MISSING DATA HANDLING IN BIOSIMILAR EVALUATION AND TIPPING POINT ANALYSIS

STEVEN HUA, BO JIN AND SHANMEI LIAO

BASS 2016

ROCKVILLE, MD

OUTLINE

- Clinical study design for biosimilar development
- Methods used for imputation
- Tipping point (TP) method
  - Study back ground
  - Introduction of simple TP method
  - Simulation
  - TP method based on multiple imputation (MI)
- Discussion/Conclusion

CLINICAL STUDY DESIGN FOR BIOSIMILAR

- Double-blind parallel design with biosimilar products and the reference product.
- Usually endpoints are either continuous or binary.
- Comparing of primary endpoints from two arms could be assessed by difference, ratio and sometimes, odds ratio.
- Determination of equivalence margin involves literature research, meta-analysis and regulatory discussion.
- Success is usually defined as 95% (90%) confidence interval (CI) of the difference or ratio being within the pre-defined margin.
- Missing data in the primary endpoint is avoidable but critical.

HANDLING OF MISSING DATA

- Complete case analysis
- Single imputation
  - Continuous endpoint: LOCF, BOCF, WOCF, etc.
  - Binary endpoint: e.g. non-responder imputation (NRI), for those with missing data at the primary timepoint, imputed as non-responder
- Multiple imputation
  - Multiple Imputation (simulation based)
  - Maximum Likelihood (model based)
- Other advance methods
  - Bayesian iterative simulation methods

TO BE CONSERVATIVE IN BIOSIMILAR EVALUATION

- Use binary endpoints as an example
- Unlike in superiority studies, NRI is not always conservative in “biosimilarity” evaluation
  - In the demonstration of biosimilarity in phase 3, if you have similar amount of missing data in two arms, NRI approach will make the response rates of the two arms even closer.

