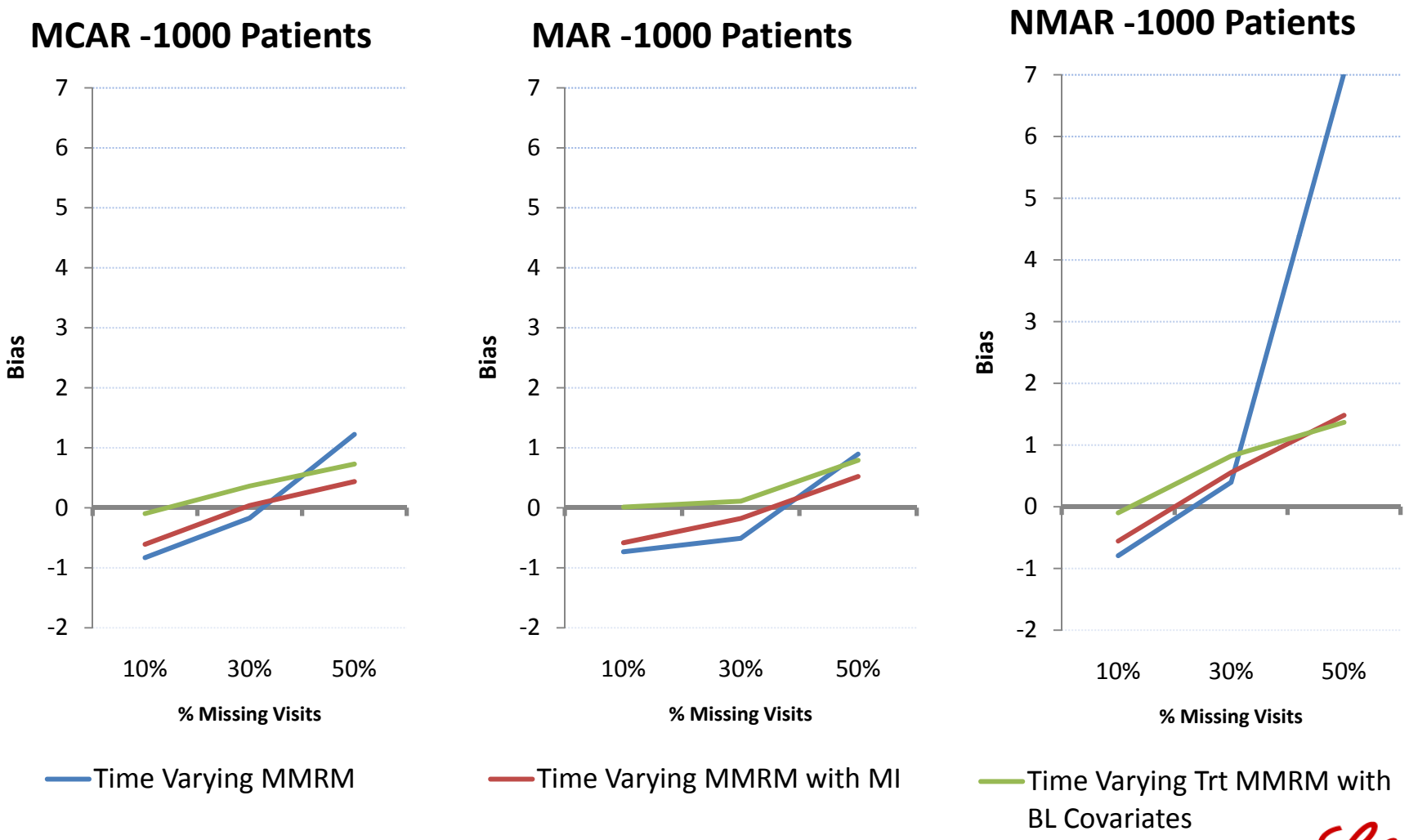
MRMM +/- MI: Bias/Small Study:

NMAR with 50% missing increases bias



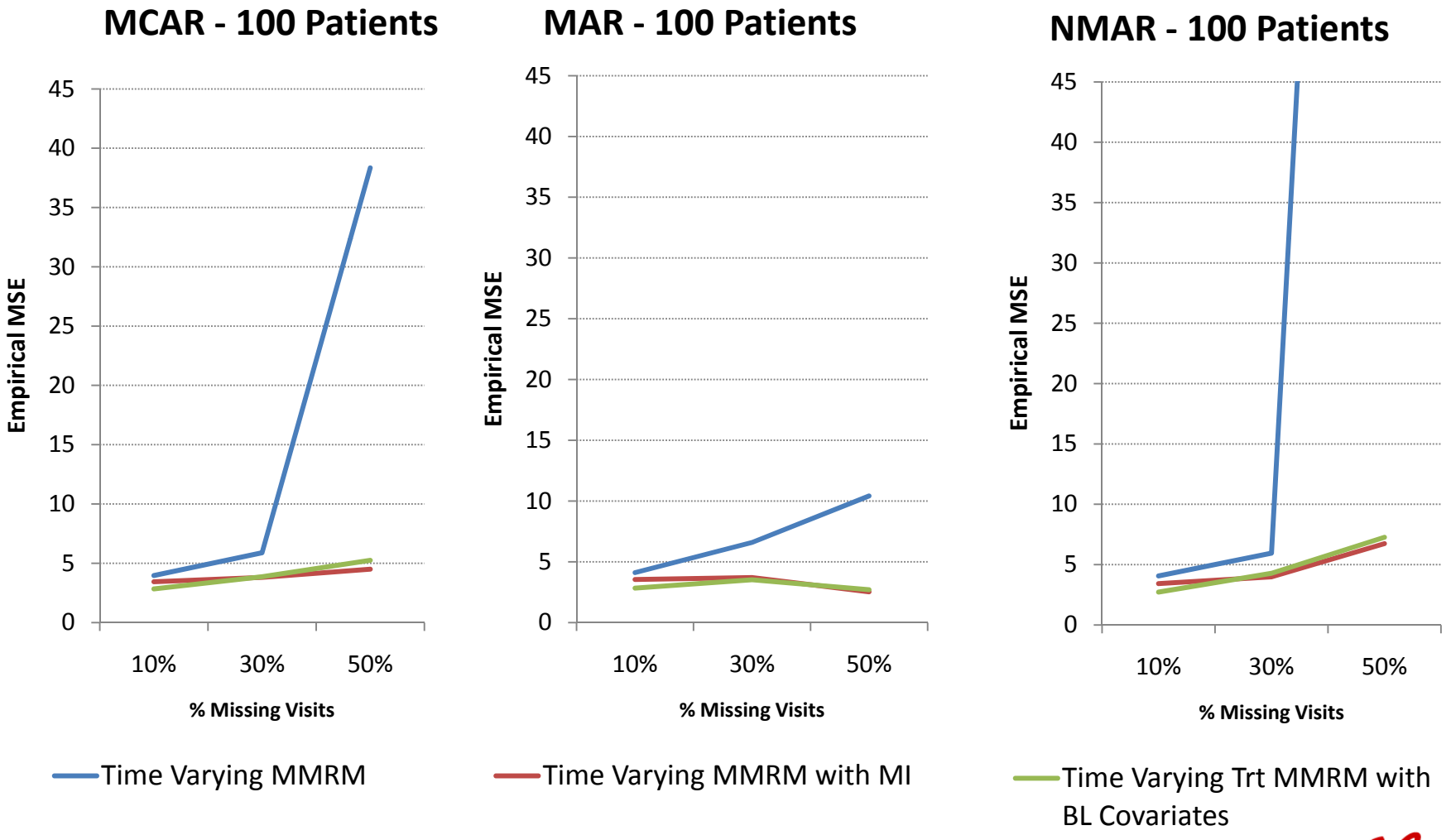
MRMM +/- MI: Bias/Large Study:

Even with larger sample size, MMRM without MI still has increased bias with NMAR missing visits of 50%



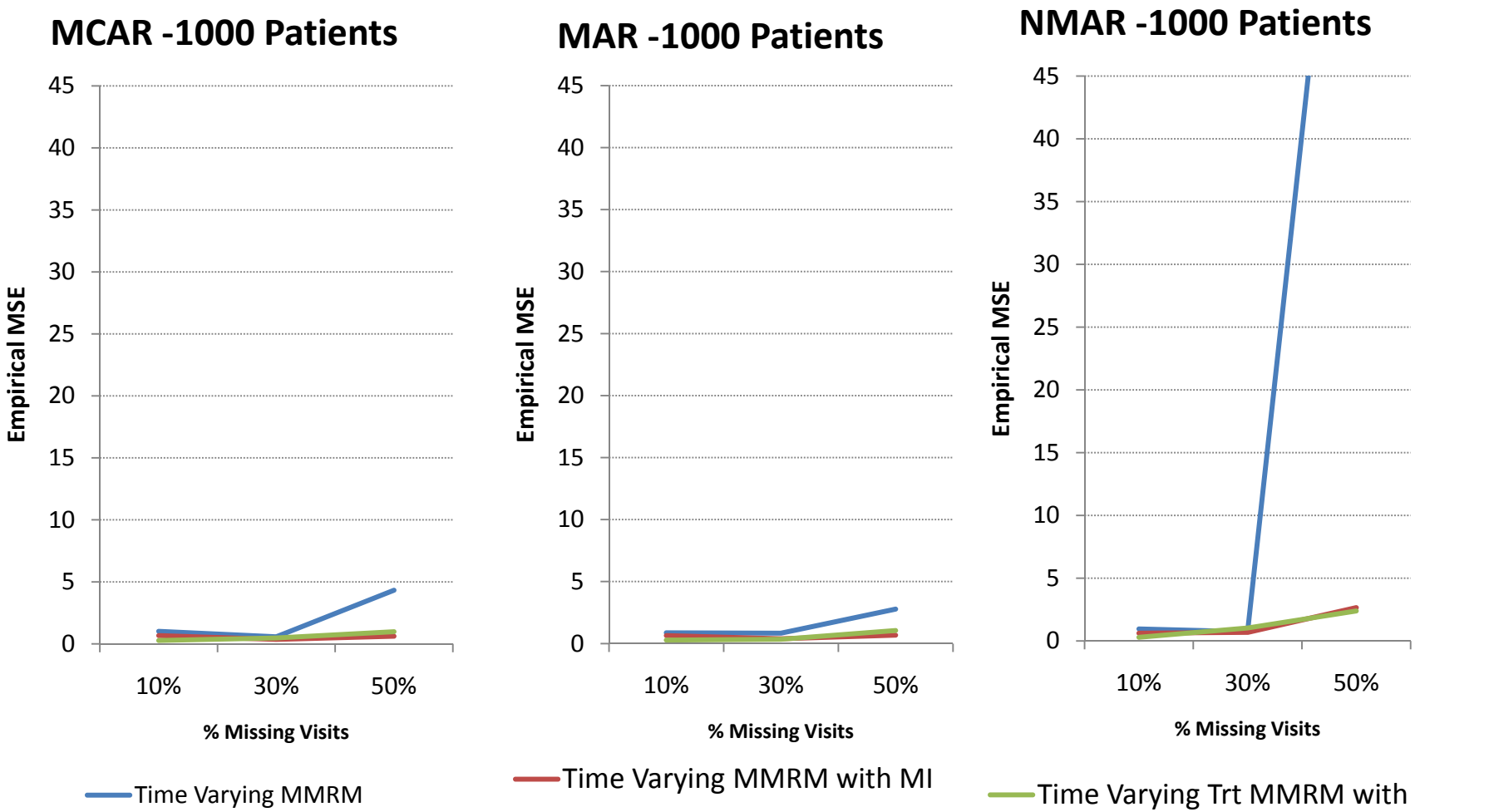
MRMM +/- MI: MSE/Small Study:

