Applied Missing Data: Causal Estimands and Clinicians

BASS XXII
November 2015

Steven Gilbert

steven.a.gilbert@pfizer.com

Causal Effects and Clinicians

Annu. Rev. Public Health. 2000. 21:121–45

CAUSAL EFFECTS IN CLINICAL AND EPIDEMIOLOGICAL STUDIES VIA POTENTIAL OUTCOMES: Concepts and Analytical Approaches

Roderick J. Little
University of Michigan, Ann Arbor, Michigan 48109-2029; e-mail: rlittle@umich.edu

Donald B. Rubin
Harvard University, Cambridge, Massachusetts 02138; e-mail: rubin@hustat.harvard.edu

Key Words statistical inference, randomization, observational studies, noncompliance, missing data

This Lecture

Concepts and not too much math

Things you need to discuss with clinical colleagues to design your study

(the reason there isn't too much math)

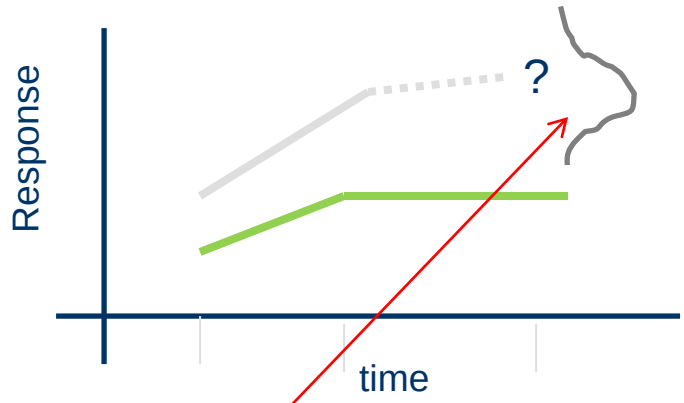
Missing data FAQs

Important Points

- **Missing data requires going back to basics, not necessarily something new.**
- **Good to think about partially observed multivariate data, not missing univariate data.**
- **Start thinking about an ‘ideal’ experiment and see how missing data ‘breaks’ that experiment.**
- **Science first, statistical methods last.**
- **Pattern Mixture Models and MI are particularly simple to explain to clinicians even if they are not always the best statistical choice.**

Multivariate vs Univariate

Visit	Y	R
1	5	1
2	6	1
3	?	0



- Distribution
- Regression
- MVN
- ICE

$$f(y_1,y_2,y_3)=f(y_1)f(y_2|y_1)f(y_3|y_2,y_1)$$

$$f(y1,r1,y2,r2,y3\ ,r3)=f(y1,r1)f(y2,r2|y1,r1)f(y3\ ,r3\ |y2,r2,y1,r1)$$

Multivariate distribution can be extended to include covariates and post-baseline response for other endpoints

Cart Before the Horse

Cart before the horse

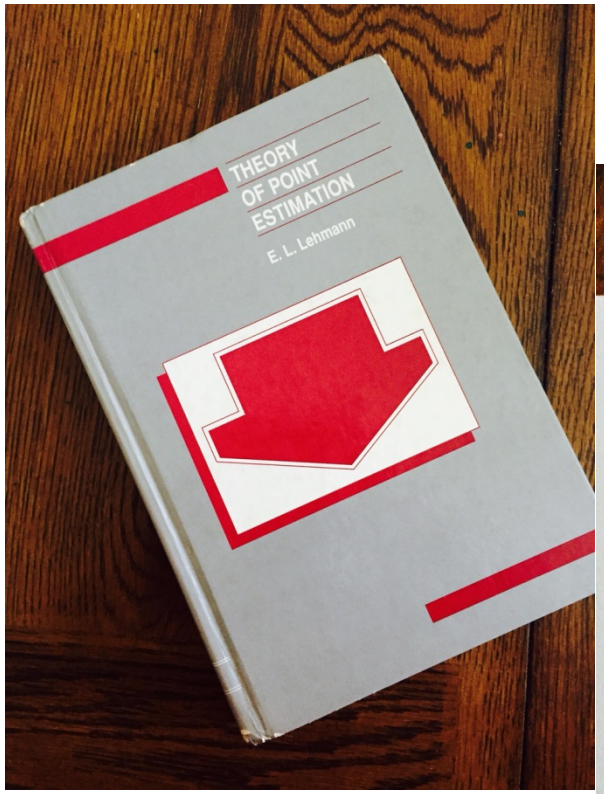
Statisticians

- Often want to start with analysis method and worry about interpretation later
- Off the shelf missing data methods (eg. Assuming the data are MAR and MMRM will provide an unbiased estimate .)

Drug Developers

- Start with the label and work backwards
- Not a bad idea but has some problems we'll discuss shortly

Back to Grad School - Math/Stats



Section 2

2.1 If $X_1, \ldots, X_n$ are iid as $N(\xi, \sigma^2)$ with σ^2 known, find the UMVU estimator of (i) ξ^2, (ii) ξ^3, (iii) ξ^4. [*Hint*: To evaluate the expectation of $\bar{X}^k$, write $\bar{X} = Y + \xi$, where Y is $N(0, \sigma^2/n)$ and expand $E(Y + \xi)^k$.]

2.2 Solve the preceding problem when σ is unknown.

2.3 In Example 2.1 with σ known, let $\delta = \Sigma c_i X_i$ be any linear estimator of ξ. If δ is biased, its risk $E(\delta - \xi)^2$ is unbounded. [*Hint*: If $\Sigma c_i = 1 + k$, the risk is $\geq k^2\xi^2$.]

2.4 If X is a single observation from $N(\xi, \sigma^2)$, show that no unbiased estimator δ of σ^2 exists when ξ is unknown. [*Hint*: For fixed $\sigma = a$, X is a complete sufficient statistic for ξ, and $E[\delta(X)] = a^2$ for all ξ implies $\delta(x) = a^2$ a.e.]

