Consideration in choosing objectives and estimands in clinical trials

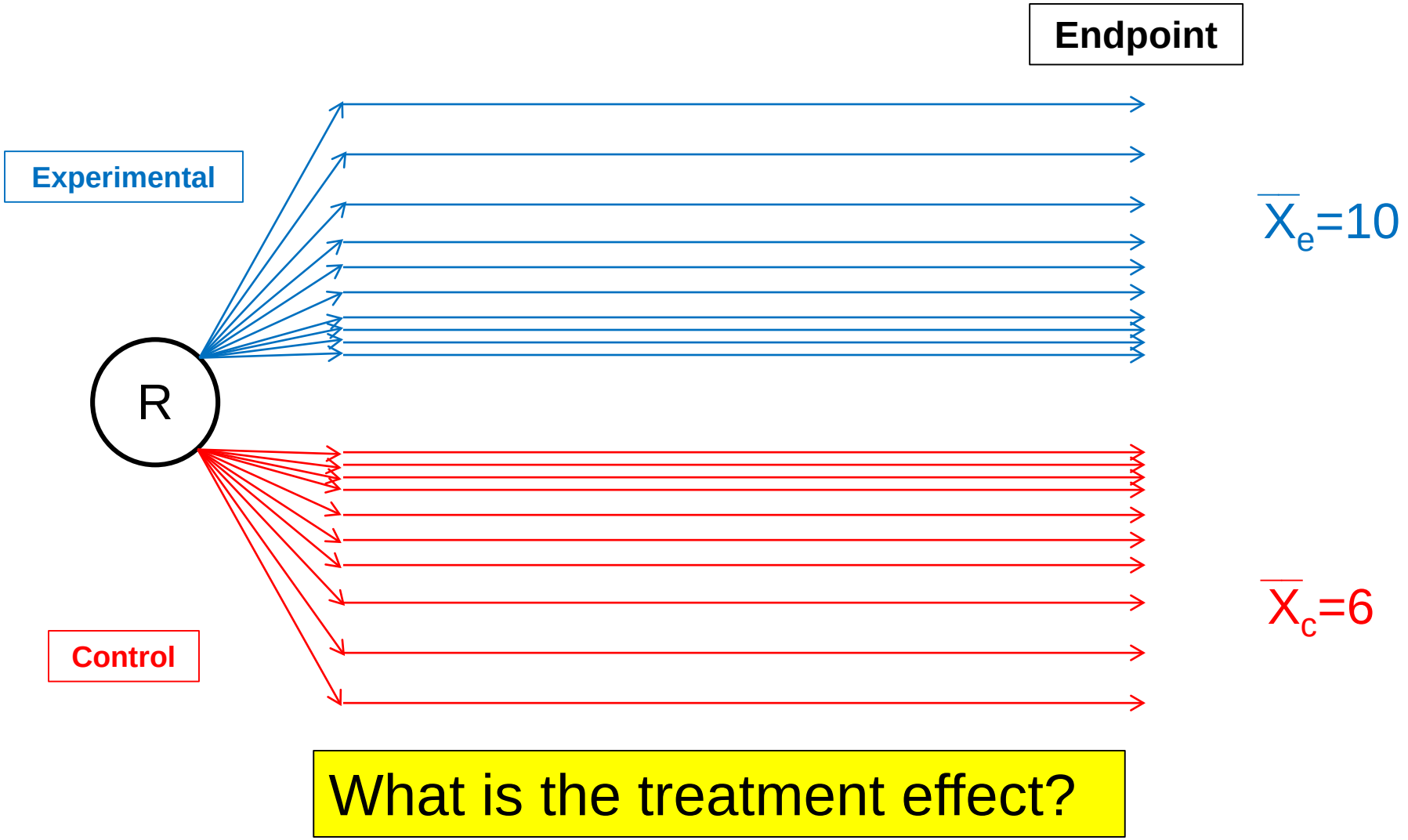
Craig Mallinckrodt

Lilly

Outline

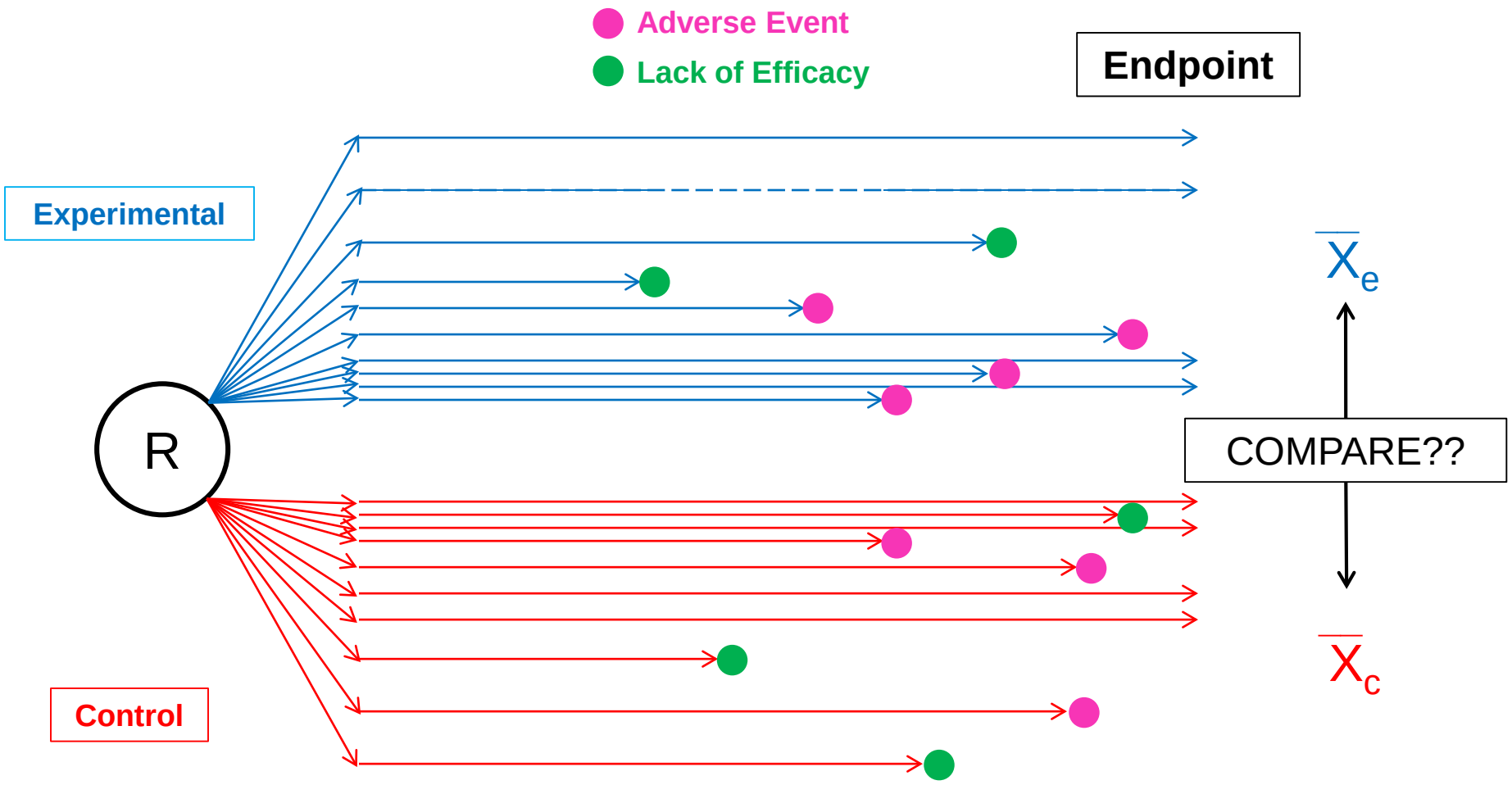
- **Objectives and estimands**
- **Estimators**
- **Sensitivity**