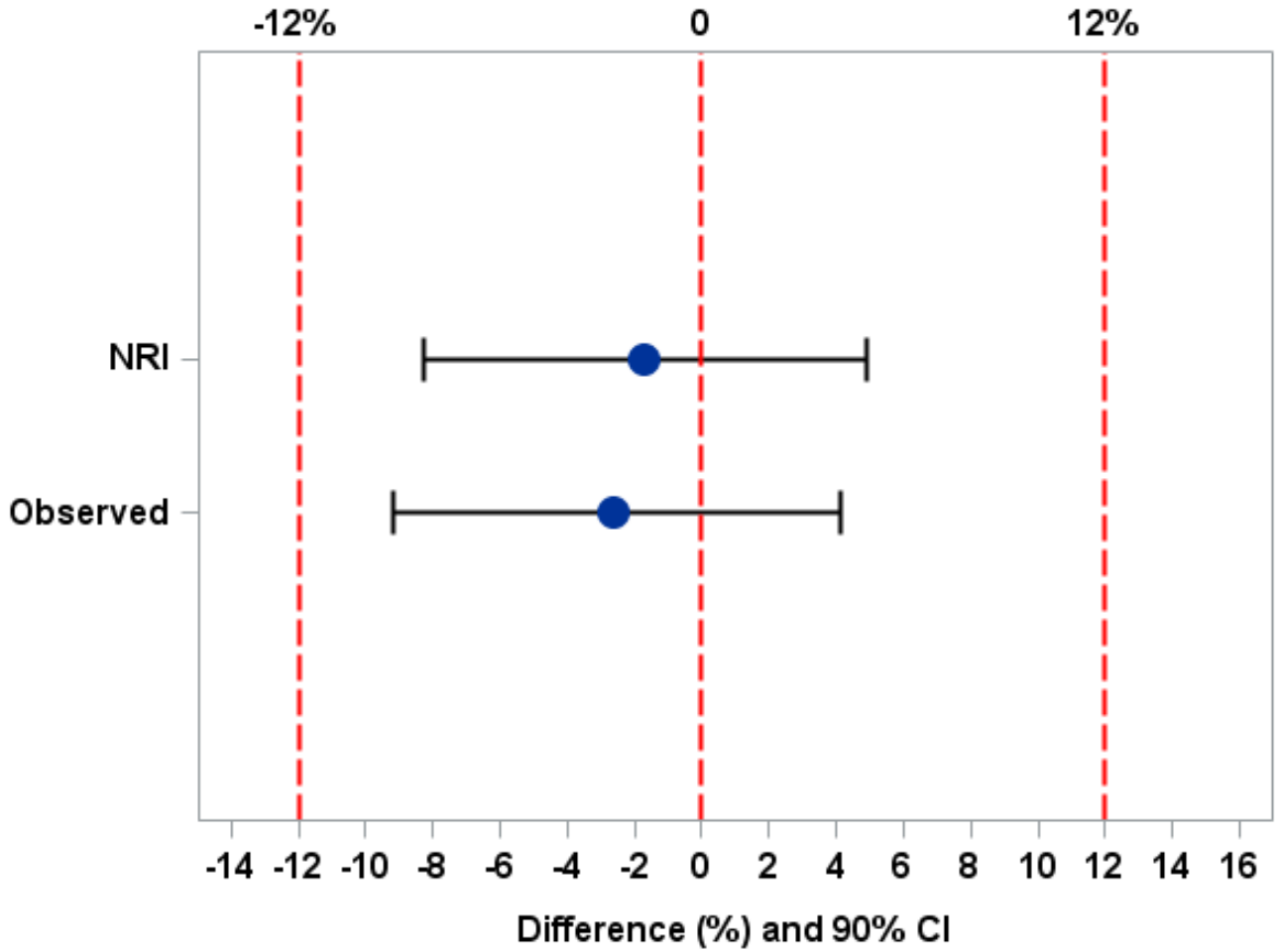
CASE STUDY

	N obs	N non-missing	Response obs	N missing	Obs rate*	NRI rate**
The Biosimilar arm	300	252	175	48	69.4%	58.3%
The Reference arm	300	250	180	50	72.0%	60.0%
Difference and exact 90% CI					-2.6% (-9.2%, 4.1%)	-1.7% (-8.3%, 4.9%)

*: obs rate=response obs/N non-missing

**: NRI rate=response obs/N obs

CASE STUDY


-12%
0
12%
NRI
Observed
-14 -12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12 14 16
Difference (%) and 90% CI

NRI MAY BE ANTI-CONSERVATIVE

- Although the two sets of analyses (NRI and observed) are consistent, NRI doesn't seem to be "sensitive" due to:
  - almost identical # of missing data between the 2 arms (48 vs. 50)
  - with similarity achieved for observed data, by assigning non-responder to all missing data in both arms, it does not impact the outcome

TP METHOD

- We need a more flexible way to evaluate the imputation's impact to the study results.
- TP method is an alternative strategy for handling missing data (Yan, et al, 2009).

The Tipping Point is "the moment of critical mass, the threshold, the boiling point"