MMRM without imputation increases MSE when 50% of visits are missing



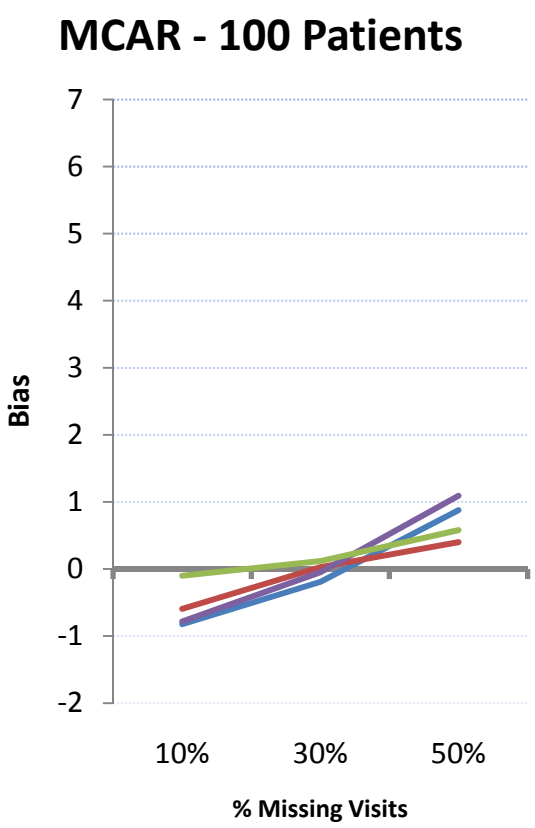
MRMM +/- MI: MSE/Large Study:

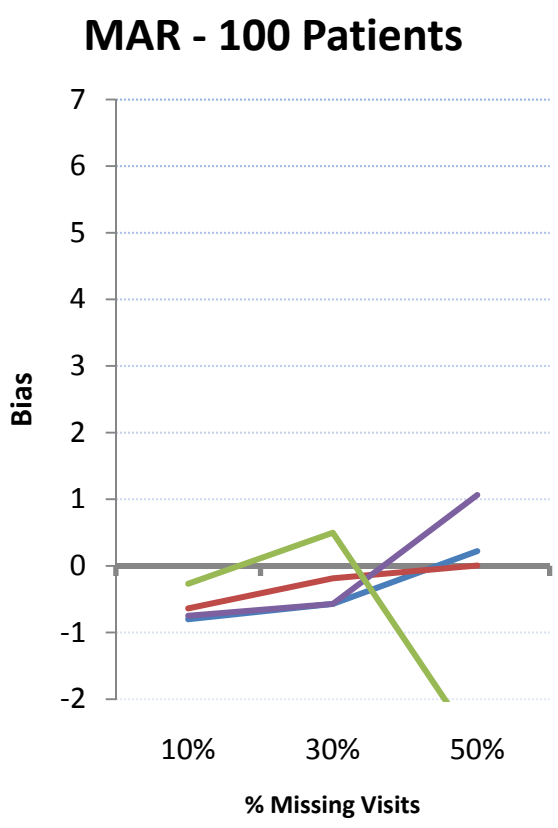
Larger sample sizes reduce MSE except for NMAR at 50% missing

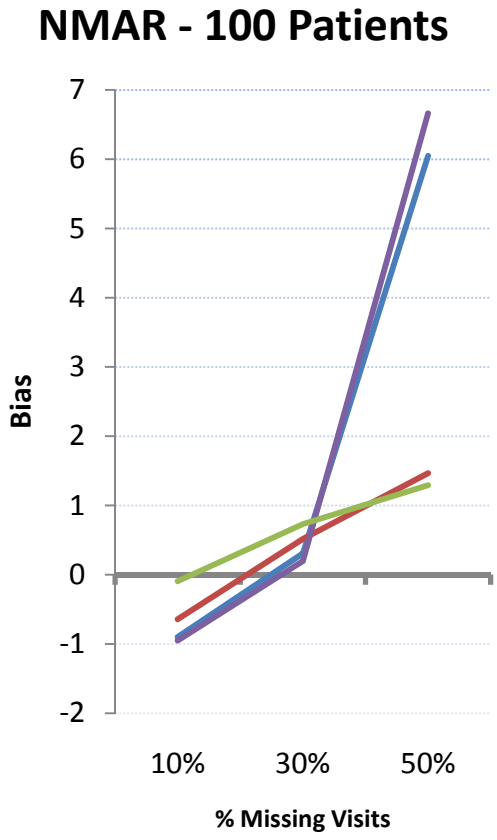


Completers and Original Treatment: Bias/Small Study:

These subgroup alternatives increased bias at 50% missing







Time Varying MMRM

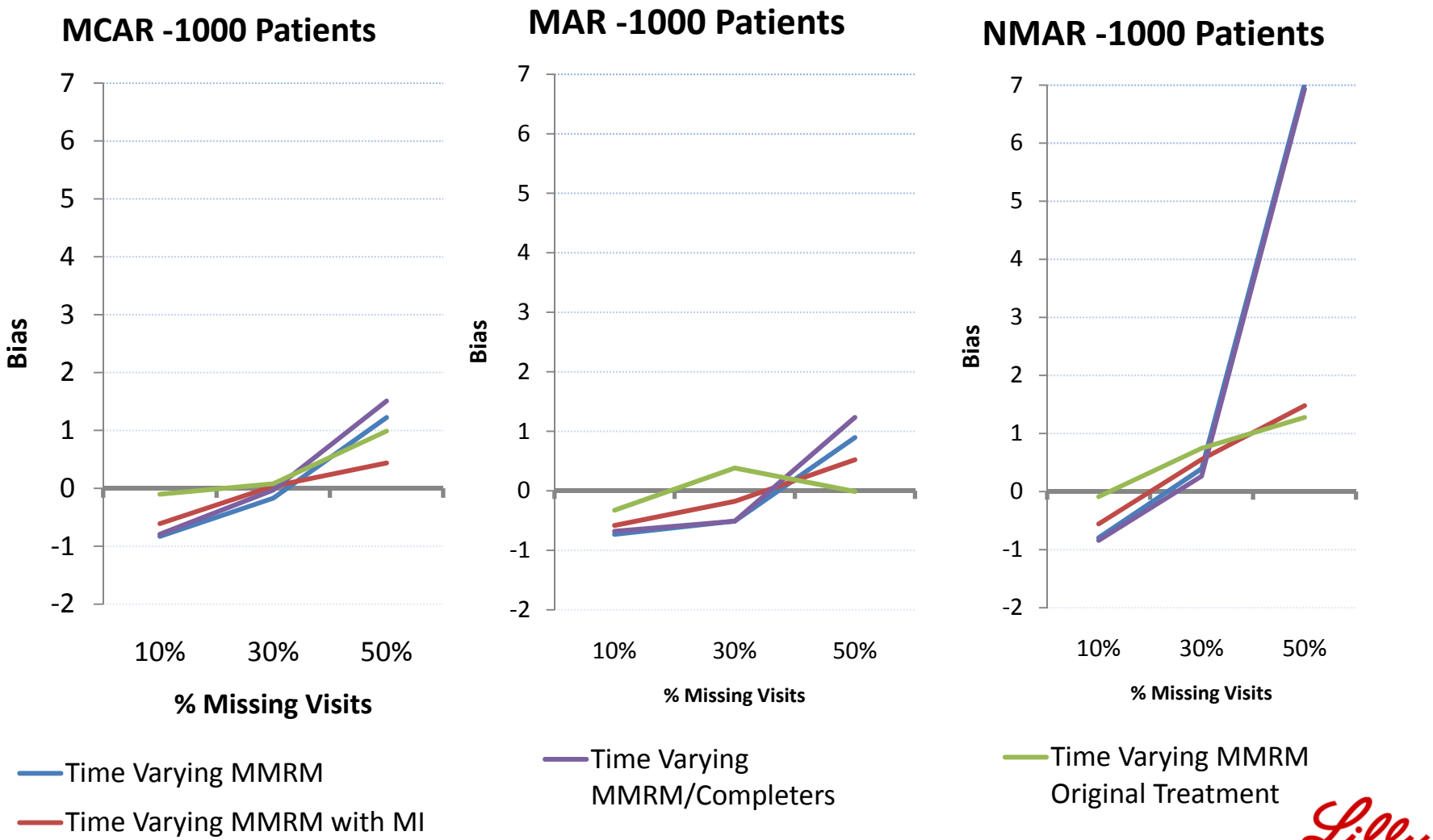
Time Varying MMRM with MI

Time Varying MMRM/Completers

Time Varying MMRM Original Treatment

Completers and Original Treatment: Bias/Large Study:

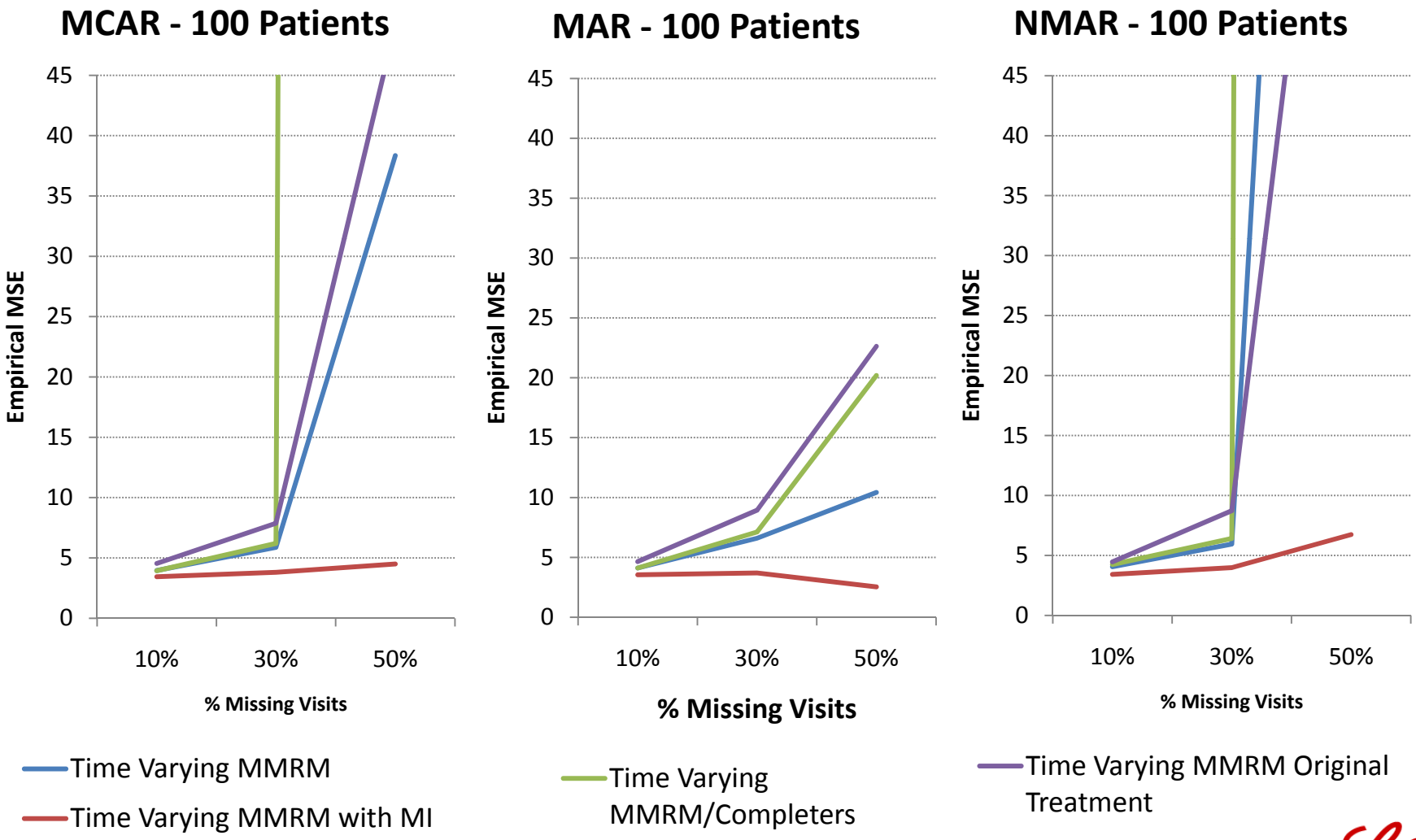
Larger sample sizes alleviate bias at 30%, but Completers increase bias still at NMAR missing visits of 50%



Lilly

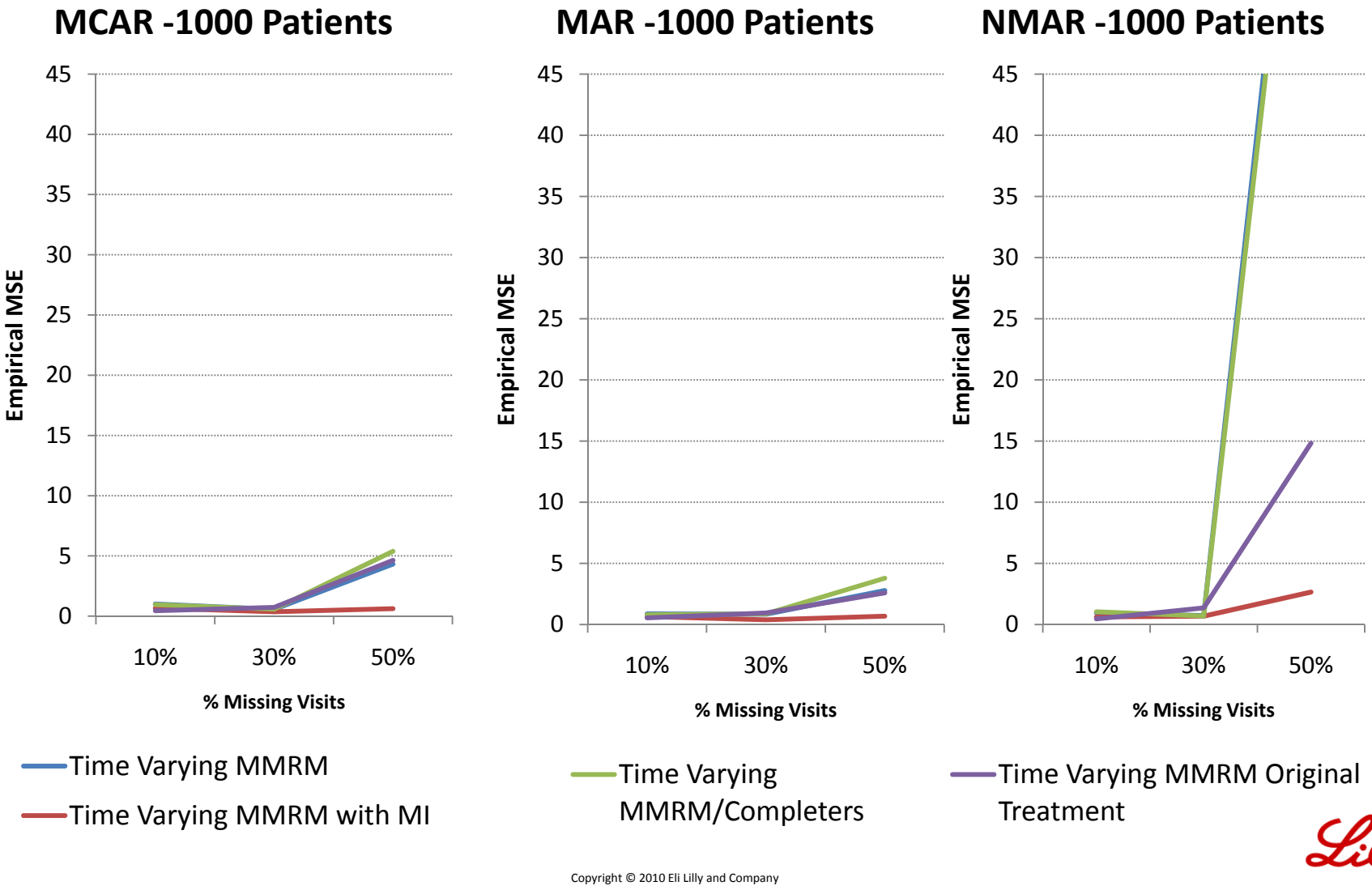
Completers and Original Treatment: MSE/Small Study:

Using subgroups increases MSE when missingness is 30%



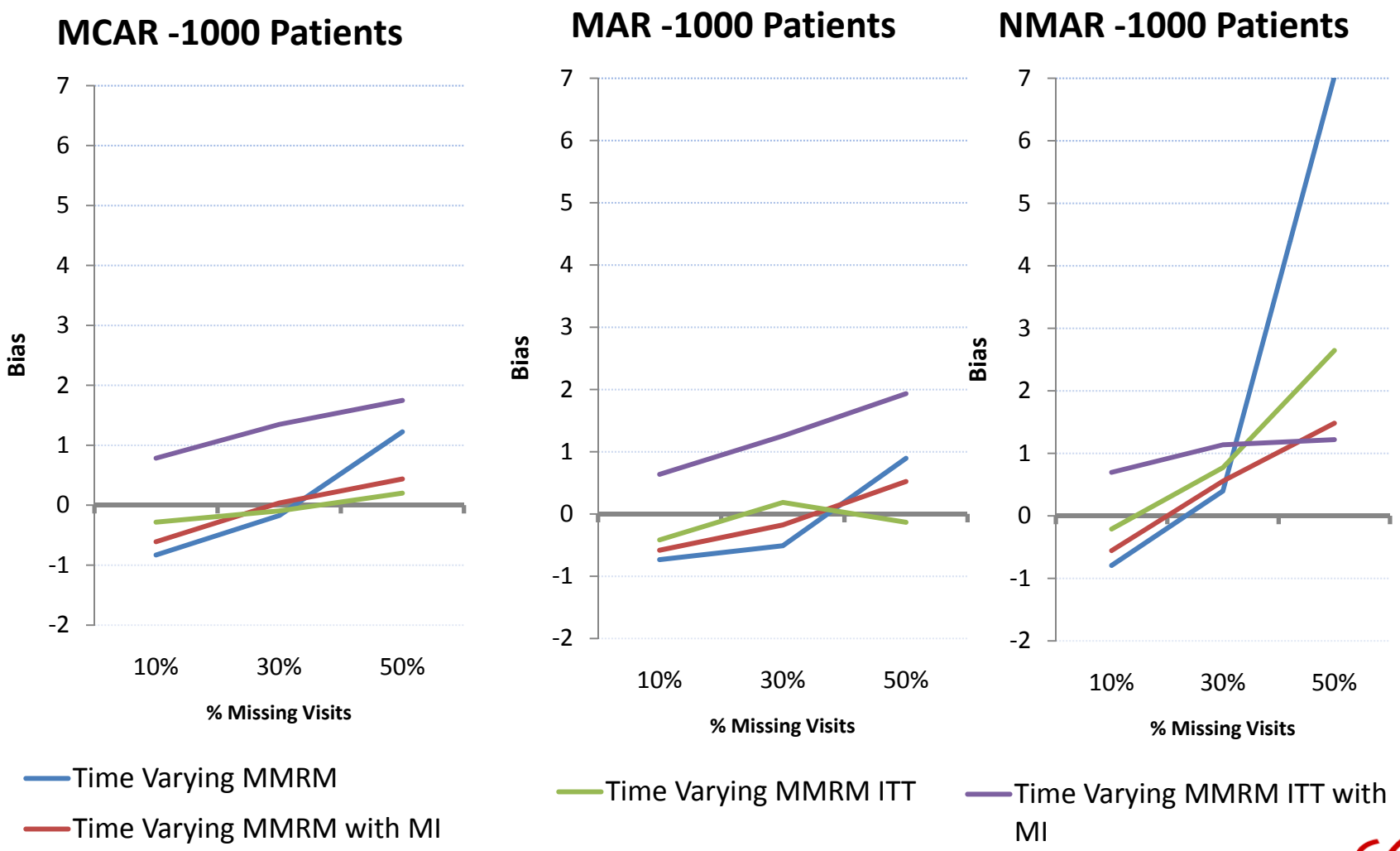
Completers and Original Treatment: MSE/Large Study