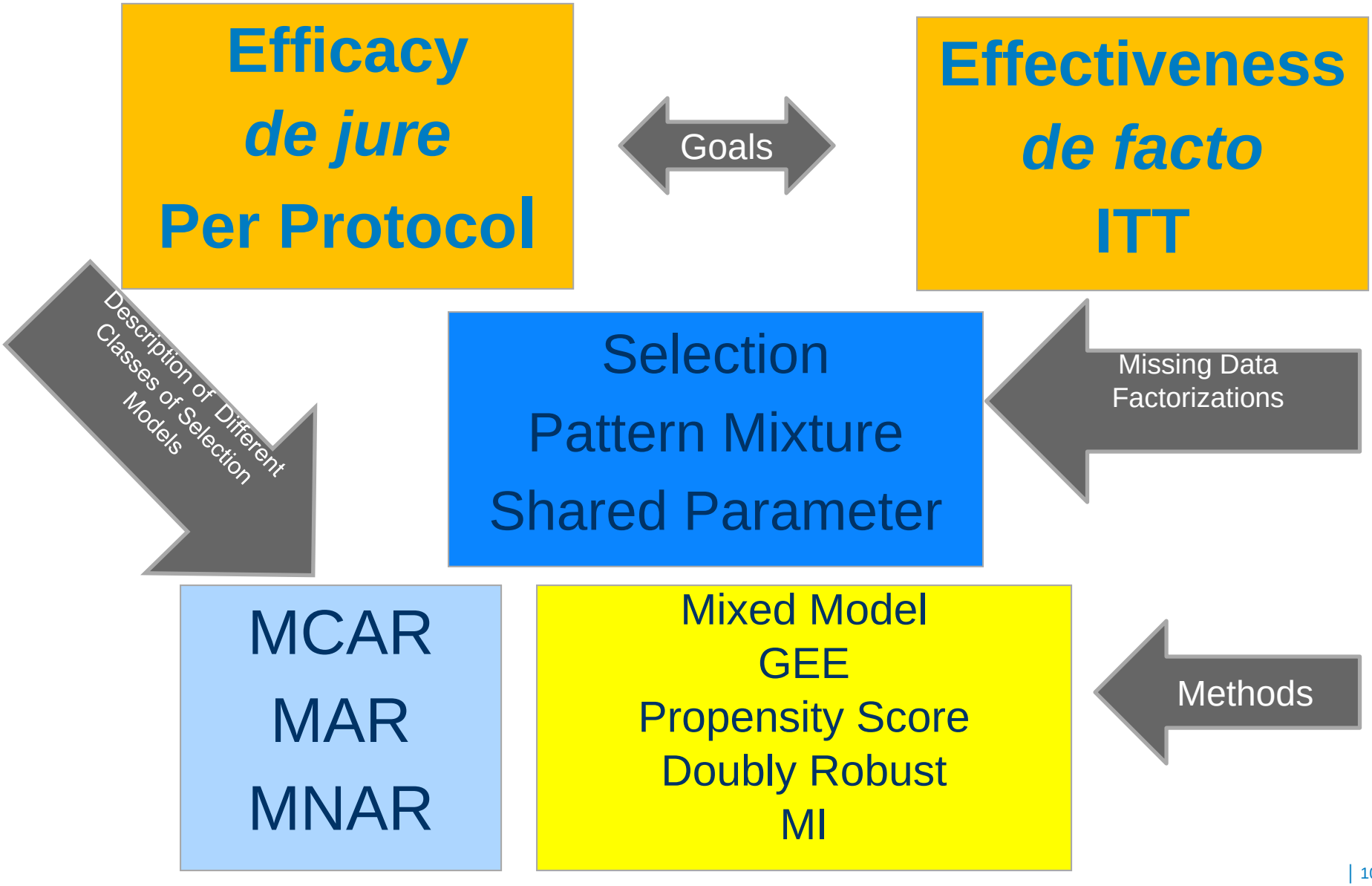
Assumes:

1. target, θ, is given
2. is well defined
3. unique

Not Always True !!

1. Target, θ, is a scientific choice
2. Not always well defined
3. You can have more than one estimand

Some Confusing Terminology



Causal Estimand

An estimand reflects what is to be estimated to address the scientific question of interest posed by a trial.

The choice of an estimand involves:

- **Population of interest**
- **Endpoint of interest**
- **Measure of intervention effect**

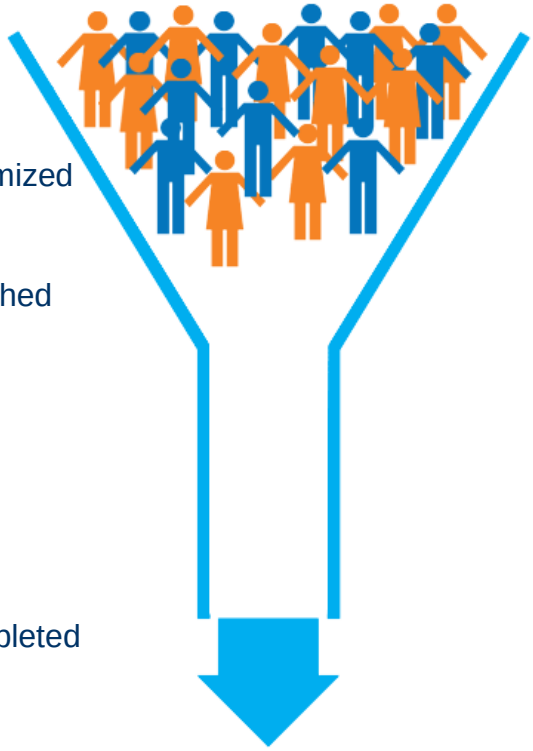
Chrissie Fletcher (one of the EFPIA representatives on the ICH E9 Working Group)

EFSPI Statistics Leaders 2015

Closer look at intervention effect

Who?

Screened



Randomized

Enriched

Completed

When?

Early Visit

Landmark Visit

End of Study

Entire Study

What?

- Randomization Effect
- Drug Effect
- Drug Effect if everyone followed protocol
- Real world drug effect
- Effect while taking drug
- Efficacy/de jure
- Effectiveness / de facto

Retrieved Dropout Data

- **Distinction between study withdrawal and treatment withdrawal.**
- **Do not need to have a patient withdraw from the study just because they stopped taking drug.**
- **Data collected while a patient is on study but off of treatment is sometimes called retrieved dropout data.**
- **Use of retrieved dropout data is controversial.**
- **FDA has become more nuanced about their stand on this.**
  - Older advice was that retrieved dropout used for outcome trials and not used for symptomatic trials
    - Treatment to prevent stroke or death, follow-up and use real data regardless of patient is complying.
    - Treatment for back pain, ignore data after drug withdrawal.

Examples of a Causal Estimands I

1. **Difference in outcome improvement at the planned endpoint for all randomized participants.**
   - ITT/*de-facto* requires retrieved dropout data
2. **Difference in outcome improvement in tolerators**
   - Uses randomized withdrawal design
3. **Difference in outcome improvement if all patients adhered**
   - De-jure / PP
4. **Difference in areas under the outcome curve during adherence to treatment, and,**
5. **Difference in outcome improvement during adherence to treatment**
   - No missing data