Randomized Clinical Trials



What is the treatment effect?

Randomized Clinical Trials



PROBLEM: Missing Data

Past Practice

- **Post-randomisation events dealt with implicitly by choices made about data collection and statistical analysis**
- **Analyses tended to be simple and ad hoc, a hold-over from the era when computing power limited options**
- **These choices defined the scientific question**
- **This practice needs to be reversed**
  - **Scientific method does not start with an analysis, from which we determine the hypothesis being tested. It starts with a question / hypothesis…**

Key Guidance

- **ICH E9 guidance (circa 1998)**
  - **does not directly address key issues**
  - **Many new developments since ICH E9 showing short-comings of common methods and attributes of newer, principled methods**
- **NRC expert panel report, 2010**
  - **135 pages, 18 recommendations**
- **ICH E9 R1 draft 2017**
  - **This presentation is an attempt to prepare for the new guidance**

Framework From NRC Guidance

- **3 Pillars**
  - **Set clear objectives & define causal estimands**
  - **Maximize adherence**
  - **Sensible primary analysis supported by plausible sensitivity analyses**
- **Discouraged simple ad hoc methods such as LOCF and BOCF**

Study Development Process Chart

- **Objectives**
  - **Decisions to be made drive objectives, which drives choice of estimands…**
- **Estimands**
- **Design**
- **Analysis**
- **Sensitivity**
- **Iterative process**

Garrett (2015) Phillips (2016) PSI / EPSI Working Group. Pharmaceutical Statistics.

Estimand – Definition

…..what is to be estimated to address the scientific questions (objectives) of interest.

Four components:

- **Population**
- **Endpoint**
- **Summary measure**
- ***How to account for inter-current events***

Inter-Current events

- **How to handle inter-current events is key to understanding the Intervention (treatment) effect**
- **Examples of inter-current eveents**
  - **In general, post-randomization events that may be related to treatment / outcome**
  - **Specific examples**
    - **Discontinuation of intervention +/- study**
    - **Addition of, or switching to rescue medication**
    - **Death**

Five Methods of Dealing with Inter-Current events

- **1) Treatment policy (ITT) – ignore inter-current events**
- **2) Composite - modified definition of the variable (or the summary measure) with inter-current event(s) a component of the outcome**
  - **NRI, mNRI, assign band rank to patients with inter-current event**
- **3) Hypothetical - specific hypothetical conditions of interest, e.g.**
  - **Outcome if no inter-current events (MMRM, MI)**
  - **Outcome if patients could be followed without treatment (reference based controlled imputations.**

Five Methods of Dealing with Inter-Current events (continued)

- **4) Principal strata - restrict population of interest to the stratum of patients in which an inter-current event would not have happened.**
- **While on treatment - values of the variable in those patients up to the time of the inter-current event in all patients**

Why Estimands are Important

- **Link between objectives and analysis**
- **Needed regardless of missing data**
- **But missing data adds complexity**
- **The definition and proper handling of missing data depend on the estimand of interest**

Different Decisions & Perspectives

Stakeholders	Types of Clinical Trials
• Regulators • Payers • Physicians • Patients • Sponsors	• Exploratory vs. confirmatory vs. post-approval • Short-term vs. long-term treatment • Symptomatic treatment vs. disease modification • Efficacy vs. safety • In-patient vs out-patient

General Categories of Objectives

- Compare treatment A vs treatment B
- Compare treatment policy A vs policy B
  - Begin with treatment A vs begin with treatment B
  - Treatment A + rescue vs Treatment B + rescue
- In causal inference “treatment policy” typically implies precisely defined treatment algorithms. Treatment X + rescue is probably too vague

General Categories of Estimands

- **Efficacy**
  - **Benefit of the drug when taken as directed**
- **Effectiveness**
  - **Benefit of the drug as actually taken**
  - **Conceptually, a composite of efficacy and adherence**
- **More general categorization (safety outcomes)**
  - **De-jure: When taken as directed**
  - **De-facto: As actually taken**

Intention to Treat

- **Primary focus of ITT in ICH E9 was on which patients to include, not as a means of dealing with missing data**
  - **Including post-rescue data does reduce the number of missing values**
- **ICH E10 states that need for rescue is an endpoint**
- **Today's more nuanced discussion of estimands compelled an update to E9**
- **That is a sign of significant progress!!!**