Larger sample sizes reduce MSE except for NMAR at 50% missing



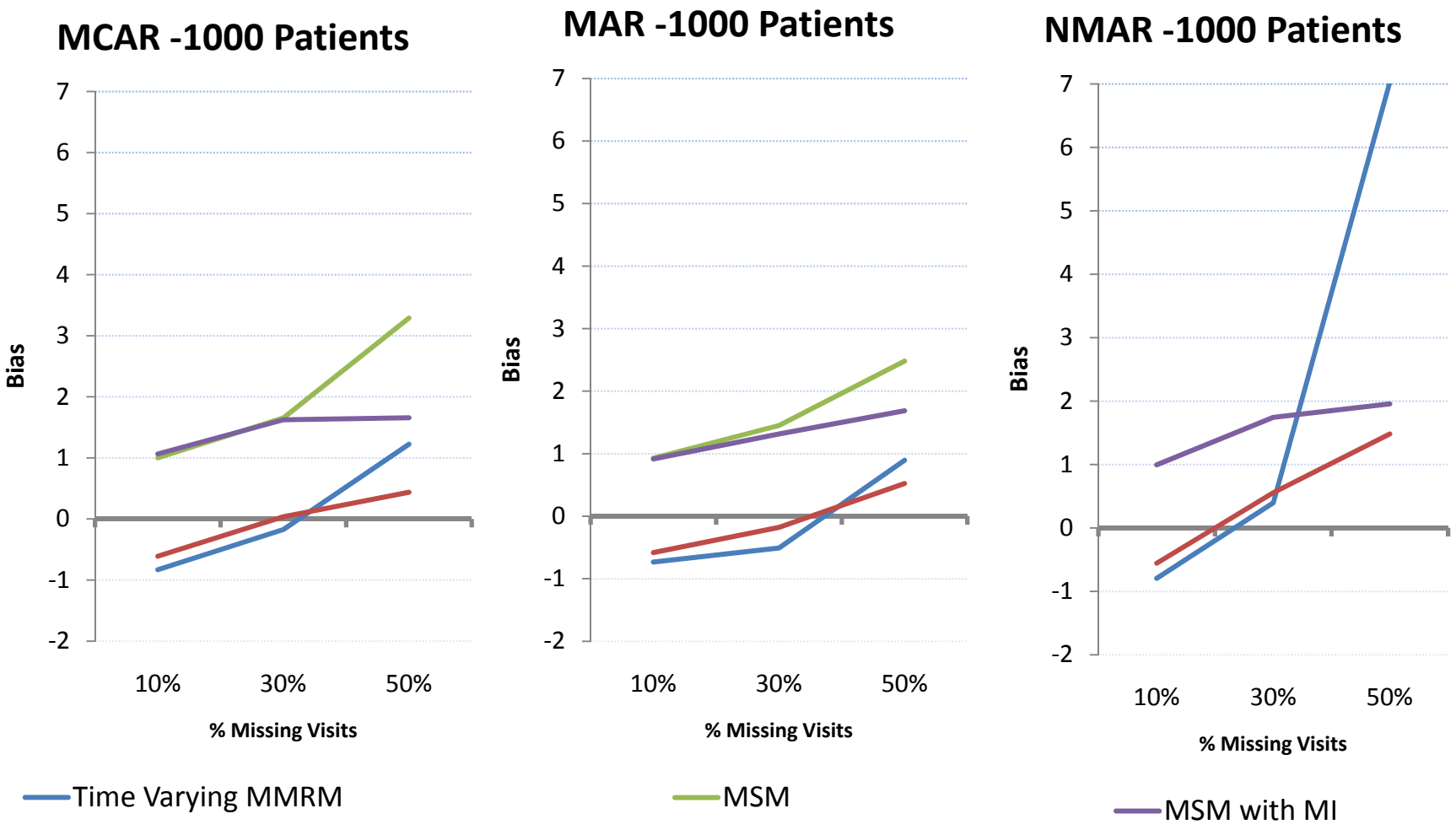
MRMM +/- ITT: Bias/Large Study:

Larger sample sizes do not reduce bias of ITT MMRM with MI



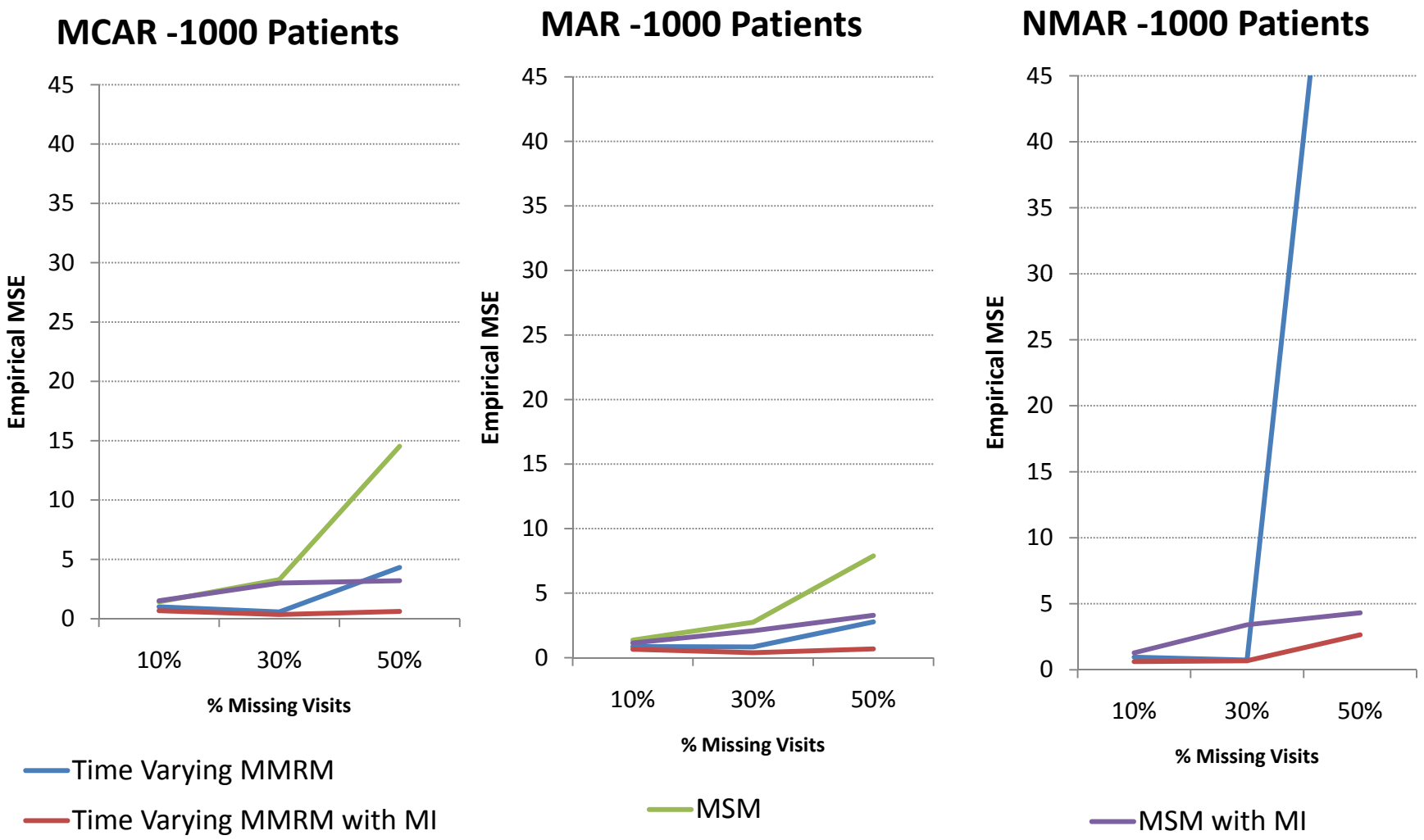
MRMM and MSM: Bias/Large Study:

Larger sample size did not decrease MSM bias



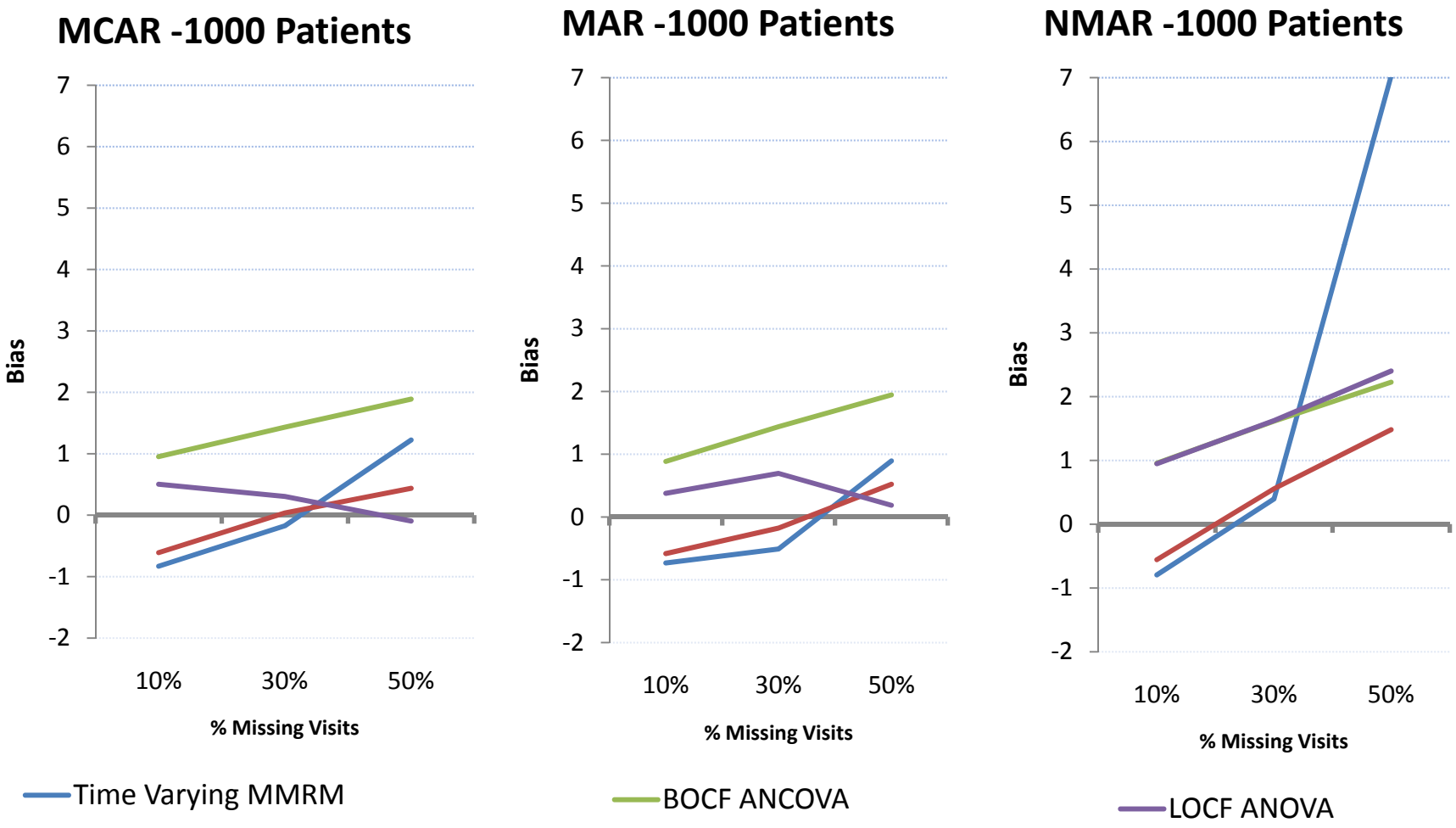
MRMM and MSM:MSE/Large Study:

Larger sample sizes reduce MSE for MSM models, particularly with MI



Bias/Large Study: MRMM, BOCF, LOCF

Even with larger sample size, MMRM without MI still has increased bias with NMAR missing visits of 50%



Part V – Conclusions

Conclusions

- In an analysis where MMRM should perform well with complete data, it generally did unless missingness was NMAR
- Multiple Imputation improved bias and precision for both MMRM and MSM, only having more bias for MMRM ITT
- Subgroups of completers and non-switchers did not perform as well as MMRM with MI, particularly at small sample sizes

Lilly

Conclusions

- We need to investigate further the bias associated with MSM and whether it is related to non-montonic patterns or errors in fitting models, which may be improved with exact methods
- We need to investigate whether MSM would outperform MMRM when “time dependent confounding” (when covariates are related to both treatment switching and the dependent variable)
- As expected due to the nature of the dependent measure, BOCF performed more poorly than LOCF, but LOCF would have performed similarly if visits varied more in their responses

Lilly

Acknowledgments

- Doug Faries
- Alan Brnabic
- Don Buesching
- Denai Milton
- Jin Xie
- Rosie Paczkowski

Lilly

References

1. Tunis S. *Evidence We Can Believe In: Pragmatic Attitudes in Clinical Research*. October 2008.
2. Committee on Comparative Effectiveness Research Prioritization – Institute of Medicine. *Initial National Priorities for Comparative Effectiveness Research*. The National Academies Press, 2009.
3. American College of Physicians. Information on Cost-Effectiveness: An Essential Product of a National Comparative Effectiveness Program, *Annals of Internal Medicine* 2008;148.
4. Institute of Medicine Roundtable on Evidence Based Medicine. *Learning What Works Best: The Nation's Need for Evidence of Comparative Effectiveness in Healthcare*. September 2007.
5. Agency for Healthcare Research and Quality (AHRQ). Definition of Comparative Effectiveness. http://www.effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness-research1/.
6. Medicare Payment Advisory Commission. *Report to the Congress: Reforming the delivery system.* Washington DC, 2008.
7. Orszag PR – Congressional Budget Office (CBO). *Research on the comparative effectiveness of medical treatments: Options for an expanded federal role*. CBO testimony before the Subcommittee on Health, Committee on Ways and Means, U.S. House of Representatives, 2007.
8. Center for Medical Technology Policy (CMTP). Comparative Effectiveness definitions. http://www.cmtpnet.org/comparative-effectiveness/comparative-effectiveness-definitions/?searchterm=None\.
9. Drummond MF, Schwartz JS, Jönsson B, Luce BR, Neumann PJ, Siebert U, Sullivan SD. Int J Technol Assess Health Care. 2008 Summer;24(3):244-58; discussion 362-68.
10. American Recovery and Reinvestment Act of 2009, PL. 111-5.

References (continued)

11. Patient Protection and Affordable Care Act of 2010, PL. 111-148. The law, intended to reform health insurance & provide coverage for uninsured Americans, includes the creation of PCORI in sec 6301.
12. Sabharwal R, Giffin R, Rangarao S. *What is Comparative Effectiveness Research and What will it Mean for Patients, Physicians, and Payers?* CMTP Report, July 2010.
13. Sonnad S, Goldsack J, Mohr P, Mullins CD, Whicher D. *Pragmatic Phase 3 Pharmaceutical Trials: Recommendations for the design of clinical trials that are more informative for patients, clinicians, and payers*. CMTP Effectiveness Guidance Document. September 2010.
14. Concato J, Peduzzi P, Huang GD, O'Leary TJ, and Kupersmith J. Comparative Effectiveness Research: What Kind of Studies Do We Need? *Journal of Investigative Medicine*. 2010.
15. Hernan M, Brumback B, and Robins J. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statist. Med.* 2002; 21:1689–1709.
16. Robins JM. Marginal structural models. In *1997 Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association: Alexandria, VA, 1998; 1–10.
17. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: The Environment and Clinical Trials*, Halloran E, Berry D (eds). Springer-Verlag: New York, 1999; 95–134.
18. Faries D and Kadziola Z. Analysis of Longitudinal Observational Data Using Marginal Structural Models. In *Analysis of Observational Health Care Data Using SAS*, edited by Faries D, Leon A, Haro J, and Obenchain R. SAS Press, January 2010.

Lilly

References (continued)

19. Moodie E, Delaney J, Lefebvre G, and Platt R. Missing Confounding Data in Marginal Structural Models: A Comparison of Inverse Probability Weighting and Multiple Imputation. *The International Journal of Biostatistics* 2008; 4(1).
20. Shortreed S and Forbes A. Missing data in the exposure of interest and marginal structural models: A simulation study based on the Framingham Heart Study. *Statist. Med.* 2010; 29: 431-43.

Lilly

Backup slides

Pragmatic CER Studies

Comparison with RCTs and Observational studies

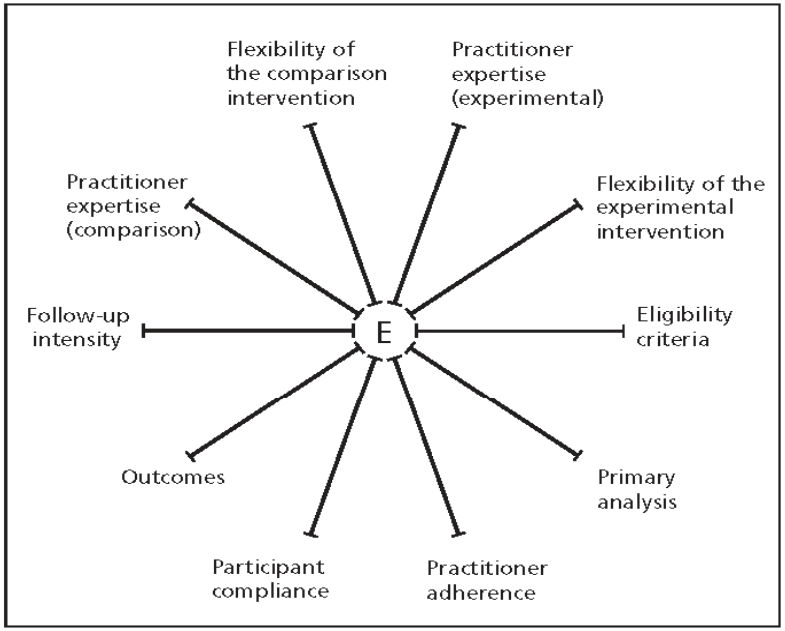
Characteristic	RCT	Pragmatic	Observational
Focus	Efficacy and safety; assess mechanistic effect; Can it work	Effectiveness and safety; assess / inform decision-making; Does it work under usual care conditions?	Effectiveness and safety; Does it work in actual practice?
Setting	Ideal / artificial	Real-world routine care (with potential minor departures)	Real-world clinical practice
Population	Strictly defined; homogenous	Typically broad; heterogeneous, comorbidities	Broad; heterogeneous, comorbidities
Randomization	Yes	Typically yes	No
Blinding	Typically yes	Typically no	No
Protocol-driven care (follow-up / safety)	Yes - fully interventional. High intensity of safety labs	Relaxed – minimally interventional Safety intensity per reg. allowance	No - non-interventional. Routine care labs only
Comparators	Placebo	Placebo, SOC, clinically-indicated care	SOC
Outcomes	Clinical surrogates; short term	Relevant (e.g., functional) outcomes; PROs	Relevant (e.g., functional) outcomes; PROs
Sample Size	Typically small	Typically larger	Typically large
Validity	High internal (↓ bias); low external (↓ generalizability)	Moderate to high internal; moderate to high external	Low internal; high external
Prospective/Retro	Prospective	Prospective	Prospective or retrospective
Comparable cost	Higher	Moderate	Lower

Pragmatic CER Studies

Level of pragmatism is measured on several domains

The pragmatic-explanatory continuum indicator summary (PRECIS) tool

- The pragmatic - randomized controlled trial (RCT) distinction is continuous rather than dichotomous.
- Typical pragmatic trials will likely land between a RCT and broad routine care (observational).
- The PRECIS tool can help assess the level of "pragmatism" when used as a discussion tool (e.g., in consultation with the EXPERT group).