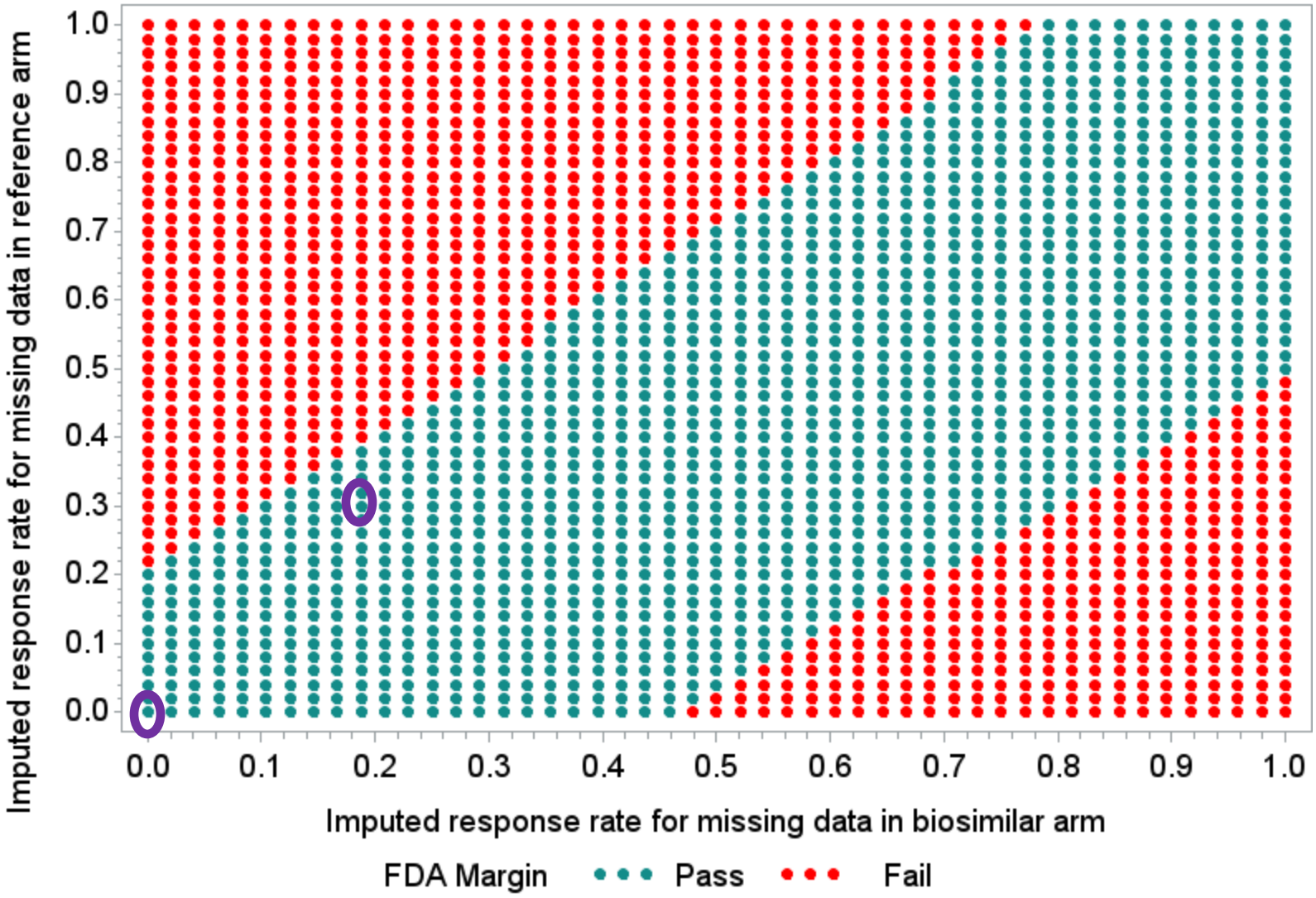
--- Malcolm Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*

- We define a TP as the difference of the number of responders between the treatment arms in the missing cohort at which the study conclusion is **changed**.

TP METHOD

- Unlike NRI, a TP analysis replaces the missing value with either responder or non-responder so that treatment comparison, or level of similarity would vary.
- It removes the concern we have for a simple NRI approach and also provides easiness in interpretation and flexibility in similarity assessment.