Rescue Medication Considerations

- **Post-rescue data in an ITT analysis can mask or exaggerate effect of originally assigned med**
  - **Post rescue data not included for treatment objectives**
  - **When data after rescue are included inference is on treatment policy / regimen**
- **Availability of rescue should not influence adherence to initial treatments - but this is a concern in placebo controlled / blinded trials**
  - **On blinded med X% chance on placebo**
  - **On rescue med 0% chance on placebo**

Example Estimands

Each based on difference between drug & control in mean change from baseline to time X

- **1 (ITT, treatment regimen): in all patients, regardless of adherence or use of rescue (as actually taken)**
- **2: (treatment effectiveness) in all patients assuming a bad outcome if non-adherent or rescued (as actually taken)**
- **3: (treatment efficacy) in all subjects assuming they remained adherent to randomized treatment and did not initiate rescue therapy (if adherent)**
  - **What to expect if hadn't switched / stopped**

Data Considerations

- **De-facto Treatment regimen** (estimand 1)
  - **Data post rescue / discontinuation included**
  - **“Pragmatic effectiveness”**
- **De-facto Initial treatment** (estimand 2)
  - **Data post rescue / discontinuation not included**
  - **Post-rescue / discontinuation data imputed or rescue / discontinuation = bad outcome**
- **De-jure Initial treatment** (estimand 3)
  - **Data after rescue / discont. not included**

Fundamental Considerations

- **De-jure estimands**
  - **What to expect if patient hadn't stopped / switched**
  - **Counterfactual for group; assess as if all patients adhere when in fact some do not**
  - **Valid estimate of what to expect if patients adhere – the majority**
  - **In order to give proper directions, must asses what happens if taken as directed**
  - **Regulators generally do not accept as primary**

Pharmaceutical Statistics; DOI: 10.1002/pst.1765

Fundamental Considerations

- **De-facto estimands**
  - **Counterfactual for individual patients**
  - **Mixture of adherent and non-adherent – each patient is one OR the other, not a mix**
  - **Valid estimate of what to expect for the group**

- **Strengths and limitations for each category**

Pharmaceutical Statistics; DOI: 10.1002/pst.1765

Design Considerations

- **For de-jure estimands, maximizing adherence**
  - **Improves sensitivity – reduces probability plausible departures from MAR overturn result**
  - **Does not influence parameter values**
- **For de-facto estimands, maximizing adherence**
  - **Influences parameter values**
    - **With NRI, If dropout reduced by design, fewer fail & treatment is more effective**
  - **If means to maximize adherence in trial are not feasible in practice, generalizability of results may suffer**

Pharmaceutical Statistics; DOI: 10.1002/pst.1765

Analysis Considerations

- **Historically, analyses tended to be simple and ad hoc**
  - **LOCF, BOCF, NRI**
  - **Hold-over from era when computing power limited options**
- **NRC said don’t use simple and ad hoc**
- **More principled alternatives now easy to implement**

Analysis Considerations: Estimand 2

- **Unifying principle: if patients don't adhere they don't benefit**
- **Implicitly assumes adherence decisions approximate clinical practice**
- **Key is how to determine what is zero benefit**
- **Dropout = failure: NRI, mNRI, BOCF**
  - **No missing data**
  - **Assumes no spontaneous improvement**
- **Controlled imputation approaches**
  - **Use placebo as definition for no benefit**

Analyses Considerations: Estimands 1 and 3

- **Direct likelihood (MMRM), multiple imputation (MI), or weighted generalized estimated equations (wGEE)**
- **Data post rescue / switch included for estimand 1, not included for estimand 3**
- **Estimand 1 will always have less missing data than estimand 3, but is it relevant**

Analysis Considerations: Estimands 1 and 3

- **MI and MMRM have same theoretical underpinnings, choice depends on non-missing data circumstances**
- **Uses observed data to predict what missing values would have been – what would have happened if patient did not stop or switch**
- **Key assumption, if observed the missing values would have had same “statistical behavior” as observed data**

Need for Multiple Estimands: Example

- **Phase 3 results of a drug support registrations to 4 regulatory authorities – who have 3 different views on primary estimands**
- **Paraphrase from a prominent HTA**
  - **Using confirmatory thinking can lead to difficulties. [We] want to see clinical benefit from several perspectives together where no adjustment for multiple testing is required**
  - **Sponsors should aim to avoid treatment switching in studies**

Example: Practical Benefit of Multiple Estimands

- **Effectiveness = Efficacy + adherence**
- **Two drugs can have = effectiveness but very different efficacy and adherence**

	Efficacy	Adherence	Effectiveness
1	High	Low	Average
2	Low	High	Average

- **Patients and prescribers need to know the difference in clinical profiles to tailor treatment**
  - **1 for more severe**
  - **2 for more sensitive**

Practical Benefit (2)

- **De-facto estimand - high dose = low dose**
  - **Inference: no benefit from high dose**
- **De-jure estimand - high dose > low dose**
- **Inference: high dose had greater efficacy but that advantage was negated by more early discontinuations**
- **Investigate subgroups**
  - **High dose for non-responders to low dose**
- **Investigate alternate dosing regimens**
  - **Titration, flex, split dosing, formulation**

Example: Including vs. not-including rescue data

- Psoriatic arthritis
- 24-wk, double blind, 1:1:1:1 randomization
- Pbo, SoC, 2 doses of experimental drug
- N ~ 100 arm
- ACR 20 (binary) @ wk-24 primary outcome
- Rescue available after wk 16
- On SoC 3/12 rescued met ACR20 @ wk 24
- On Pbo 13/45 rescued met ACR20 @ wk 24

Example Results

	Adalimumab ($n = 101$)	Placebo ($n = 106$)	Difference
Without post-rescue data (estimand 2)	58 (57.4%)	32 (30.2%)	27.2 %
With post-rescue data (estimand 1)	61 (60.4%)	45 (42.5%)	17.9 %

- **If rescue success on placebo had been 17/45 significance would have been lost**
- **For 90% power estimand 1 requires 175 / arm, more than 2x the placebo exposure compared with estimand 2, which requires 75 / arm**