The blank "wheel" of the pragmatic-explanatory continuum indicator summary (PRECIS) tool. "E" represents the "explanatory" end of the pragmatic-explanatory continuum.

This figure is adapted with permission from Thorpe et al. CMAJ 2009;180(10):E47-E57, copyright 2009, Canadian Medical Association or its licensors.

Pragmatic CER Studies

Explanation of PRECIS domains

Domain	Explanation
Eligibility Criteria	The most extremely pragmatic approach to eligibility would seek only to identify study participants with the condition of interest from as many sources (for example, institutions) as possible. As one moves toward the attitude of a RCT, additional restrictions will be placed on the study population.
Primary Analysis	The pragmatic approach to the primary analysis would typically be an intent-to-treat analysis of an outcome of direct relevance to the study participants and the population they represent. This analysis would make no special allowance for non-compliance, nonadherence or practice variability.
Practitioner Adherence	The pragmatic approach takes account of the fact that providers will vary in how they implement an intervention. A purely pragmatic approach, therefore, would not be concerned with how practitioners vary or "customize" a trial protocol to suit their setting.
Participant Compliance	The pragmatic approach recognizes that noncompliance with any intervention is a reality in routine medical practice. Because measurement of compliance may possibly alter subsequent compliance, the pragmatic approach in a trial would be not to measure or use compliance information in any way.
Outcomes	The pragmatic approach would be to not have a central outcome assessment, and to rely on usual training and measurement to determine the outcome status.
Follow-up Intensity	The pragmatic position would be not to seek follow contact with the study participants in excess of the usual practice for the practitioner.
Practitioner Expertise (Comparison)	The pragmatic approach aims to find out the benefits and harms of the intervention in comparison with usual practice in the settings of interest.
Flexibility of the Comparison Intervention	A pragmatic trial would typically compare an intervention to "usual practice" or the best alternative management strategy available.
Practitioner Expertise (Experimental)	A pragmatic approach would put the experimental intervention into the hands of all practitioners treating (educating, etc.) the study participants.
Flexibility of the Experimental Intervention	The pragmatic approach leaves the details of how to implement the experimental intervention up to the practitioners. In addition, the pragmatic approach would not dictate which co-interventions were permitted or how to deliver them.

Methods

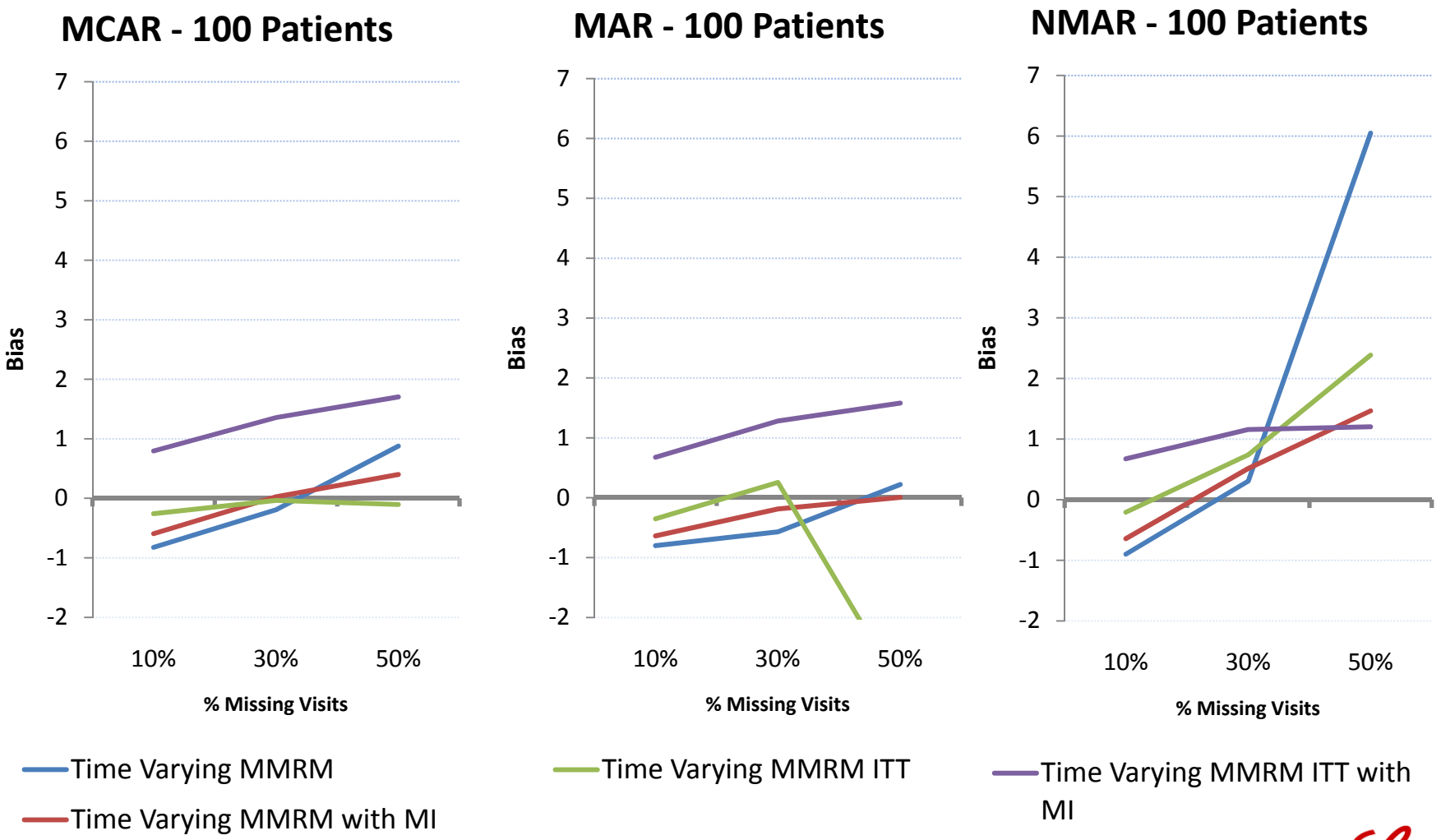
Marginal Structural Models (MSM)

- Intent to treat (ignore the switching)
  - Good for policy questions (consequences of Tx initiation), not for CER
- Subset analysis (eliminate the switching)
  - Adds bias and reduces precision
- MSM (model the switching)
  - Weighted repeated measures models
  - A 'longitudinal' version of Propensity Scoring
  - Good for time-varying confounding (e.g., disease severity)
    - A time dependent variable that predicts both the outcome and subsequent treatment, and is also predicted by past treatment
  - Advantages
    - ✓ incorporates all of the data
    - ✓ adjusts for selection biases
    - ✓ adjusts for bias due to dropouts (monotone MAR)
    - ✓ 'relatively' easy to implement and explain