Tipping point analysis of difference in response rate

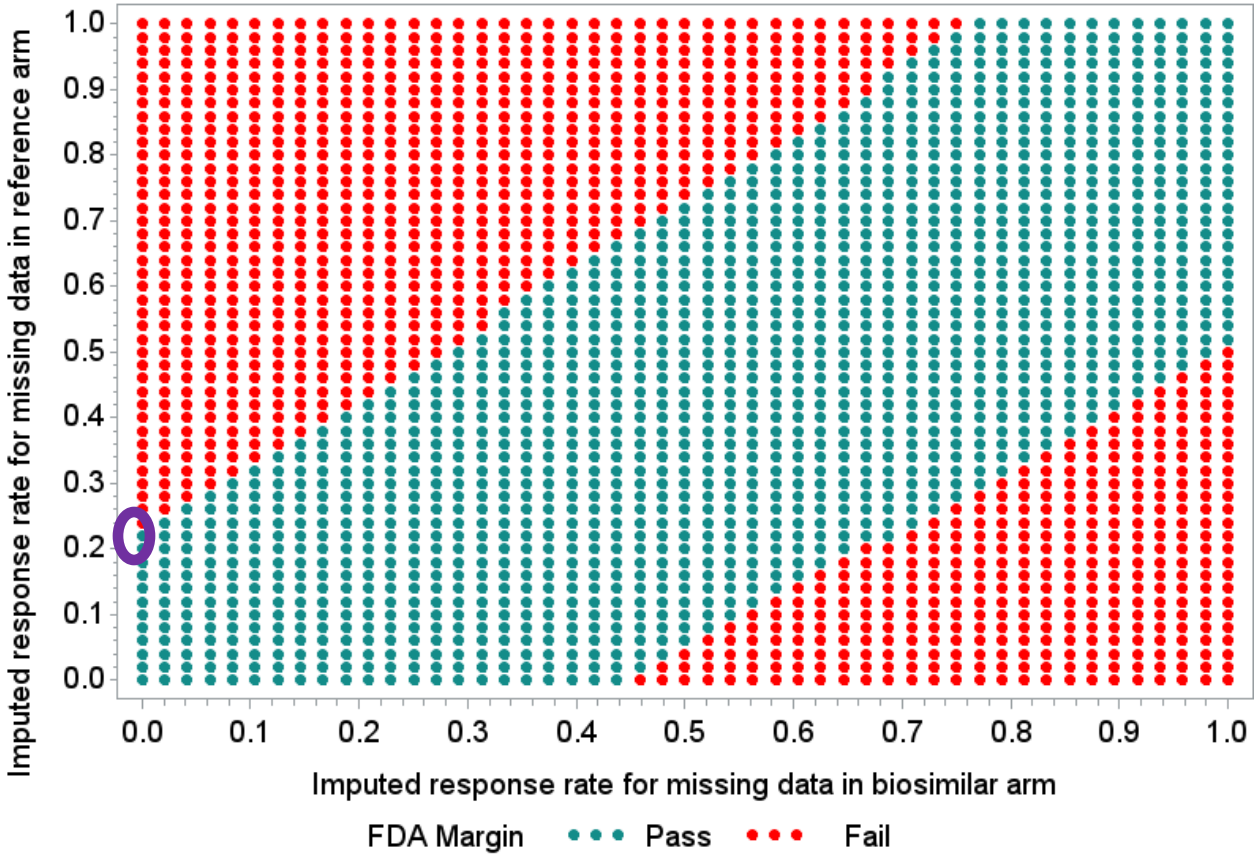


INTERPRETATION OF TP ANALYSIS

- Each dot represents one combination of the imputed response rates for the missing data for the two arms.
  - (0, 0) dot means NRI applied to all missing data.
  - (0.2, 0.3) dot means impute 0.2 for response rate to missing data in the biosimilar arm and 0.3 in the reference arm.
- Based on the pre-defined margin and calculated CI based on each imputation combination, we plot the result of the study (Pass or Fail) accordingly.
  - Both dots (0,0) and (0.2, 0.3) lead to a Pass result.

INTERPRETATION OF TP ANALYSIS

- Next, we need to show it is unlikely to be in the red dot area
- The smallest difference in imputation response rate between the two arms on the boundary is about 21%, we call this as TP.
- We need to show true response rates between arms >= 21% is less possible



TRUE DIFF >= 21% IS LESS LIKELY

- The 2-sided 90% CI based on all observed data is (-9.2%, 4.1%), much less in absolute values than 21%.
- Using a Bayesian method for all observed data, assuming a non-informative prior, the 2-sided 90% credible bound is (-11.3%, 6.2%), much less in absolute values than 21%.
- We can also use historical data
  - i.e. show magnitude of risk difference between the reference vs. the standard of care (SOC) is >= 21%. Given the similarity of the biosimilar and the reference demonstrated in non-clinical and PK, it is less likely that the biosimilar will just have same effect as the SOC.

SIMULATIONS FOR OTHER SCENARIOS

- Previously we shown how tipping point works in similar missing percentage and similar response rate cases. Now lets explore other scenarios.
- Big difference in either response rate or missing percentage will all cast question to biosimilarity between the biosimilar and the reference, so only moderate/mild differences are considered.
- The TPs for each scenario are summarized in following table.