General Principles / Approaches

- **Objectives**
  - **Pre-approval, Trtmnt objectives often more relevant than Trtmnt policy. Post-approval, Trtmnt policy increases in relevance**
  - **Trtmnt policy objectives common for hard endpoints due to ethical need for rescue before the endpoint**
  - **Trtmnt objectives more relevant for symptomatic treatments. Trtmnt policy objectives important for disease modifying drugs because effect persists after discontinuation**

General Principles / Approaches

- **Estimands**
  - **Diverse stake holders**
    - **Specifying a primary estimand is essential**
    - **Often Important to assess multiple estimands in a trial**
  - **Greater focus on de-jure (efficacy) early (Ph2), shifting to de-facto (effectiveness) later**
  - **De-jure estimands are useful – for efficacy and safety**

Outline

- Objectives and estimands
- Estimators
- Sensitivity

Primary Analyses

- **Dropout is outcome, complete data, but…**
  - **Treatment success (dropout = failure)**
  - **Trimmed mean** (Permutt Pharmaceutical Statistics i2017)
  - **Impute “placebo like” outcome**
- **Dropout yields missing data**
  - **Often MAR-based: Direct likelihood (MMRM), MI, wGEE)**
  - **Data post rescue / switch included for estimand 1, not included for estimand 3**

Missing Data Mechanisms

MCAR - missing completely at random

- **Conditional on the independent variables in the model, neither observed or unobserved outcomes of the dependent variable explain dropout**
- **At a given time point, dependent variable not different for dropouts vs. completers**

Missing Data Mechanisms

MAR - missing at random

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

- **Conditional on the covariates and observed outcomes, the statistical behavior of the unobserved data is what it would have been if observed**

Missing Data Mechanisms

MNAR - missing not at random, non-ignorable

- **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**
- **Conditional on the covariates and observed outcomes, the statistical behavior of the unobserved data is *not* equal to that if it had been observed**

Consequences

- **Missing data mechanism is a characteristic of the data AND the model**
- **Differential dropout by treatment indicates covariate dependence, not mechanism**
- **Mechanism can vary from one outcome to another in the same dataset**
- **Can never definitively distinguish MAR vs MNAR**

Missing Data in Clinical Trials

- **Efficacy outcomes are seldom MCAR because the observed outcomes typically influence dropout (DC for lack of efficacy)**
- **Trials are designed to observe all the relevant information, which minimizes MNAR data**
- **Hence in the highly controlled scenario of longitudinal confirmatory trials, missing data *may* be mostly MAR**

Modeling Assumptions

- **MAR *always* more plausible than MCAR because MCAR is a subset of MAR**
  - **MAR methods will be valid in every case where MCAR methods are valid**
  - **MCAR methods will not be valid in every scenario where MAR methods are valid**
- **MNAR can never be ruled out**

Modeling Conundrum

- **Can’t assume MCAR**
- **We don’t have the missing data about which the assumptions are made, Therefore…**
- **Validity of MAR can’t be verified**
- **Key assumptions in MNAR models can’t be verified**

General Guidance: Missing Data

- **Develop a *sensible* analysis and assess sensitivity to assumptions – not sensitivity of the method**
- **Sensible**
  - **No bias. All the variation in the estimated effect is random**
  - **Consistent**
  - **SE accurately reflects variability in estimate**

General Guidance: Missing Data

- **Can do better than MAR only via assumptions**
- **Strive for validity of MAR**
- **Implement MAR primary**
- **Use sensitivity analyses to assess the degree to which departures from MAR influence conclusions**

MAR Methods

- **Direct likelihood**
  - **Missing data can be ignored if MAR**
- **Multiple imputation**
  - **Explicitly models missing data**
- **Inverse probability weighting**
  - **Explicitly models dropout and uses inverse probability weights in a subsequent analysis**

Pro and Con: Direct likelihood

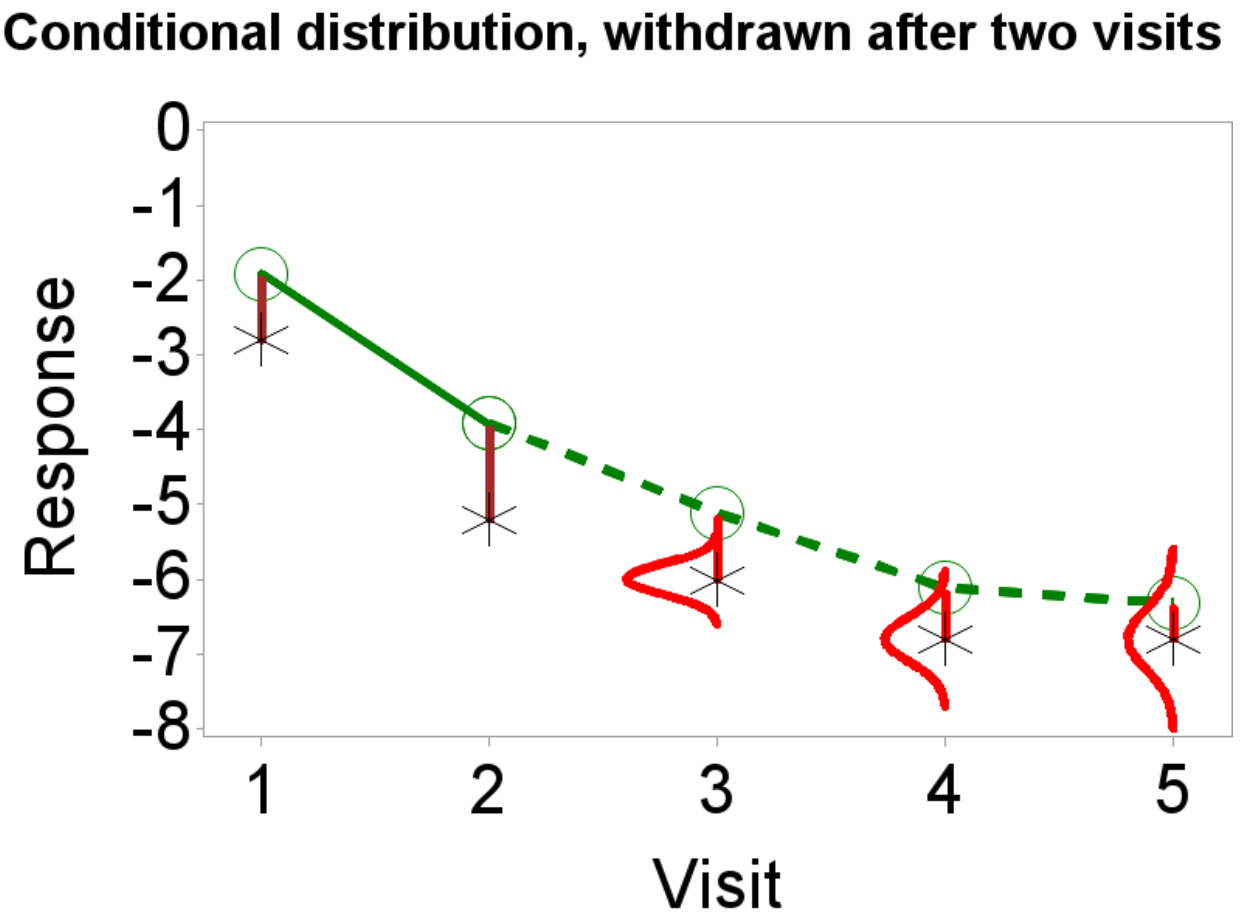
- **Simplest**
- **Least flexible**
- **Good choice with restrictive models**