Lilly

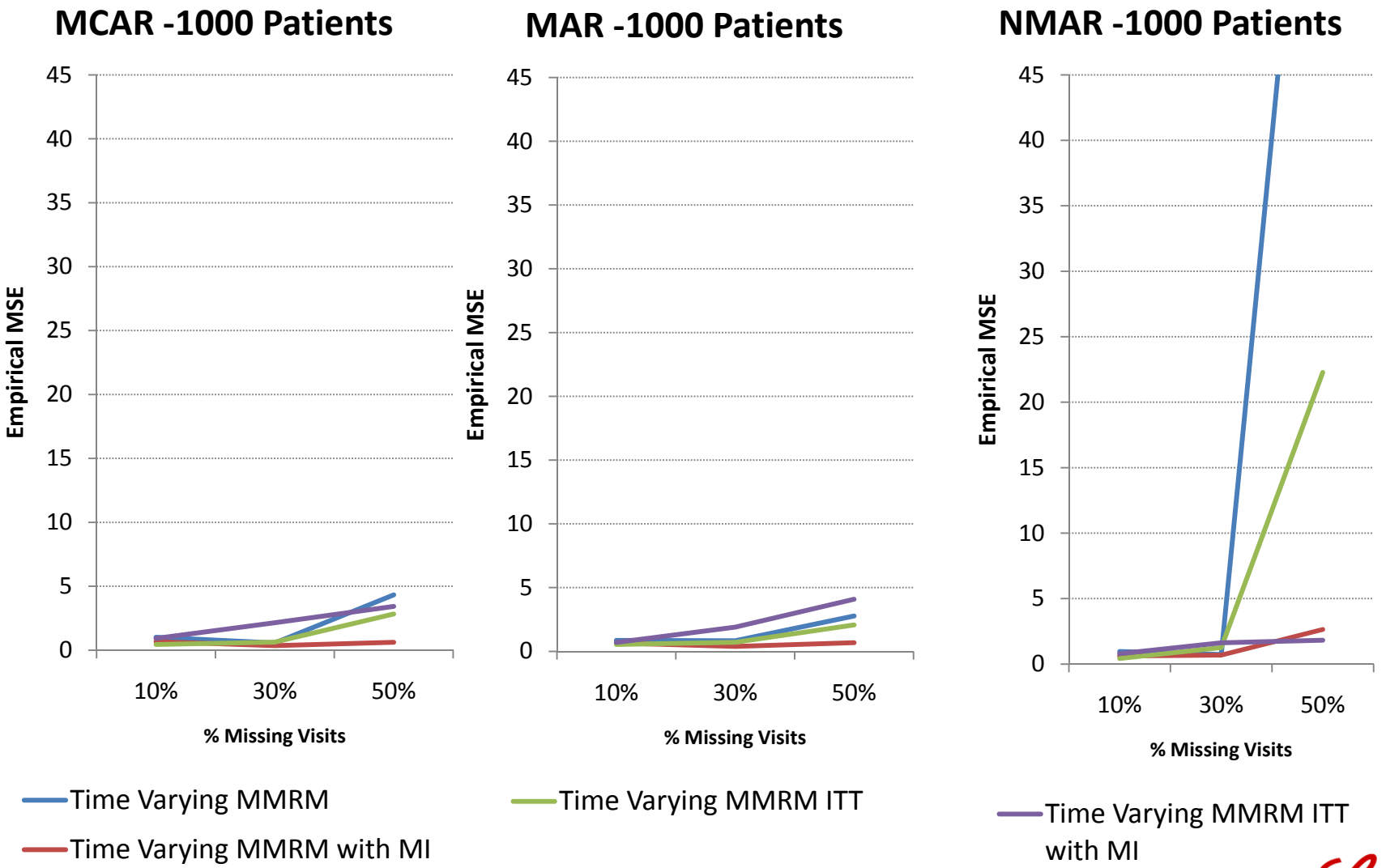
Bias/Small Study: MRMM +/- ITT

ITT, particularly with MI, increases bias



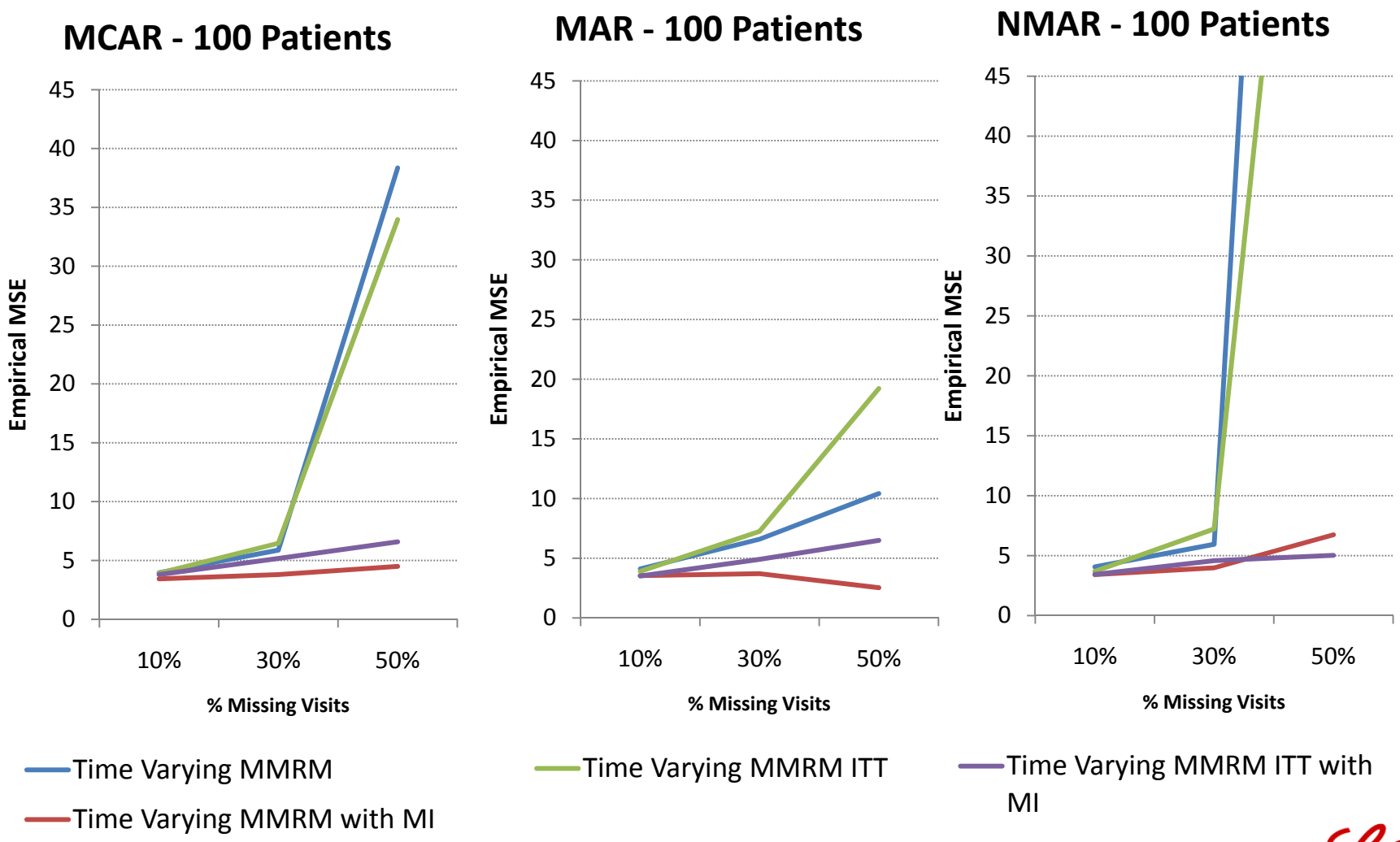
MSE/Large Study: MRMM +/- ITT

Larger sample sizes improve efficiency of ITT models



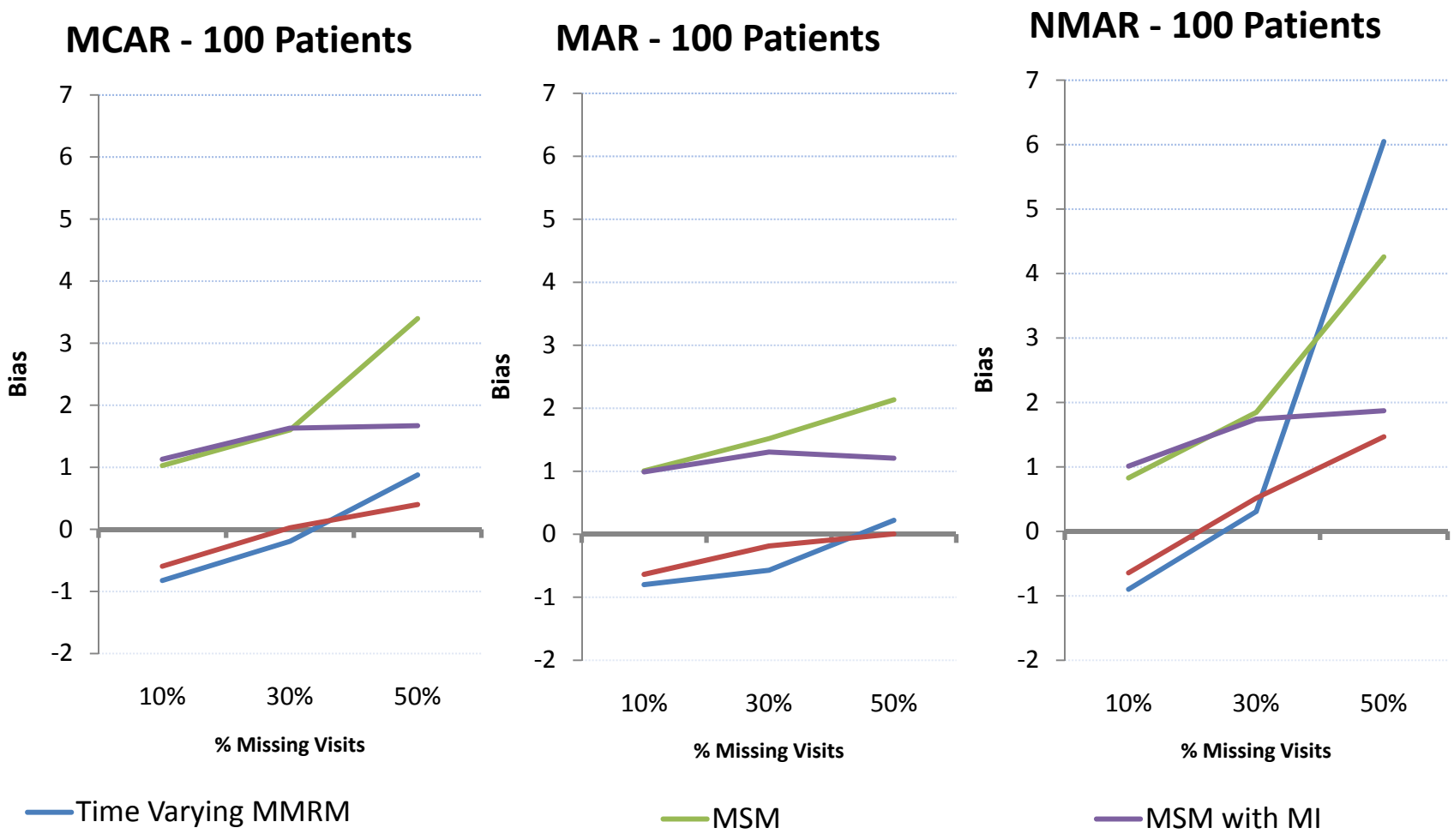
MSE/Small Study: MRMM +/- ITT

Efficiency of ITT reduced when 50% MAR missing



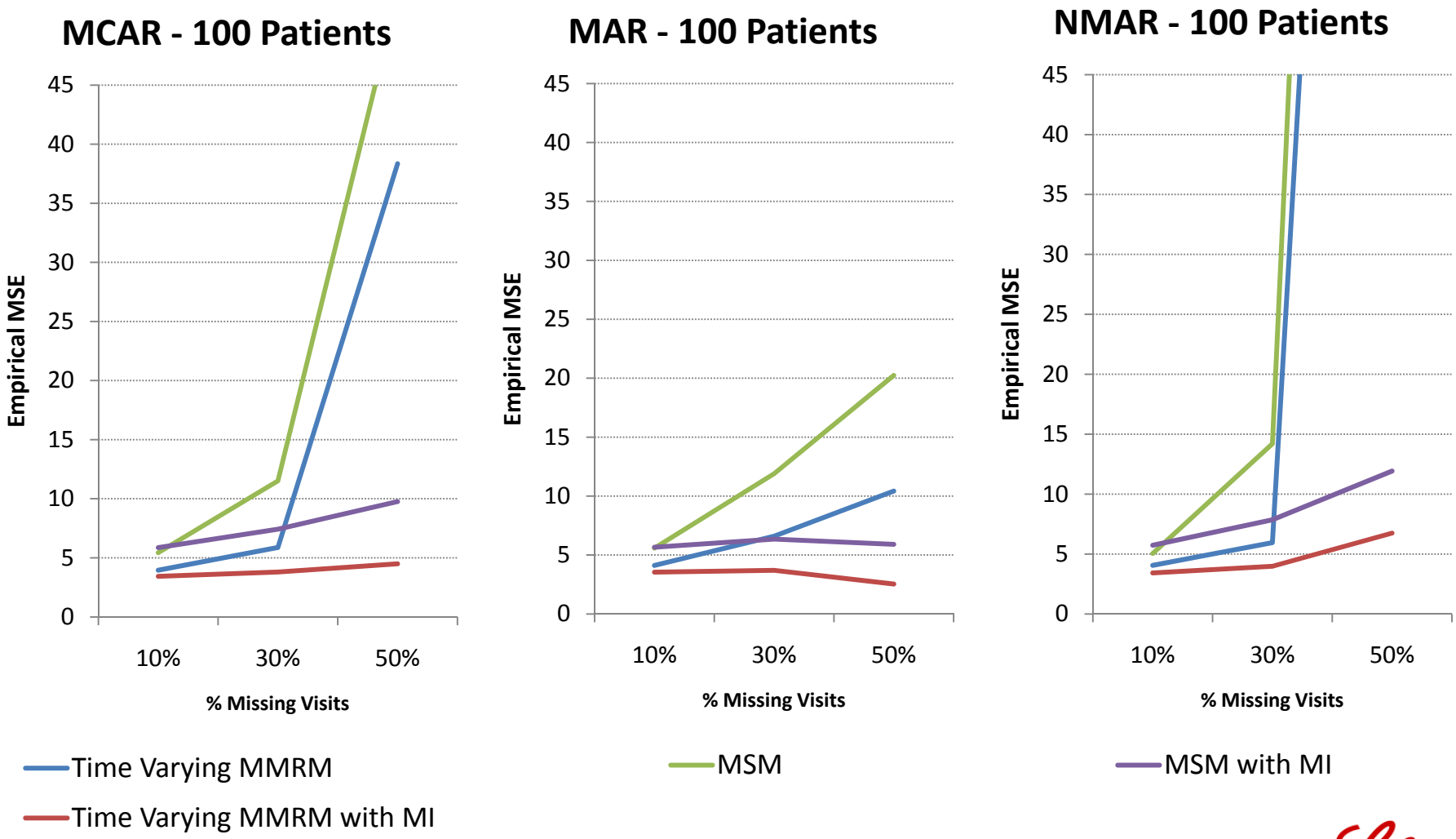
Bias/Small Study: MRMM and MSM