Examples of a Causal Estimands II

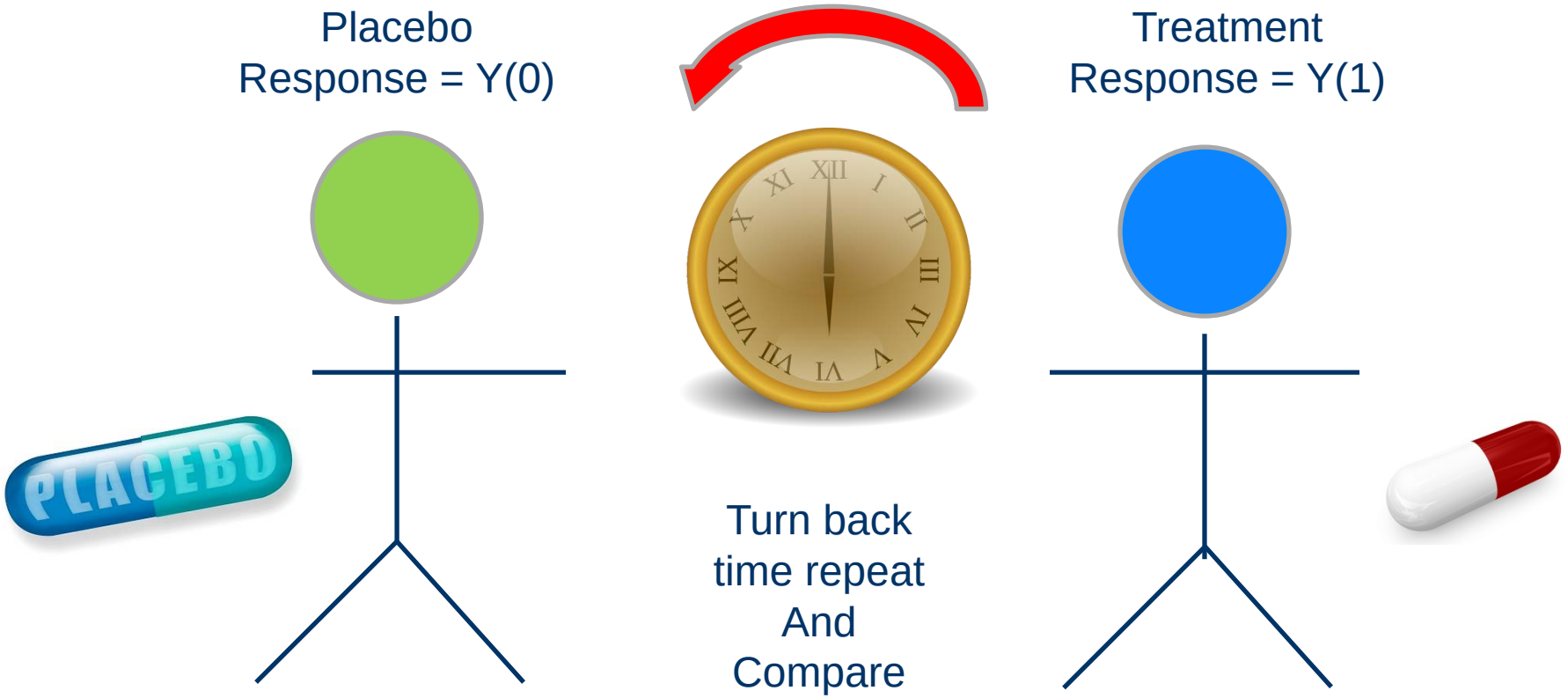
6. **Difference in outcome improvement in all randomized patients at the planned endpoint of the trial attributable to the initially randomized medication**
   - Similar to estimand 1. but it is not confounded with rescue medication
   - Not strictly an ITT or PP estimand

Missing Data: Turning Guidance Into Action

Table 1. Estimands and their key attributes

Estimand	Hypothesis	Inference	Population	Endpoint	Use of data after withdrawal of randomized study medication
1	de-facto (effectiveness)	Treatment policy	All patients	Planned endpoint	Included in primary analysis
2	de-jure (efficacy)	Initially randomized medication	Tolerators	Planned endpoint	Not included in primary analysis
3	de-jure (efficacy)	Initially randomized medication	All patients	Planned endpoint	Not included in primary analysis
4	de-facto (effectiveness)	Initially randomized medication	All patients	Undefined	Not included in primary analysis
5	de-facto (effectiveness)	Initially randomized medication	All patients	Undefined	Not included in primary analysis
6	de-facto (effectiveness)	Initially randomized medication	All patients	Planned endpoint	Likely imputed

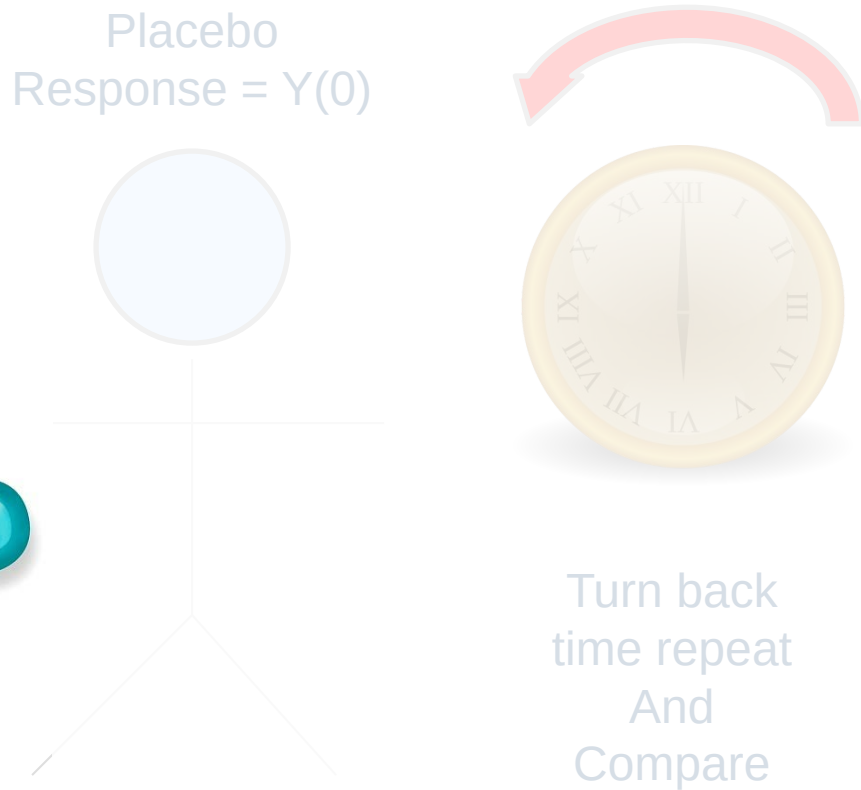
Aside – Ideal Experiment and Potential Outcomes



Individual Treatment Effect

Everything is identical except treatment

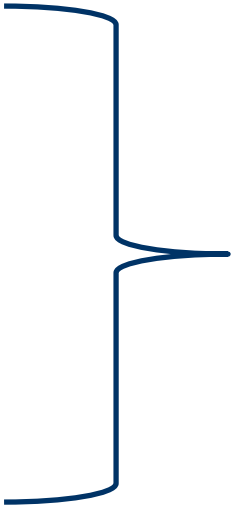
At end of experiment the only difference is the assigned treatment



Everything is identical except treatment

Less than ideal cases

Placebo	Treated
Placebo + Rescue Med	Drug
Placebo	Drug + early withdrawal
Placebo	Drug + Rescue Med
Placebo + Early Withdrawal	Drug + Rescue + early withdrawal



What is the treatment?