The Biosimilar arm vs. the Reference arm	Same percentage in missing 16%	Higher percentage in missing 20%	Lower percentage in missing 12%
Similar response rate 70%	33%	19%	23%
Higher response rate 73%	19%	15%	1%
Lower response rate 67%	19%	2%	17%

SCENARIOS WITH VERY SMALL TP

- From the two red colored scenarios above, It is interesting to see that when the response rate and the missing percentages are in the opposite direction, it is very hard to prove biosimilarity in your biosimilar compared to the reference, since you will have very small TP.
- Small difference used in the imputation for the missing data (0% vs. 3% and 2% vs. 0%) can not prevent the large difference end-up in the overall rate after imputation.

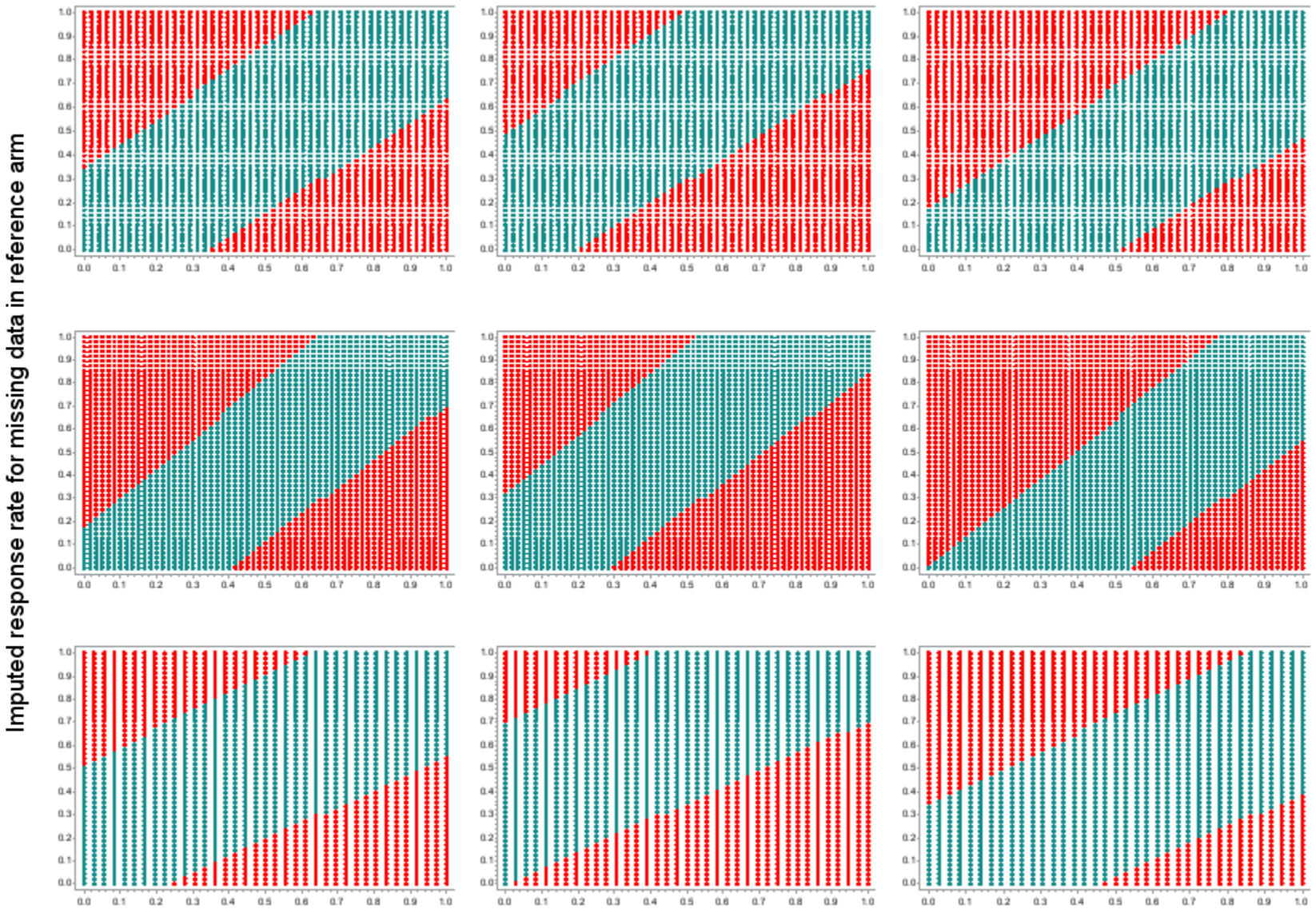
Scenarios	Arms	N per arm	Missing percentage	Response rate observed	Imputed rate for missing	Overall rate after imputation
Lower pct in missing, higher pct in response, TP=2%	biosimilar	300	0.12	0.73	0%	64%
	reference	300	0.16	0.7	3%	59%
Higher pct in missing, lower pct in response, TP=1%	biosimilar	300	0.2	0.67	2%	54%
	reference	300	0.16	0.7	0%	59%

