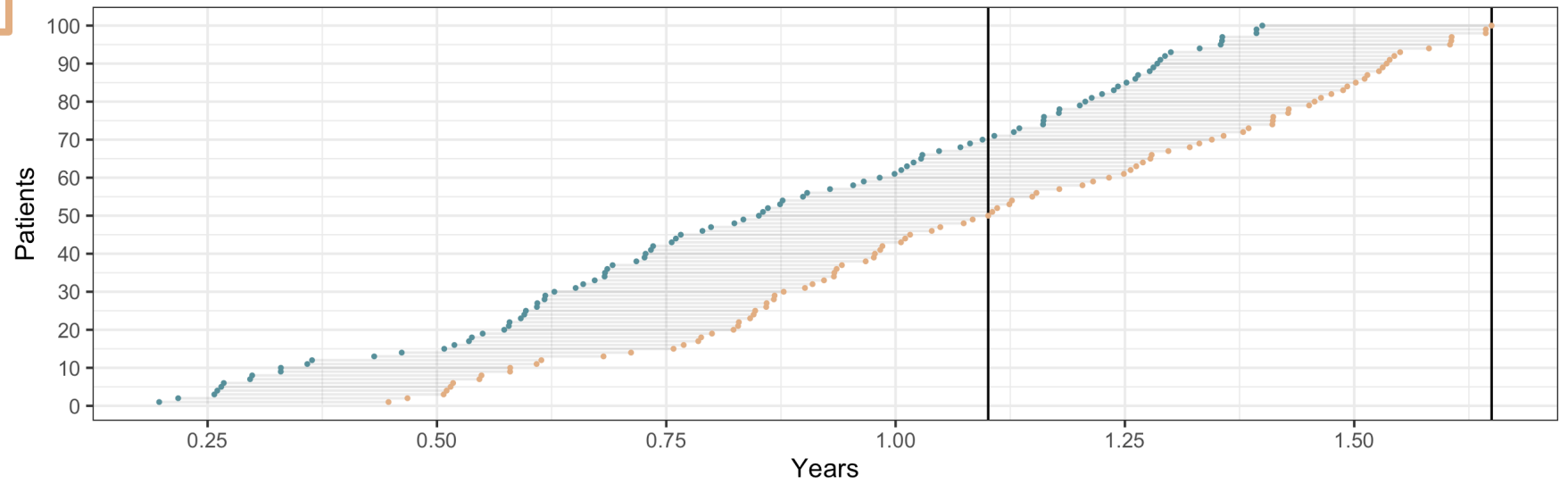
Forecasting with Confidence: Harnessing Predictive Probabilities in Practice

2024 BOSTON

SCT | 45TH ANNUAL MEETING

Predicting Future Data from Current Data

Example Trial



Interim Analysis
50
observations

Final Analysis
100
observations

Given 50 observed patients, what is the probability of success at 100?

What do current data show?

	posterior probability	conditional power	predictive probability
Assumptions	incorporates prior information	frequentist calculation, no priors	incorporates prior information
Information	currently observed data	currently observed data	currently observed data
Goal	summarizes current information + prior	predicts trial success assuming a **precise** future treatment effect	predicts trial success based on a **distribution of possible future treatment effects**

Saville, Detry, & Viele 2023

Given the observed interim data, how likely is the trial to win if all future data show an assumed treatment effect?

	posterior probability	conditional power	predictive probability
Assumptions	incorporates prior information	typically frequentist, no priors	incorporates prior information
Information	currently observed data	currently observed data	currently observed data
Goal	summarizes current information + prior	predicts trial success assuming a **single** future treatment effect	predicts trial success based on a **distribution of possible future treatment effects**

Saville, Detry, & Viele 2023

Given the observed data and distribution of treatment effects, how likely is the trial to win?