In our example, MSM tends to increase bias, but somewhat improved with MI



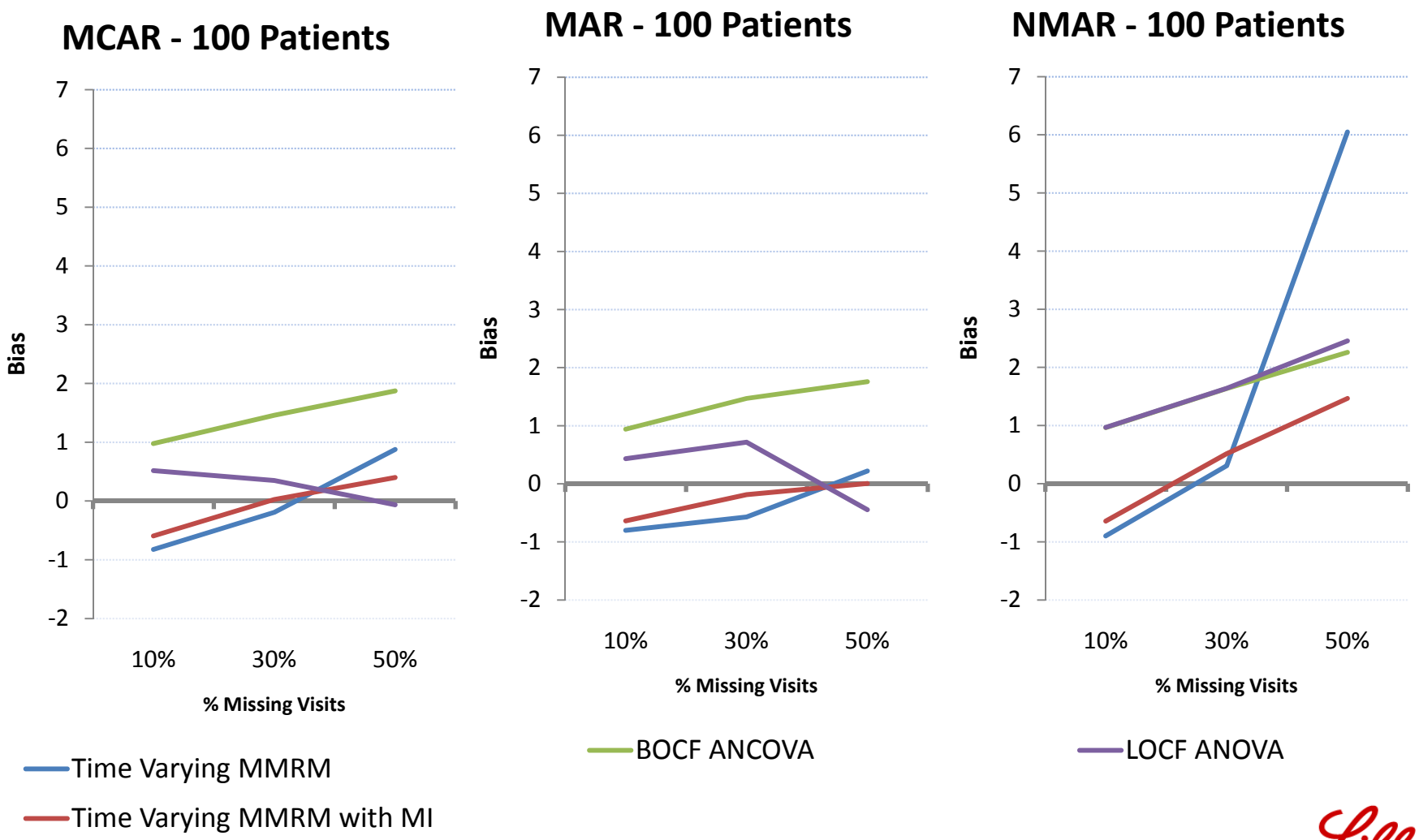
MSE/Small Study: MRMM and MSM

MSM efficiency improves with MI

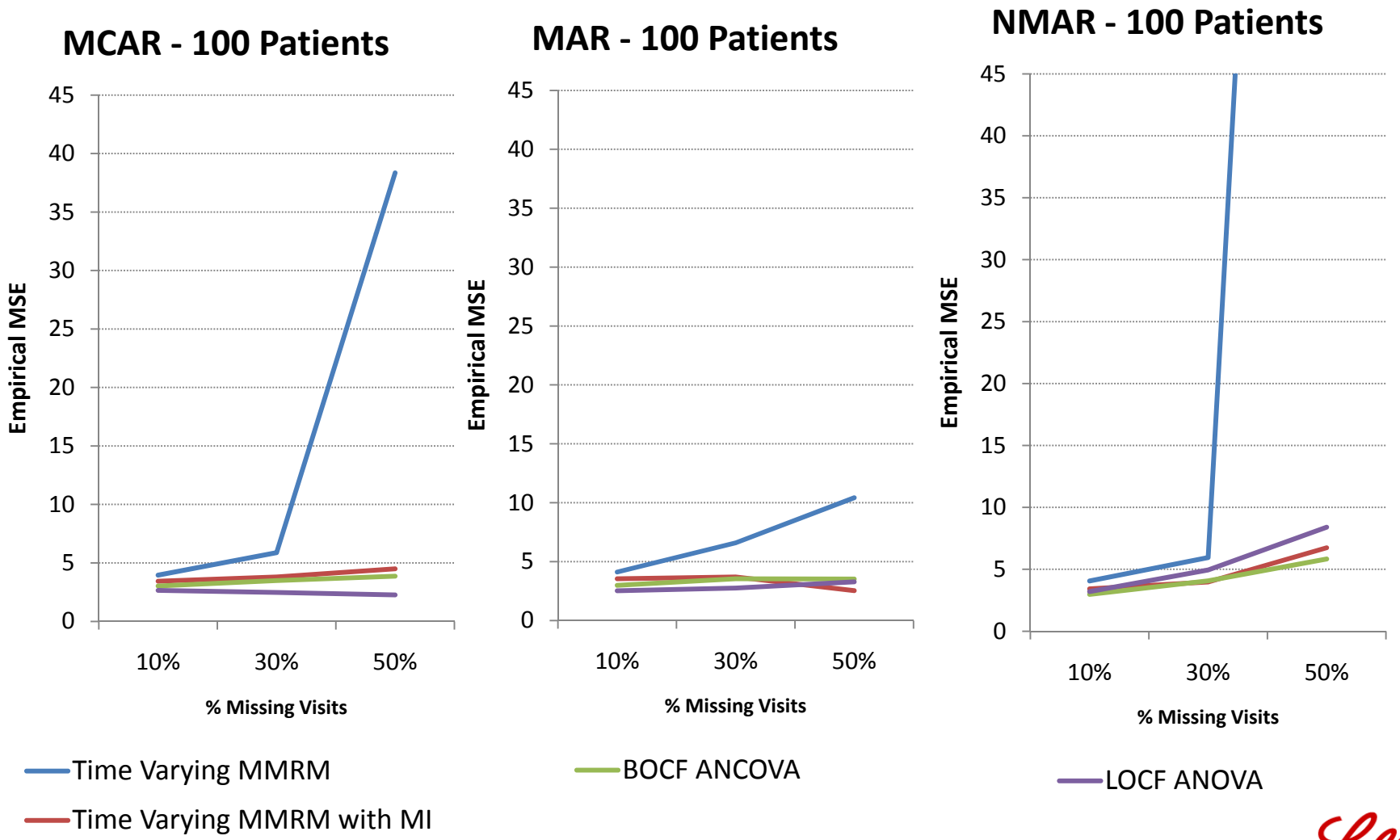


Bias/Small Study: MRMM, BOCF, LOCF

BOCF was biased in all cases, and LOCF for NMAR



MSE/Small Study: MRMM, BOCF, LOCF



MSE/Large Study: MRMM, BOCF, LOCF

Larger sample sizes reduce MSE except for NMAR at 50% missing