Pro and Con: MI

- **Most flexible and intuitive**
- **A (tiny) bit more cumbersome**
- **Asymptotically converges to DL result as N and M increase – if models are the same**
- **Good choice for inclusive models or when covariates are missing**

Situations where MI is especially useful

- **Missing baseline covariates: direct likelihood sub-optimal because all subjects with missing covariate are discarded**
  - **Use MI to impute baseline covariates**
- **Direct likelihood difficult to apply to categorical data**
  - **Use MI to impute**



Green line shows the means A_{JK} (circles), dotted after withdrawal.
Brown residuals are for two observed values (star) before withdrawal.
Red “residuals” show location of means (star) for conditional distribution.
Red Normal curve indicates actual conditional distribution

MI with nonotone missing data

- **Impute** missing data (typically) using Bayesian predictive distributions, conditional on observed data, resulting in multiple (*m*) completed data sets
- **Analyze** the *m* completed data sets using an analysis that would have been appropriate for complete data, resulting in *m* estimates
- **Combine** the *m* estimates into a single inferential statement by using combination rules (or "Rubin's rules") that account for uncertainty due to imputation of the missing values, therefore providing valid inference

MI for non-monotone missing data

- **The three basic steps assume monotone missing data**
- **Intermittent data are conceptually easier (MAR), but logistically are more complex**
- **Two approaches**
- **use “intermittent approach” for all missing values**
- **Fill in intermittent then use monotone approach – if intermittent rare**

General Considerations for MNAR

- **No definitive MNAR model**
- **Every MNAR model has an MAR "Bodyguard"**
  - **Equal fit to the data, but different "missingness models"**
- **There is never any data to support the MNAR part of a model**

Outline

- Objectives and estimands
- Estimators
- Sensitivity

General Approaches for Assessing Sensitivity to Departures From MAR

- **Compare results from multiple (MNAR) models**
  - **Inferences difficult because results may differ because both models wrong, 1 wrong, chance differences** (Statist Med. DOI: 10.1002/sim.6753
- **Add a sensitivity component or parameter(s) to the primary analysis** Ther Innov & Reg Sci 48(1): 68-80.
  - **Vary sensitivity (MNAR) parameter(s) within the primary analysis model**
  - **Tipping point and plausible worst case approaches**