SCENARIOS WITH VERY SMALL TP- COMPARED TO ANALYSIS WITH NRI OR OBSERVED RATES

Biosimilar arm vs. reference arm	Same percentage in missing 16%	Higher percentage in missing 20%	Lower percentage in missing 12%
Similar response rate 70%	33%	19%	23%
Higher response rate 73%	19%	15%	1%
Lower response rate 67%	19%	2%	17%

- While if we conduct tests on response rates based either NRI or observed cases, ALL nine scenarios will be concluded as biosimilar given margin (-0.12, 0.12).
  - Although based on the NRI cases, the two scenarios highlighted in red also have 90% CI with bounds VERY close to the margin, i.e. (-0.1199, 0.0132) and (-0.0119, 0.1186)

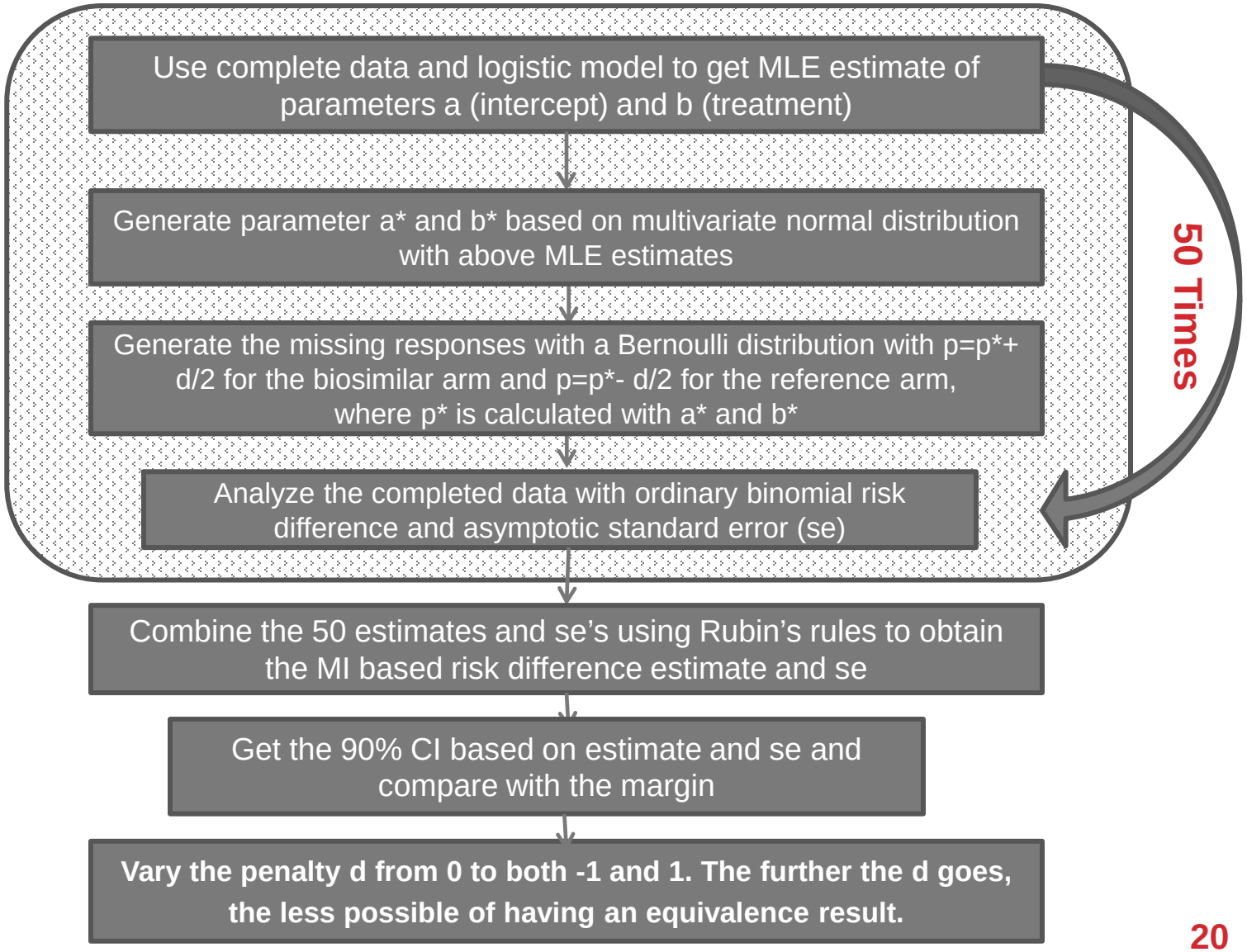
FOR DIFFERENT SCENARIOS



TIPPING POINT BASED ON MI

- A further extension of the TP analysis is based on MI.
- Ratitch, O'Kelly, and Tosiello (2013) proposed the methods of TP analysis based on MI for superiority trials with continuous outcome.
- A brief illustration of how this can be applied to binary data in biosimilarity study is shown in next slide.

TIPPING POINT BASED ON MI



DISCUSSION/CONCLUSIONS

- Pros of TP
  - Conservative in terms of biosimilar evaluation
  - Easy to be applied and interpreted
- Cons of TP
  - Leave more work in evaluating the possibility of meeting TP
  - More properties need to be evaluated

STATISTICAL REFERENCE

- Yan X, Lee S, Li N. (2009) Missing data handling methods in medical device clinical trials. J Biopharm Stat. Nov;19(6):1085-98.
- Carpenter J and Kenward MG (2007) Missing Data in Clinical Trials — a Practical Guide. UK National Health Service, National Coordinating Centre for Research on Methodology.