Acute vs Chronic conditions
Benefit over trial vs Landmark Analysis

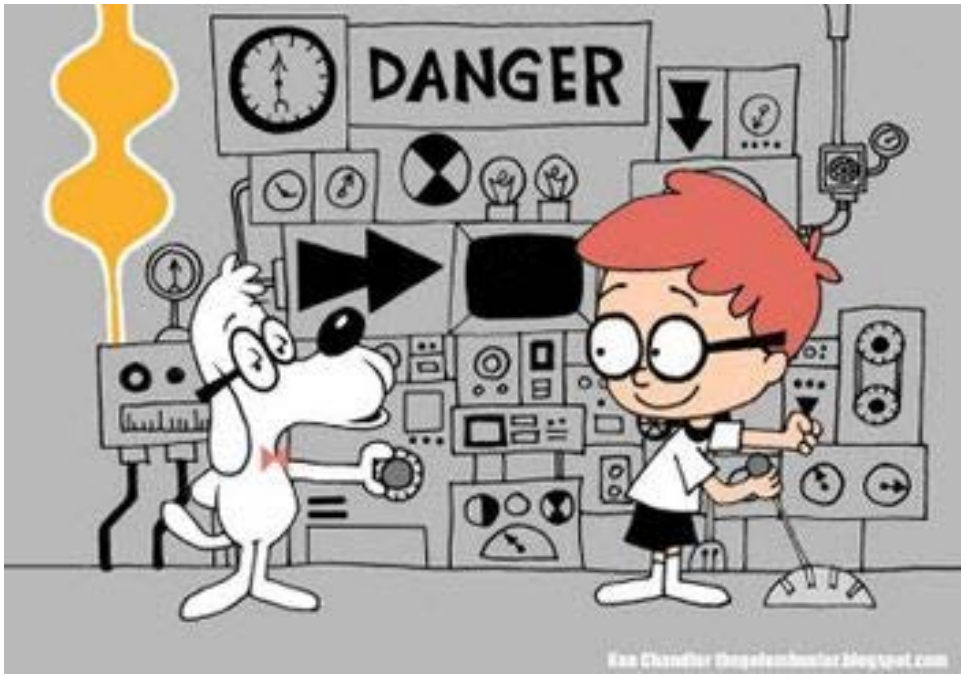
Pop Quiz

Is a per protocol analysis a valid way to get estimate a pure treatment effect?

Compare only those placebo and treated subjects who followed the protocol with no violations

We'll get back to this in a few slides

Another Problem with Real World Experiments



No Time Machine

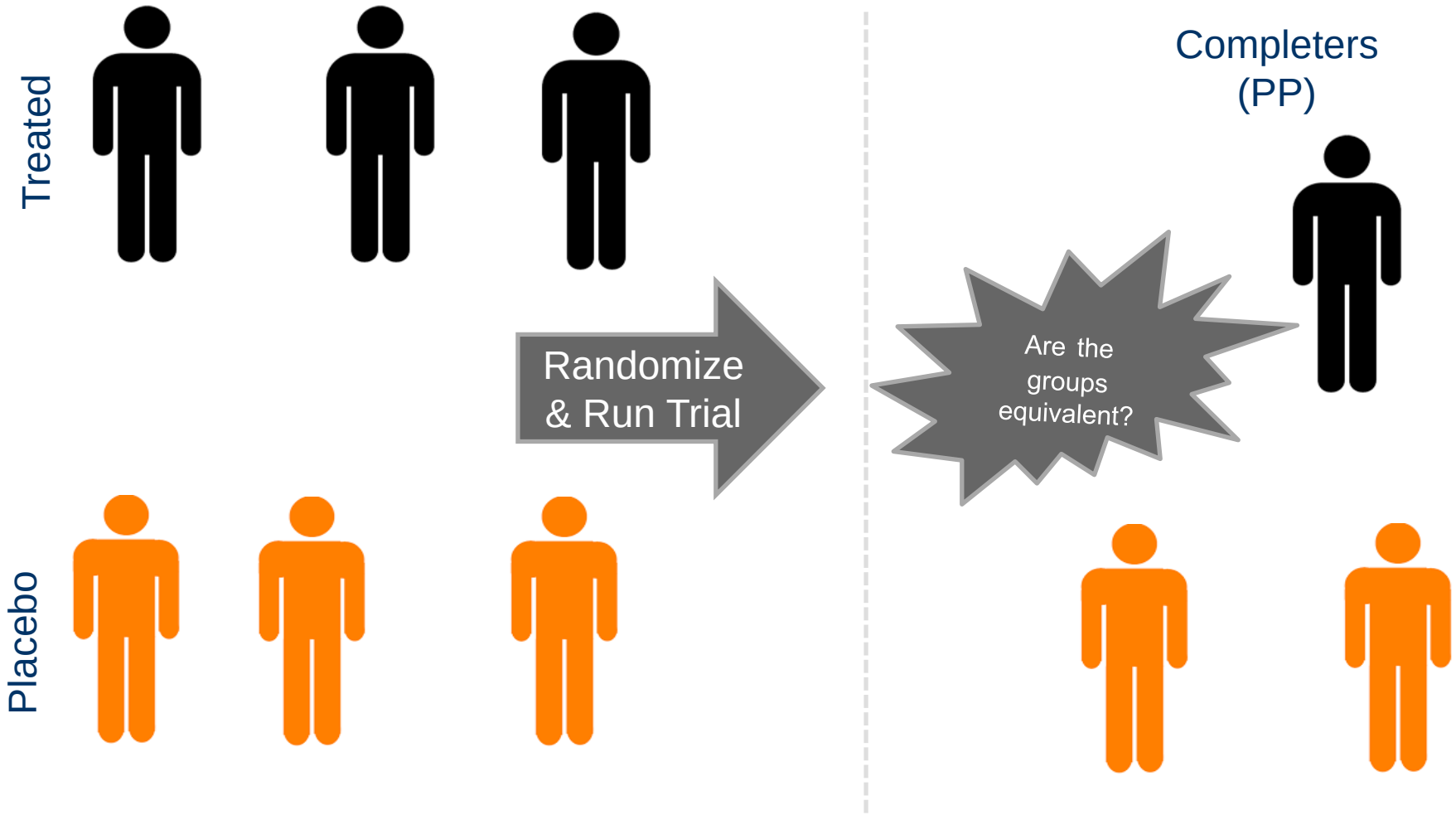
Real World Experiments

1. **Approximate Ideal Experiment as Group Level**
   - Sample/randomize ‘statistically’ equivalent groups and compare at group level
2. **Approximate Ideal Experiment at Individual Level**
   - Case control
   - Cross Over

Compare Like With Like

Compare Like Groups Not Individuals

Selection Bias?