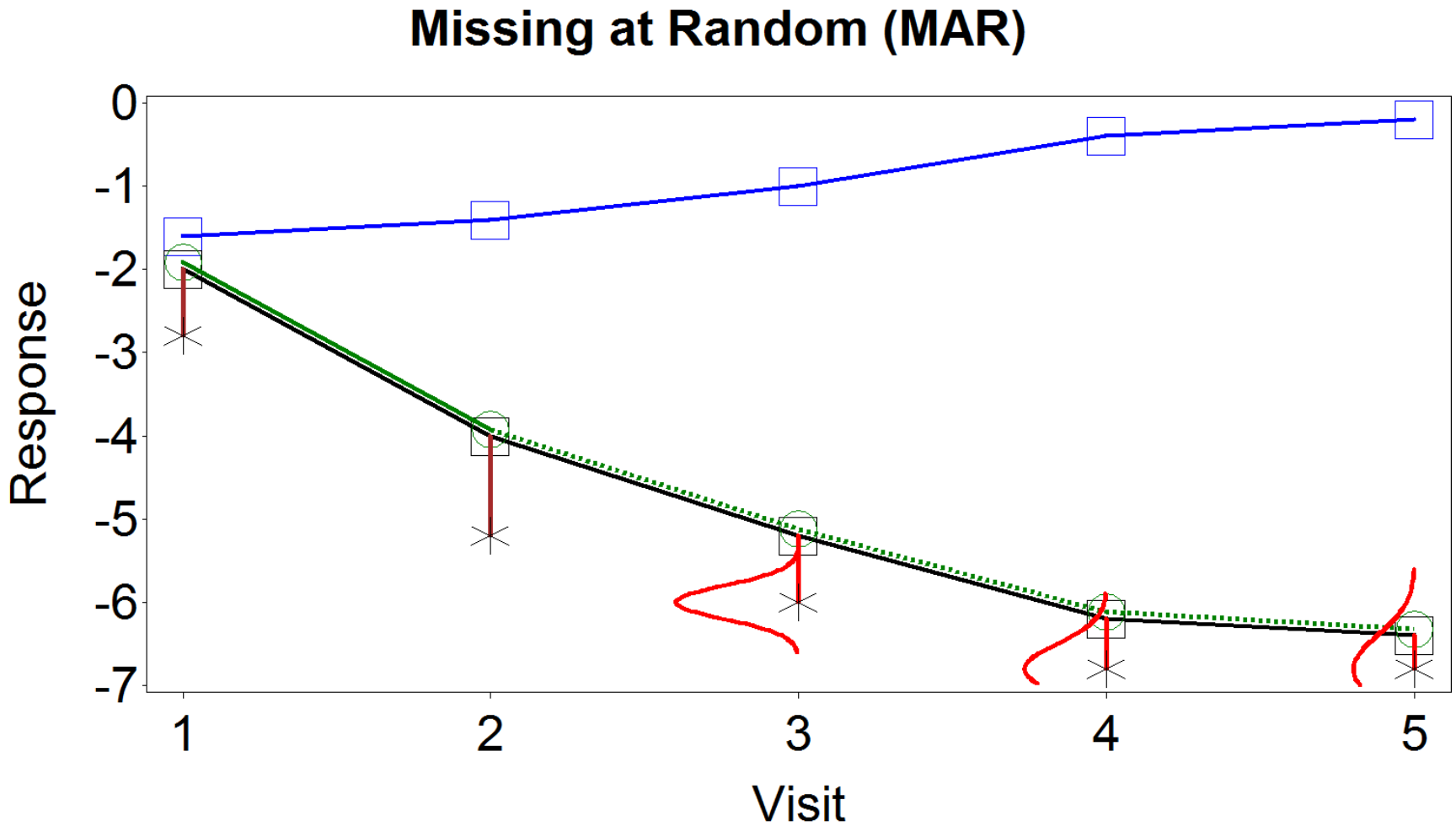
“Controlled Imputation” Family

- **MI and likelihood-based approaches**
- **Assumptions can be transparent and debated**
- **General idea is to create departures from MAR**
  - **Reference based – Plausible worst case**
    - **Jump to reference, copy reference, copy increment from reference** J Bio pharm Stat 23:1352-1371
  - **Delta adjustment – Tipping point or plausible worst case** Clinical Trials with Missing Data. (2014). Wiley, Chichester
    - **Conditional (sequential)**
    - **Marginal**

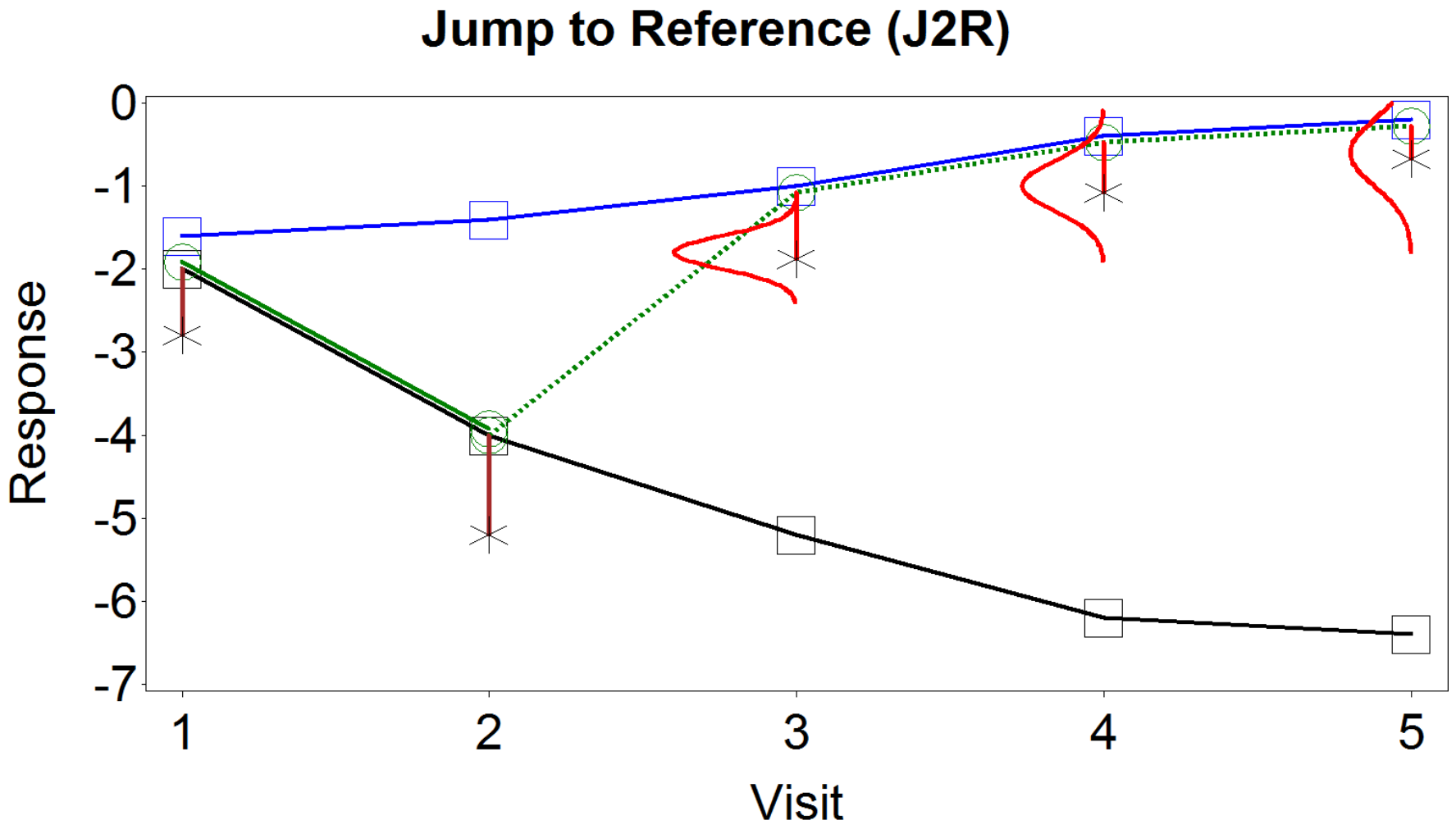
Controlled Imputations: Reference Based

- **Jump to reference**
  - **The statistical behavior of drug treated patients after dropout immediately becomes that of reference patients" (E.G., placebo)**
    - **Use for drugs with short on target half-life**
- **Copy reference**
  - **…gradually transitions to placebo**
    - **Use for drugs with long on target half life**
- **Copy increment**
  - **After dropout, change for drug = change for placebo**
    - **Use for disease modifying drugs**