- Ratitch B, O'Kelly M, Tosiello R (2013). Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceut. Statist*., 12 337–347.

BACKUP SLIDES

TIPPING POINT BASED ON MI

- Use binary endpoint and risk difference as an example:
  1. Use complete data and logistic model $p(y_i=1|x_i)=\exp(a+bx_i)/[1+\exp(a+bx_i)]$ to get MLE estimate of parameters a and b as $\hat{a}$ and $\hat{b}$ and cov($\hat{a}$, $\hat{b}$), where y_i 's are the response and x_i's are the treatments (the biosimilar or the reference)
  2. Generate parameter a* and b* based on multivariate normal distribution with $\hat{a}$ and $\hat{b}$ and cov($\hat{a}$, $\hat{b}$)
  3. Generate the missing outcome with a Bernoulli distribution of p*+ d/2 for the biosimilar arm and p*- d/2 for the reference arm, where p* is calculated with a* and b*.
     - d is called the penalty term and is within (-1, 1).
     - If d adjusted probability is above 1 or below 0 then it is set to be 1 or 0.
  4. Completed data from step 3 will be analyzed with ordinary binomial risk difference and asymptotic standard error (se).
  5. Repeat step 2 to step 4 50 times, resulting in 50 risk differences and standard errors. Combine these using Rubin's rules (Carpenter and Kenward, 2007) to obtain the multiple imputation-based risk difference estimate and se.
  6. Get the 90% CI based on estimate and se from step 5 and see if it is within the margin.
  7. Vary the penalty d from 0 to both -1 and 1. The further the d goes, the less possible of having an equivalence result.
- This search of a grid of d values is similar to the tipping point strategy suggested by Ratitch, O'Kelly, and Tosiello (2013) for continuous data from a superiority trial.