Per Protocol Analysis

Are treated per protocol patients comparable to placebo per protocol patients?

If not the analysis is problematic to interpret.

Missing Data Causes Break from Ideal Experiment

1. **Break the ‘case-mix’ (Group Level):**
   - Patients at end of trial are no longer a representative sample of the population
   - Differential dropout in treated and placebo arms and can not be compared directly
2. **Break the Individual Treatment Effect (Individual Level)**
   - Other factors beyond assigned treatment change between treated and untreated patients

Pattern Mixture Models

- **In a nutshell**

P[Y,R|X]=P[Y|X]P[R|Y,X] <- Selection Factorization

=P[Y|R,X] P[R|X] <- Pattern Mixture Factorization

Distribution of Data given you are in pattern R=r

Not trivial to fit

Patterns generally specified *a priori* based on

- Completer status
- Reason for dropout
- Time of dropout
- Treatment

Final estimate is weighted average of conditional estimates. Need to account for reason data is missing (pattern). No counterpart to ignorability

Trivial to fit distribution – (binomial or multinomial)

How to fit P[Y|R=r]

- **Identifying Restrictions** **(see Molenberghs and Verbeke)**
  - MCAR, MAR, MNAR are descriptions of selection model factorizations
  - PMM use concepts such as CCMV, ACMV, NCMV, restrictions and non-future dependence
  - Link between MAR and ACMV (available case missing value) restriction
- **Practical Work often based on simpler concepts**
  - Fit longitudinal model with linear or polynomial time trends, categorical pattern and interaction between the two
    - Not usually used in clinical trials because of unknown response patterns. Instead a saturated mode is often used.

How to fit P[Y|R=r]

- **Many clinical trials use a saturated linear mixed model with categorical time, treatment and categorical time by treatment interaction. This provides separate estimates for each combination of time and treatment. These can also be combined in a PMM.**

FDA Stat Review of Drug for Chronic Pain

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	206-627
Drug Name:	Hysingla ER (hydrocodone bitartrate) extended-release tablet
Indication(s):	Management of moderate to severe pain when a continuous, around-the-clock analgesic is needed for an extended period of time
Applicant:	Purdue Pharma. L.P.
Date(s):	Letter date: April 28, 2014 PDUFA date: October 28, 2014
Review Priority:	Priority
Biometrics Division:	II
Statistical Reviewer:	Yan Zhou, Ph.D.
Concurring Reviewer:	Janice Derr, Ph.D.

Quote from Review

Usually, in chronic pain studies, there are high percentages of subjects who discontinue the study early due to various reasons …. The current approach favored by the division is that a drug intended to treat chronic pain is not efficacious if subjects cannot stay on the treatment for the trial duration. Thus, strategies to handle missing data should not attribute any treatment benefit to subjects discontinuing from the study. In July 2010, the National Academy of Sciences (NAS) released a report on the prevention and treatment of missing data. The NAS report discourages single imputation methods. In light of the NAS report, the applicant's primary analysis utilized a mixed effects model with repeated measures (MMRM) incorporating a pattern mixture model (PMM) framework to account for missing data.

Quote from Review

The primary comparison between Hysingla ER and placebo was ..

based on a weighted average of the estimated mean PI scores for three patterns of subjects, while weight was the proportion of each pattern of subjects within a treatment. The primary analysis method had a desirable feature in that a bad outcome was attributed to subjects that discontinued the study due to AEs. The method additionally account for sources of variability introduced by different patterns of missing data.

How do we account for variability?

- Pattern mixture analysis is essentially just a stratified analysis
- Final estimate requires putting the stratum specific estimates back together
- Example

  PMM = [proportion in pattern] X [pattern estimate]

  = L` **β**

  In standard linear model analysis L vector is fixed and **β** vector is random

  In PMM analysis both are random

Pattern estimates from MMRM based on model parameters, **β**

Can output estimates and covariance matrix of estimates from software

Proportions are also statistics based on simple multinomial model, with estimates (**π**'s) and covariance matrix

PMM estimate is a linear combination of the β's and **π**'s.

- In linear model framework you can use
  - Delta method
  - Bootstrap
- MI, uses PMM to define imputations
  - Usual rules for combining MI (Rubin's rules)

Back to FDA Report

Table 6: Primary efficacy analysis

Study Period/Week Mean Pain Intensity	Placebo[a] (N=292)	HYD (N=296)
Baseline		
n	292	296
Mean (SD)	7.4 (1.19)	7.4 (1.13)
Prerandomization		
n	291	296
Mean (SD)	2.8 (1.15)	2.8 (1.16)
Double-blind Week 12		
n	199	218
Mean (SD)	3.7 (2.04)	3.3 (1.93)
Pattern 1: Completed Week 12**		
n (%)	210 (72)	229 (77)
LS Mean (SE)	4.17 (0.131)	3.47 (0.128)
Pattern 2: Discontinued Study Drug due to Adverse Event or ASHA Related Event**		
n (%)	11 (4)	18 (6)
LS Mean (SE)	7.40 (0.048)	7.40 (0.048)
Pattern 3: Discontinued Study Drug due to Other Reasons**		
n (%)	71 (24)	49 (17)
LS Mean (SE)	3.90 (0.114)	3.38 (0.112)
Repeated Measures Analysis/Least Squares Means (SE) at Double-blind Week 12 from PMM		
LS Mean (SE)	4.23 (0.126)	3.70 (0.128)
Treatment Comparison at Double-blind Week 12		
Difference in LS means from Placebo (Mean (SE))		-0.53 (0.180)
P value vs Placebo		0.0016
95% CI for difference from Placebo		(-0.882, -0.178)

Source: Clinical Study Report Table 14.2.1.1.1 and Table 14.2.1.1.2.

4.17*0.72 + 7.4*0.04 +3.90*0.24=4.23

3.47*0.77 + 7.4*0.06 +3.38*0.17=3.7

FAQs

1. Which method is better?
2. What's so bad about LOCF?
3. Doesn't MAR make the least assumptions?
4. Is MI just for MAR?
5. How do I pick a sensitivity analysis?
6. What's wrong with starting from the label and working backwards?
7. What's the most important thing to know about missing data?

1. Which method is better?

- **Chronic condition. Endpoint is based on investigator score at 12 weeks. Which is better?**
  1. Responder if score drops by X and still taking drug?
  2. Longitudinal analysis including at all measured weeks (1,2,4,8,12).

  How do we decide?
  Are they estimating the same thing?