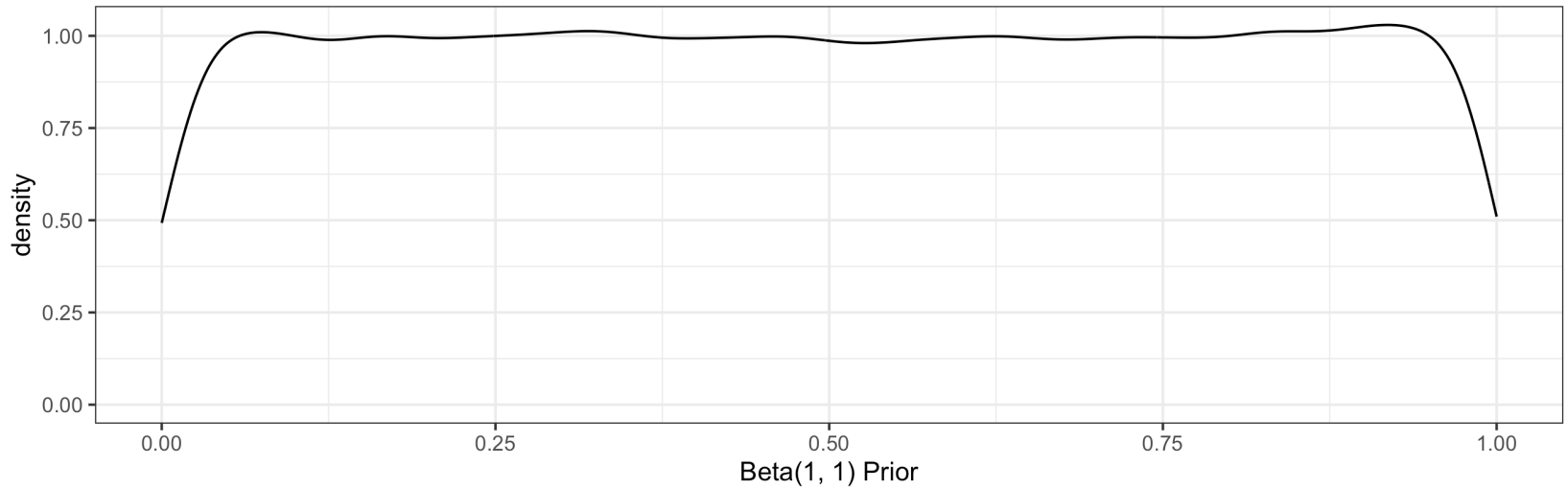
	posterior probability	conditional power	predictive probability
Assumptions	incorporates prior information	typically frequentist, no priors	incorporates prior information
Information	currently observed data	currently observed data	currently observed data
Goal	summarizes current information + prior	predicts trial success assuming a **precise** future treatment effect	predicts trial success based on a **distribution of possible future treatment effects**