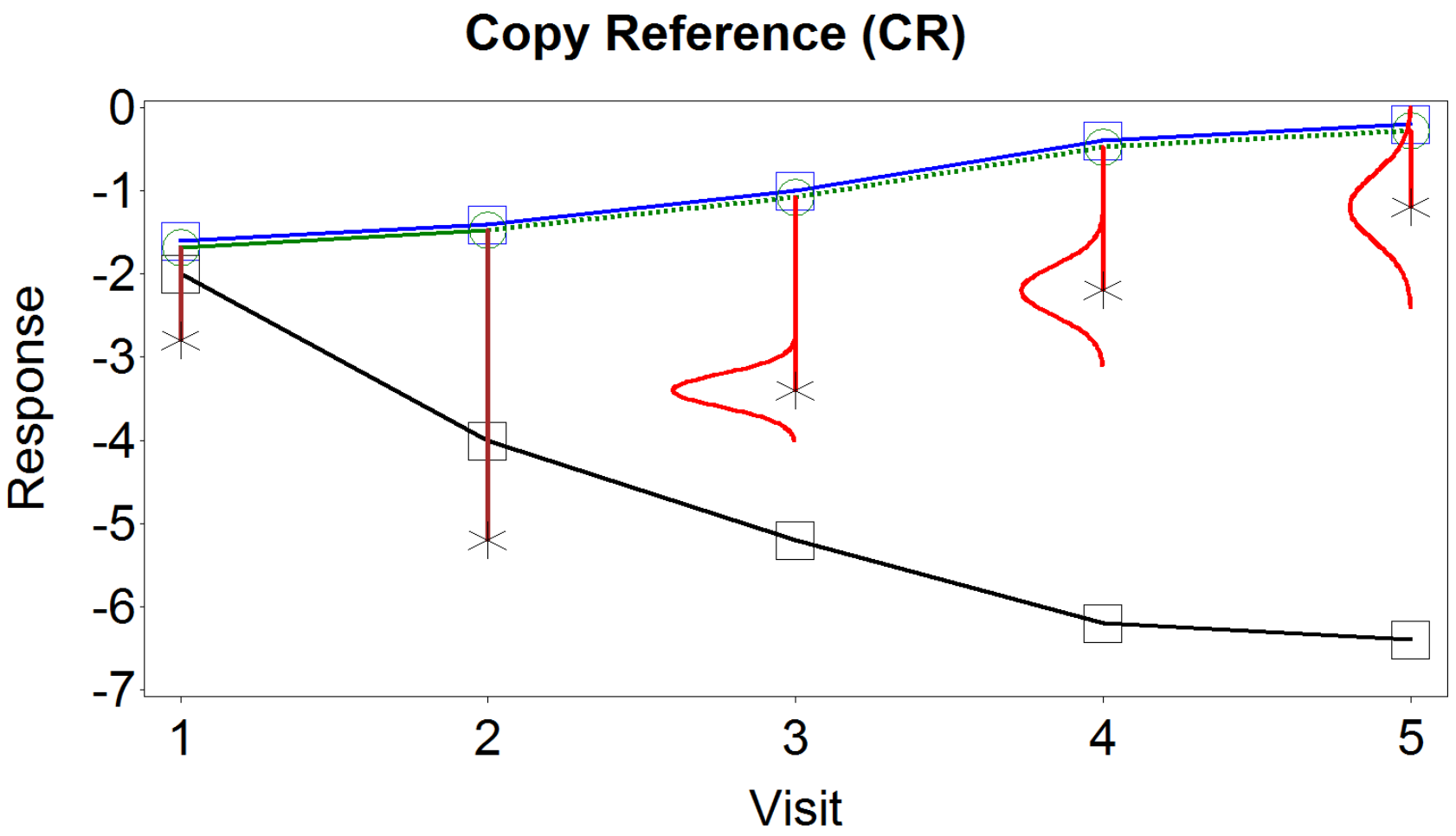
Withdrawal after two visits: Black line Active. [Terms A_{jk}] Blue line Reference. [Terms A_{jk}]. Dotted Green line (imputation). [Terms B_{pjk}]