1. Answer

- **One method is *not* better than the other. They both have different estimands and should not be compared on statistical grounds such as power, etc.**

1. **Responder analysis, uses a composite endpoint. Responder if both positive response and you can tolerate drug. (*de facto* type of estimand)**
2. **Longitudinal analysis. Usual assumption is that with the observed responses up to drop out and included covariates, the data are missing at random (MAR). The implication is that the model assumes that patients who drop early will have a response similar to completers with a similar response profile up to dropout and similar covariates if they had adhered to the protocol. (*de jure* type of estimand)**

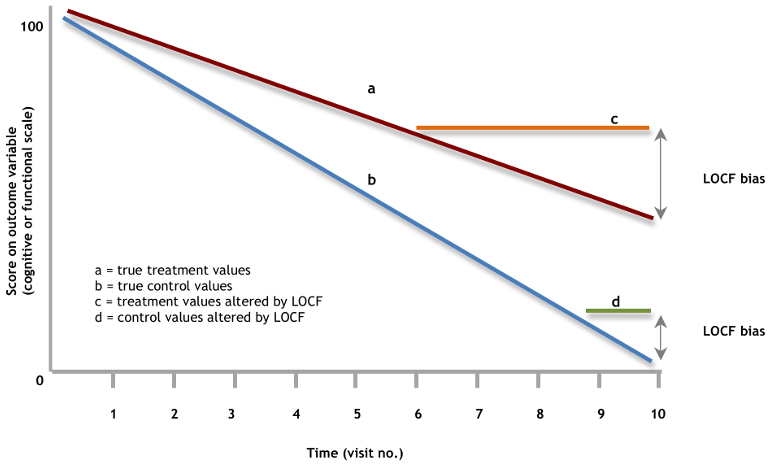
2. What's so bad about LOCF?

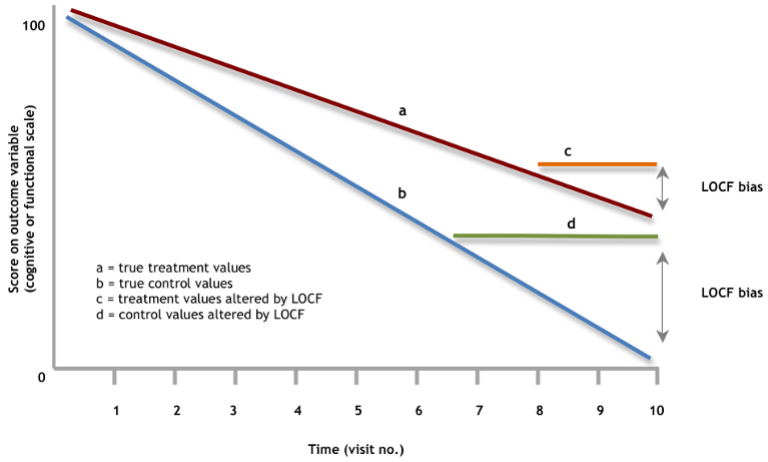
- **LOCF = Last Observation carried forward**

Patient	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5
1	3	4	5	6	5
2	2	3	*	*	*

Patient	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5
1	3	4	5	6	5
2	2	3	3	3	3

2. It confounds time and mean





Have last-observation-carried-forward analyses caused us to favour more toxic dementia therapies over less toxic alternatives? A systematic review

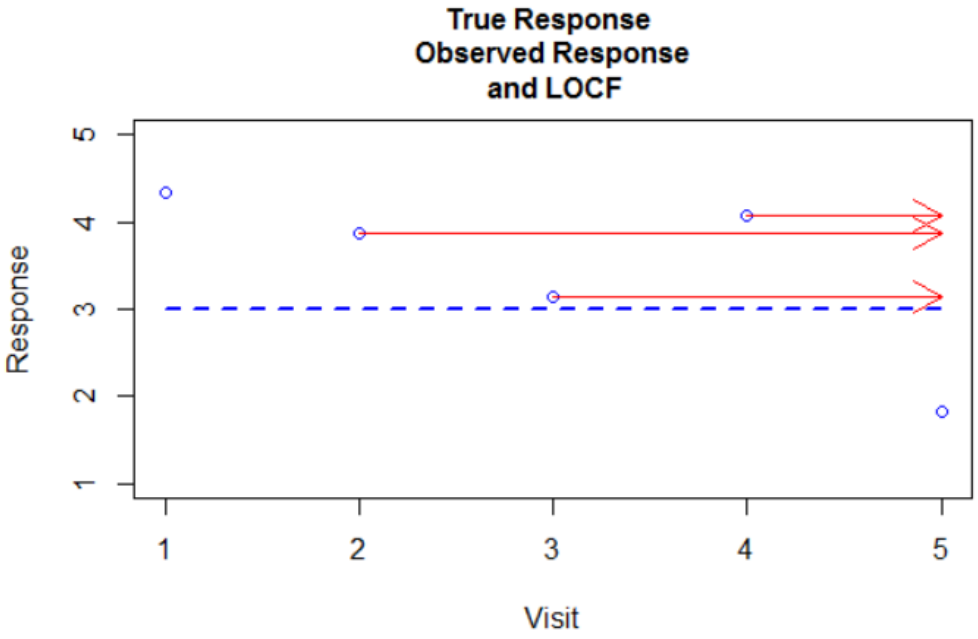
FRANK J MOLNAR, MALCOLM MAN-SON-HING, BRIAN HUTTON, DEAN A FERGUSSON

2. It's not principled

Principled = Take account of uncertainty by using probability models

NB: we record measurements not true values

Want to account for ***predictive distribution*** .



Estimated mean + variability of estimate + measurement error

MI does this through simulation

Likelihood methods do this implicitly under certain circumstances (MAR)

2. But…….

LOCF is not a valid imputation method

But….

It is a perfectly valid measurement of response at the last observation under treatment

If this makes sense for your estimand …….

3. Doesn't MAR make the least assumptions?

When in doubt add variables *(works for time to event data).*
Let R be a missing data indicator, $Y=(Y_O , Y_M)$ the data

Turns missing into partially observed multivariate data

$P[Y,R|X,\theta, \varphi]=P[Y|X,\theta]xP[R|Y,X,\varphi]$ <- **Selection Model**

- **MCAR** $P[R|Y,X,\varphi] = \mathbf{P[R| X,\varphi]}$ <- doesn't depend on observed responses
- **MAR** $P[R|Y,X,\varphi] = \mathbf{P[R| Y_O , X,\varphi]}$ <- depends on observed responses
- **MNAR** $P[R|Y,X,\varphi] = \mathbf{P[R| Y_O Y_M X,\varphi]}$ <- depends on missing responses

Need One More Concept – **Ignorability**

$$P[Y_O,R|\theta, \varphi]= \int P[Y_O , Y_M |X,\theta]P[R| Y_O , Y_M,X,\varphi]dY_M$$
$$= \int P[Y_O , Y_M |X,\theta]P[R| Y_O ,X,\varphi]dY_M$$ <- because of MAR
$$= P[Y_O |X,\theta]P[R| Y_O ,X,\varphi]$$ <- likelihood analysis observed data alone ok because of separation of parameters

3. Big Assumption

Data exists, or potentially exist and is unobserved

(what is treadmill walk time for a patient who dies ?)