Saville, Detry, & Viele 2023

Computing Predictive Probabilities – Closed Form

centered at prior estimate

$$\theta \backsim \text{Beta}(\alpha, \beta)$$



Bayesian Adaptive Methods for Clinical Trials. Berry et al. 2011

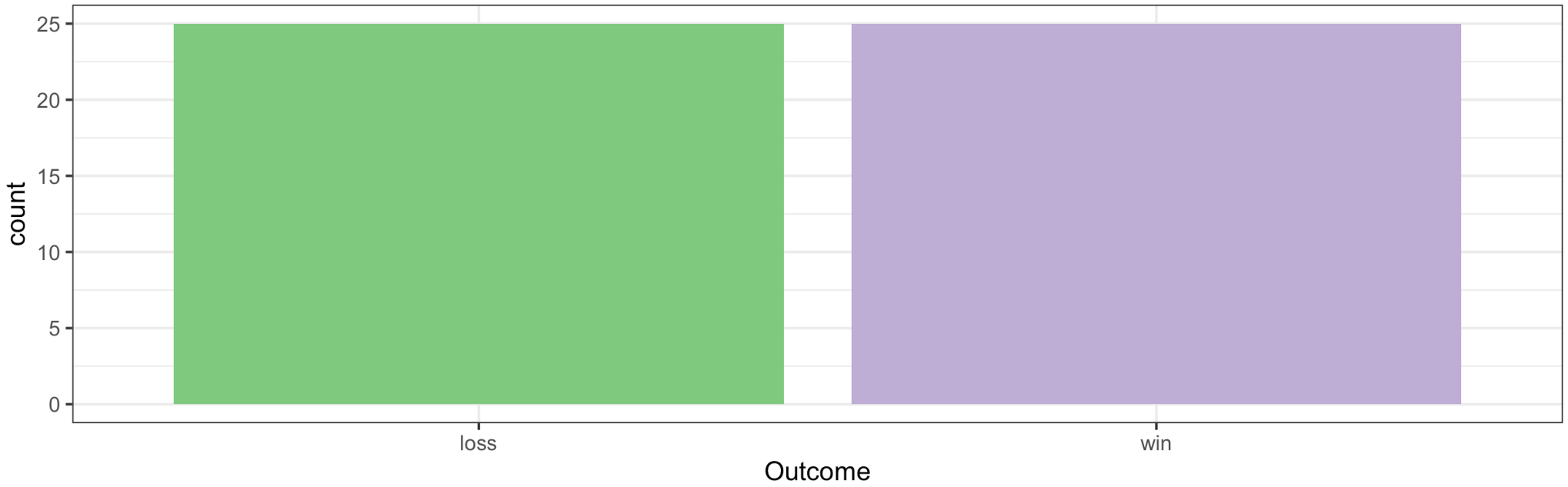
Computing Predictive Probabilities – Closed Form

centered at prior estimate

$$\theta \backsim \text{Beta}(\alpha, \beta)$$

observed data at N = 50
25 wins, 25 failures

$$x_1 \backsim \text{Binomial}(n_1, \theta)$$



Bayesian Adaptive Methods for Clinical Trials. Berry et al. 2011

Computing Predictive Probabilities – Closed Form

centered at prior estimate

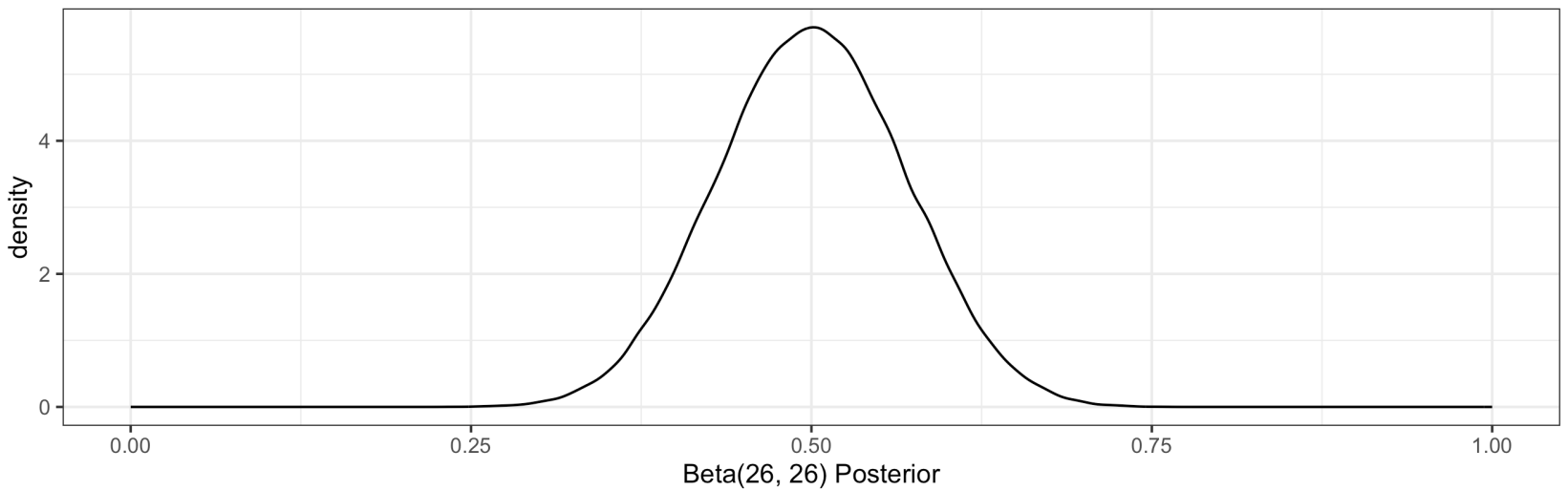
$$\theta \backsim \text{Beta}(\alpha, \beta)$$