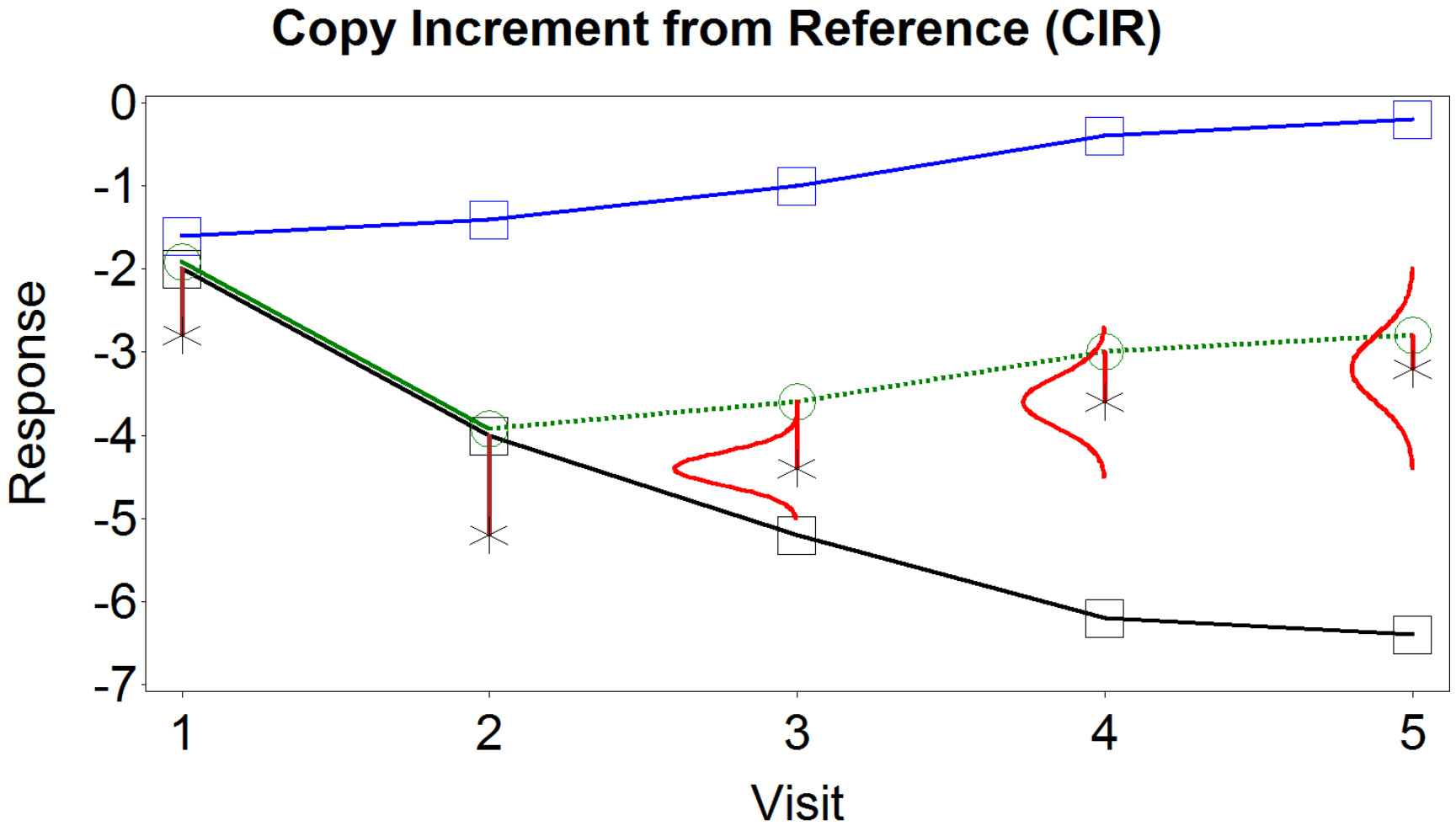
Jump to reference: Regression on residuals – based on pre-withdrawal deviations from mean of assigned arm



Copy Reference (CR): Based on pre-withdrawal deviations from reference arm mean



Copy Increment from Reference



Interpretations

- **Effectiveness context**
  - **Assumes benefit diminishes / disappears**
  - **Uses placebo group to define “no benefit, thereby accounting for study effect & placebo effect**
  - **Valid if patients improve or worsen**
  - **Free of confounds in follow up data**
- **Efficacy context**
  - **Worst reasonable case MNAR**
- **Standard software, standard tests**

Delta Adjustment Methods

- **Conditional (sequential, visit-by-visit)**
  - **Subtract a constant (delta) from visit X imputed value that then further influences imputed values at visit > X**
    - **First missing visit only** **(diminishing effect)**
    - **All missing visits** **(accumulating effect)**
- **Marginal**
  - **Complete all imputations then add delta** **(constant effect)**

Delta Adjustment Frameworks

- **Plausible worst case**
  - **Choose a meaningful delta (e.g., average treatment effect)**
  - **If results significant after delta adjustment, conclude results are robust**
- **Tipping point**
  - **Progressively increase delta until primary analysis is overturned**
  - **If value required to overturn significance is not plausible results are robust**

Example - Depression

- **Two real but contrived data sets (n=100/arm)**
  Therapeutic Innovation and Regulatory Science; 2014, 48(1): 68-80.
- **Drug arm patients randomly selected from 3 active arms**
- **Placebo arms mostly as is (with minor replication)**
- **Nearly identical designs**
  - **8-week, double blind, randomized 1:1:1:1**
  - **Assessments @ weeks 1,2,4,6,8**
  - **Similar inclusion / exclusion**
  - **Low dropout from EU study with ext and titration**
  - **High dropout from US fixed dose, no extension**

Completion Rates

	Placebo	Drug
High dropout data set	**60%**	**70%**
Low dropout data set	**92%**	**92%**

Primary Objective / Estimand / Endpoint / Analysis / Sensitivity

- **Treatment Objective**
- **Efficacy (de-jure) Estimand 3**
- **Mean change from baseline, contrast at week 8, 17 item Hamilton depression scale total score**
- **Direct likelihood primary – MMRM (MAR)**
- **Delta adjustment sensitivity** (marginal)

De-jure Estimand (3) – Mean change

Data	LSMEAN Change Drug	Placebo	Diff	SE	P value
High	8.24	5.94	2.29	1.00	0.024
Low	12.32	10.50	1.82	0.70	0.010

Delta Adjustment Results – Marginal Delta

	Low Dropout			*High dropout*		
Value of Delta	Endpoint Contrast	Std Error	P value	Endpoint Contrast	Std Error	P value
0	1.86	0.70	0.008	2.27	1.12	0.042
1	1.79	0.70	0.011	1.96	1.13	0.083
2	1.71	0.70	0.015	1.64	1.14	0.151
3	1.64	0.71	0.020			
4	1.57	0.71	0.027			
5	1.50	0.72	0.037			
6	1.42	0.72	0.049			
7	1.35	0.73	0.065			

Effect of delta proportional to fraction missing

- Change per unit delta ~ 0.08 for low dropout
- Change per unit delta ~ 0.30 for high dropout

Reference-Based Imputation Results

	LSMEANS		LSMEAN	Std	
	Placebo	Drug	Difference[1]	Error	P value
High dropout					
MAR	-5.95	-8.24	2.29	1.00	0.024
J2R	-5.97	-7.57	1.60	0.99	0.110
CR	-5.96	-7.71	1.75	0.98	0.075
CIR	-5.95	-7.78	1.83	0.97	0.004
Low dropout					
MAR	-10.56	-12.40	1.84	0.70	0.009
J2R	-10.55	-12.26	1.71	0.70	0.016
CR	-10.55	-12.27	1.72	0.70	0.015
CIR	-10.55	-12.27	1.72	0.70	0.015

Difference MAR vs. J2R over 6x greater in high dropout