3. Doesn’t MAR make the least assumptions?

So… under ignorability (MAR + separation) you can ignore the missing data problem and use your usual likelihood based analysis

Example text from a protocol or analysis plan
“ …*under the assumption the data are missing at random (MAR) a mixed model repeated measures analysis (MMRM) will provide an unbiased estimate of the treatment effect*.”

Assumptions

1. This analysis is estimating the correct estimand (*de jure*). That is an MMRM assumes everyone stays on protocol after withdrawal.
2. MAR assumes correct model with all needed covariates.
   - *It is a property of both the data and the model.*
3. Assuming separability of parameters.

3. Doesn’t MAR make the least assumptions?

What they mean:
I just told the programmer to run a standard analysis in SAS and ignored the missing data. I didn’t do anything out of the ordinary so I didn’t make any assumptions.

What they forgot:
The important part is the estimand and MAR assumption. The analysis provides a false sense of objectivity (not making assumptions) because Proc Mixed is easy to use. The same estimand and assumptions could have been used with a multiple imputation approach, which would seem a lot more ‘hands on’. The only real difference is the calculation of the results, not the underlying statistical or scientific assumptions and interpretation. In this case they may think they did a lot because there are multiply imputed datasets created.

Some analyses have more ‘degrees of freedom’ to specify, but all make equally important assumptions.

4. Is MI just for MAR?

- **No!**
- **$P[Y_O,R] = \int P[Y_O, Y_M, R]dY_M$**
  - $\int P[R]\, P[Y_O, Y_M, |R]dY_M$
  - $\int P[R]\, P[Y_O|R]\, \mathbf{P[Y_M|Y_O, R]}\, dY_M$
  - Can numerically integrate expression by taking random draws from the conditional predictive distribution and averaging.
  - This can be done under any assumptions regarding MCAR, MAR, MNAR etc.

MI makes multiple predictions/imputations of the missing data based on the observed data and assumptions. As long as there is a model for the observed data, we can use MI

4. Is MI just for MAR?

- **A very general approach is to use MI with pattern mixture models.**
- **Pattern Mixture Factorization**
  - P[Y,R]=P[Y|R]P[R]
  - R indicates pattern
    - Eg R = Completer, Lack of Efficacy, Adverse Event, Loss to Follow up etc. These patterns must be specified ahead of time
    - Distribution of R is trivial to estimate. Percentage of patients in each category.
    - P[Y|R] is description of the potential observed data in a pattern. Very easy to discuss with clinicians.

4. Is MI just for MAR?

- **Some examples**
  - If R={Adverse Event and Treated} then P[Y|R] can be modeled from the distribution of treated subjects at baseline
    - Principled form of BOCF
  - If R={Withdrawal and Treated} then P[Y|R] can be modeled from the distribution of placebo subjects at the same time points conditional on baseline covariates
    - MNAR type of analysis. Assumes patients do not receive any benefit from drug after withdrawal (Jump to Control)
  - If R={Withdrawal and Treated} then P[Y|R] can be modeled from a regression (linear, polynomial, spline) of treated group
    - MAR type of analysis. Assumes patients remain on treatment

5. How do I pick a sensitivity analysis?

- **Sensitivity analyses should be planned to assess the impact of missing data on the study results**
  - Merely running additional analyses that make the same missing data assumptions is not useful

- **For example, if a primary analysis assumes missing at random (e.g., MMRM), then a sensitivity analysis involving multiple imputation under the same assumption is uninformative**

Lisa M. LaVange, PhD
Office of Biostatistics
OTS/CDER/US FDA

5. How do I pick a sensitivity analysis?

- **Start with your primary analysis and perturb your assumptions to see how much that changes the results.**
- **Examples**
  - Jump to reference
  - Copy reference
  - Jump to treatment
  - Etc.
  - Delta adjustment and tipping point

Plausible Worst Case

How far can I perturb the imputations before I lose significance

6. What's wrong with starting with the label and working backwards?

Nothing
Actually it is a Great Idea

But …

6. What's wrong with starting with the label and working backwards?

Estimand approach is more detailed in many ways than a label

In particular

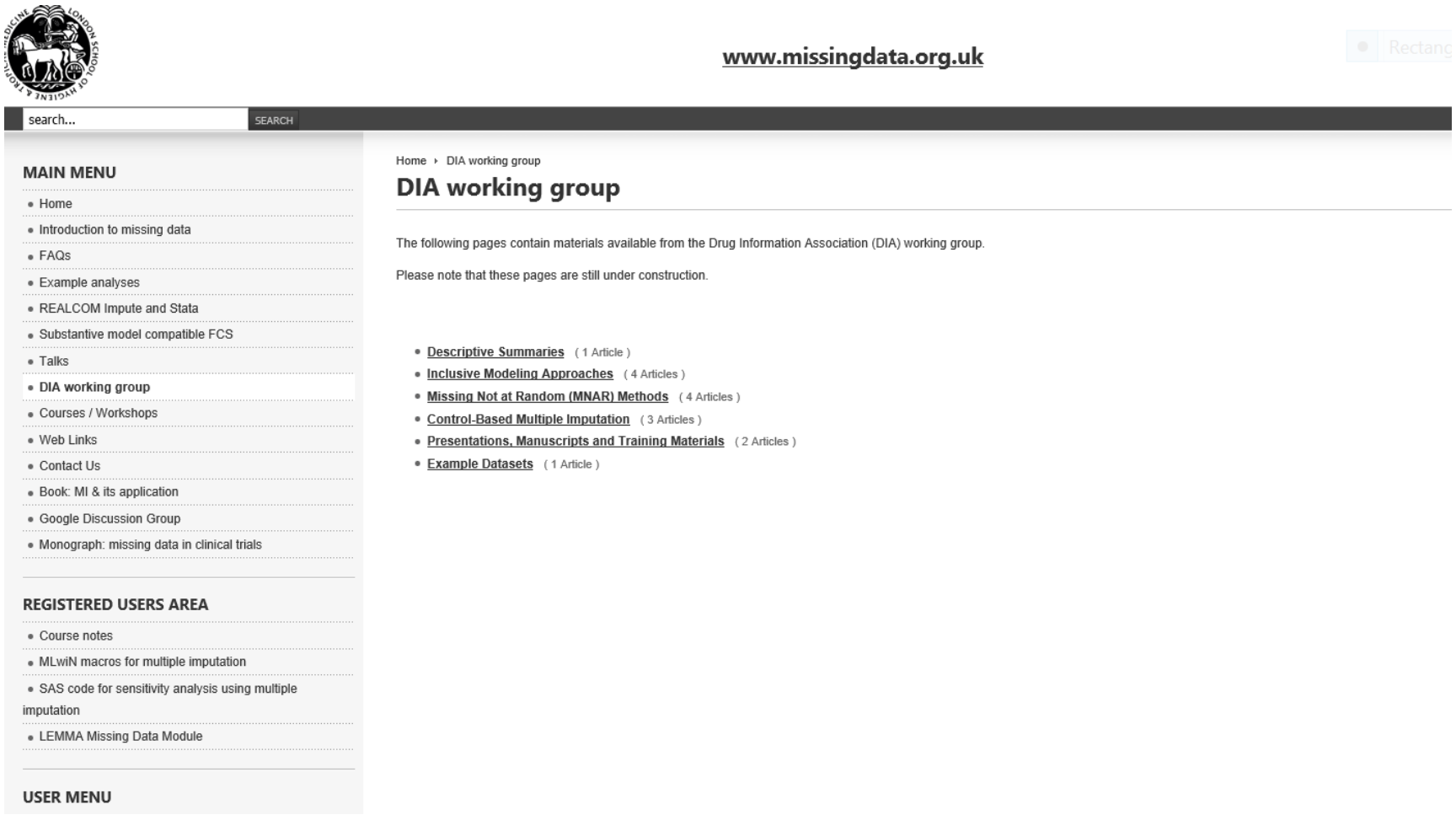
It specifies under what conditions the treatment effect will be measured