observed data at N = 50
25 wins, 25 failures

$$x_1 \backsim \text{Binomial}(n_1, \theta)$$

posterior distribution

$$\theta | x_1, n_1 \backsim \text{Beta}(\alpha + x_1, \beta + n_1 - x_1)$$



Bayesian Adaptive Methods for Clinical Trials. Berry et al. 2011

Computing Predictive Probabilities – Closed Form

centered at prior estimate

$$\theta \backsim \text{Beta}(\alpha, \beta)$$

observed data at N = 50
25 wins, 25 failures

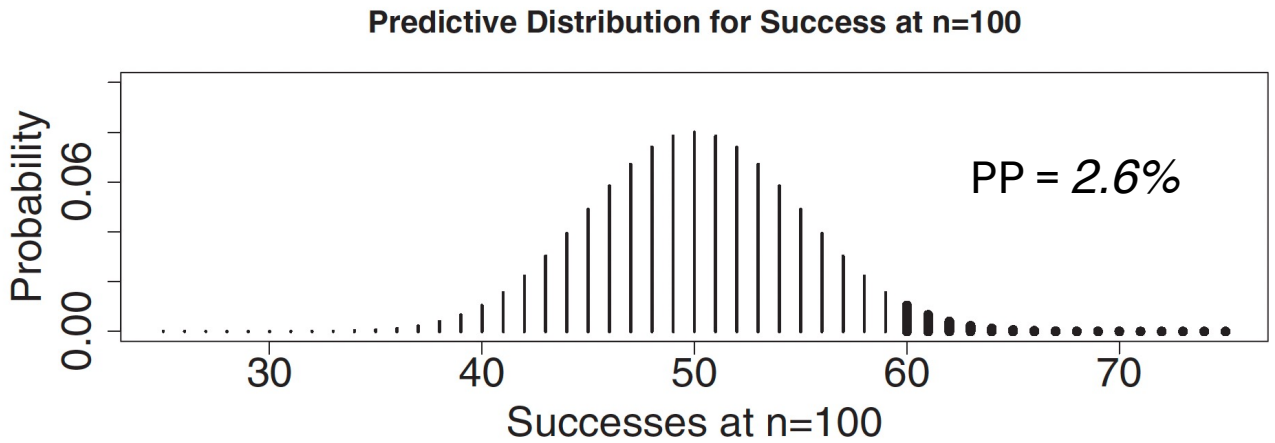
$$x_1 \backsim \text{Binomial}(n_1, \theta)$$