Danger is that regulatory groups may require an estimand that is primarily *de facto* , when past approvals were based on estimands that were *de jure*. The result is that you will likely have a smaller treatment effect to report to physicians who will not know the difference between the two. This can result in a new and better drug reporting results that are less impressive than an older competitor.

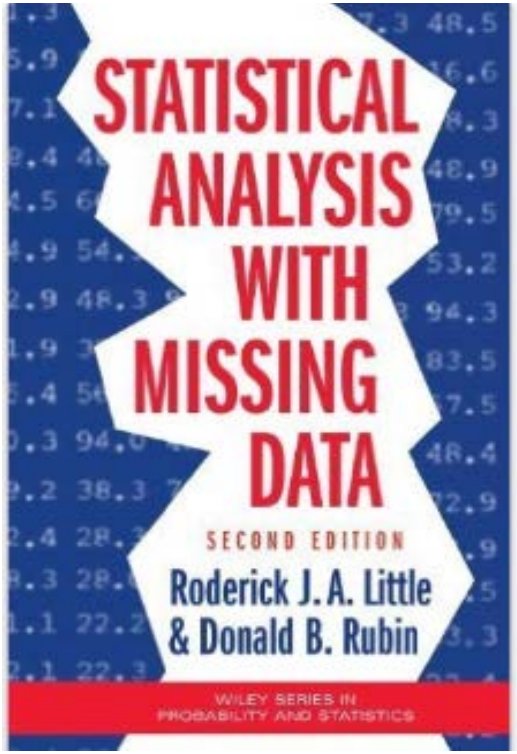
7. What's the most important thing to know about missing data?

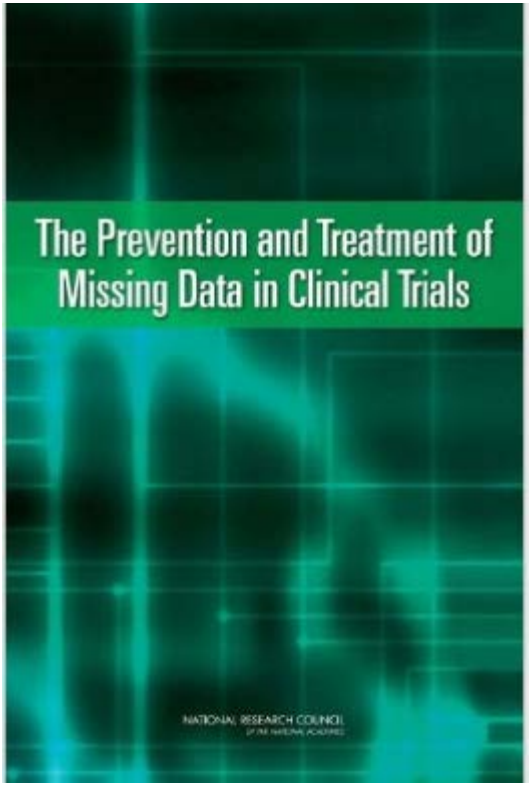
- ***Don't miss it in the first place***
  - Patient and clinician education
  - Enrichment
  - Reduce burden on patient
  - Rescue medication
  - Open label extensions for incentive
  - Designs that allow a higher probability of receiving treatment
- **Gather enough data to help with withdrawal if it happens**

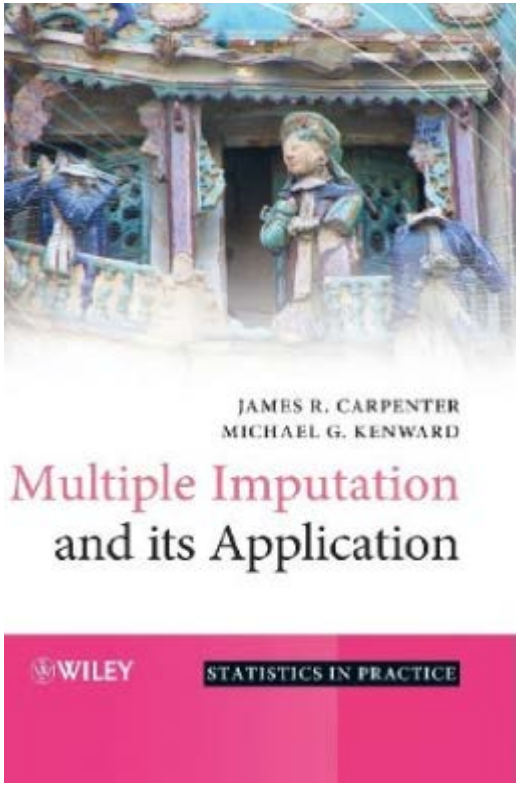
Resources


www.missingdata.org.uk
search...
SEARCH
MAIN MENU
Home
Introduction to missing data
FAQs
Example analyses
REALCOM Impute and Stata
Substantive model compatible FCS
Talks
DIA working group
Courses / Workshops
Web Links
Contact Us
Book: MI & its application
Google Discussion Group
Monograph: missing data in clinical trials
REGISTERED USERS AREA
Course notes
MLwiN macros for multiple imputation
SAS code for sensitivity analysis using multiple imputation
LEMMA Missing Data Module
USER MENU
Home ▸ DIA working group
DIA working group
The following pages contain materials available from the Drug Information Association (DIA) working group.
Please note that these pages are still under construction.
Descriptive Summaries (1 Article)
Inclusive Modeling Approaches (4 Articles)
Missing Not at Random (MNAR) Methods (4 Articles)
Control-Based Multiple Imputation (3 Articles)
Presentations, Manuscripts and Training Materials (2 Articles)
Example Datasets (1 Article)

Recommended Books







Recommended Books