posterior distribution

$$\theta | x_1, n_1 \backsim \text{Beta}(\alpha + x_1, \beta + n_1 - x_1)$$

predictive distribution for next n_2 observations

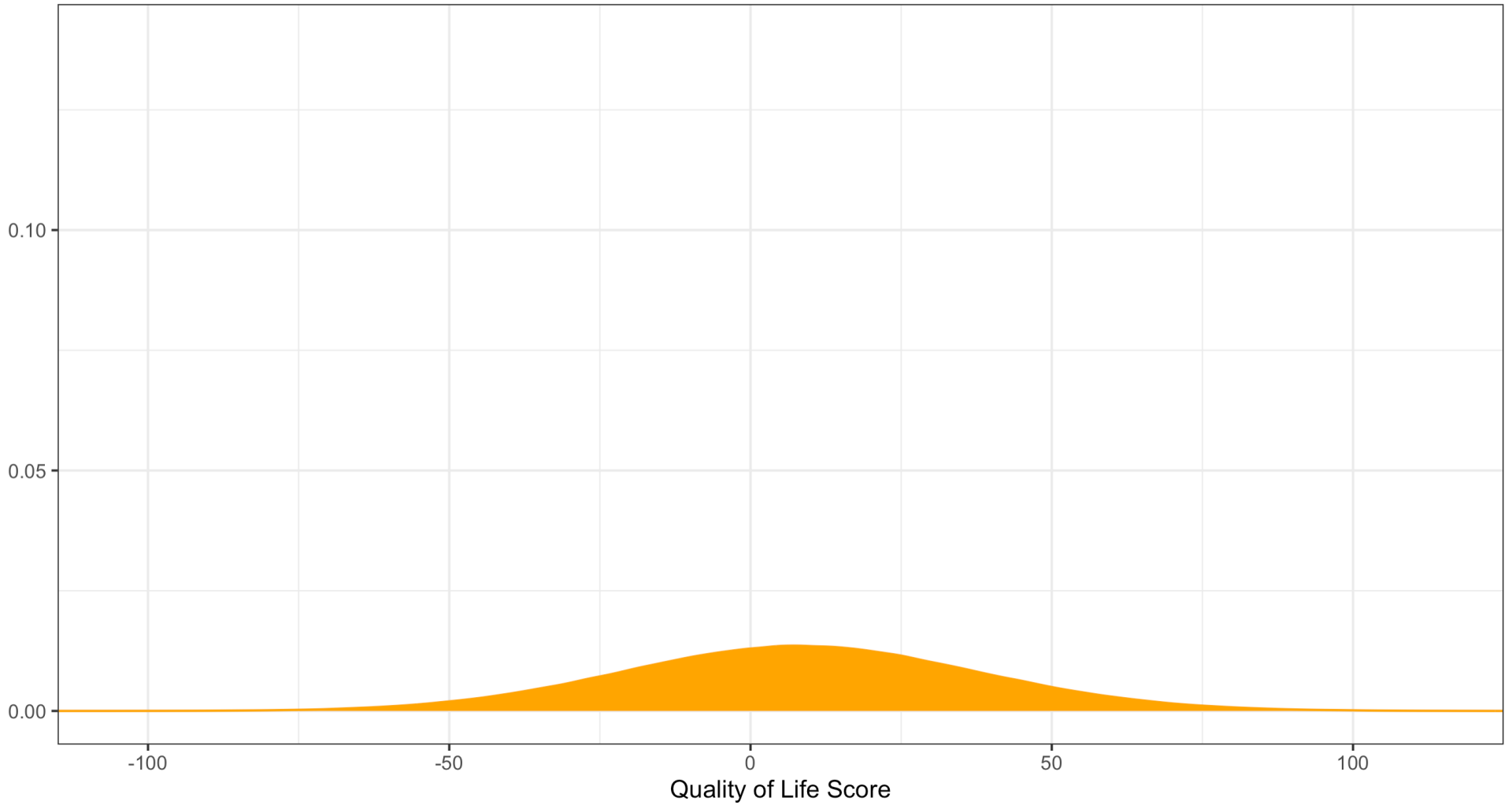
$$\text{x}_2 \mid n_1, \alpha + x, \beta + n_1 - x_1 \backsim \text{Beta-Binomial}(n_2, \alpha + x_1, \beta + n_1 - x_1)$$



Bayesian Adaptive Methods for Clinical Trials. Berry et al. 2011

Calculating a Predictive Probability of Success: Monte Carlo Integration

prior information
- clinical expertise
- previous studies
- purposefully diffuse

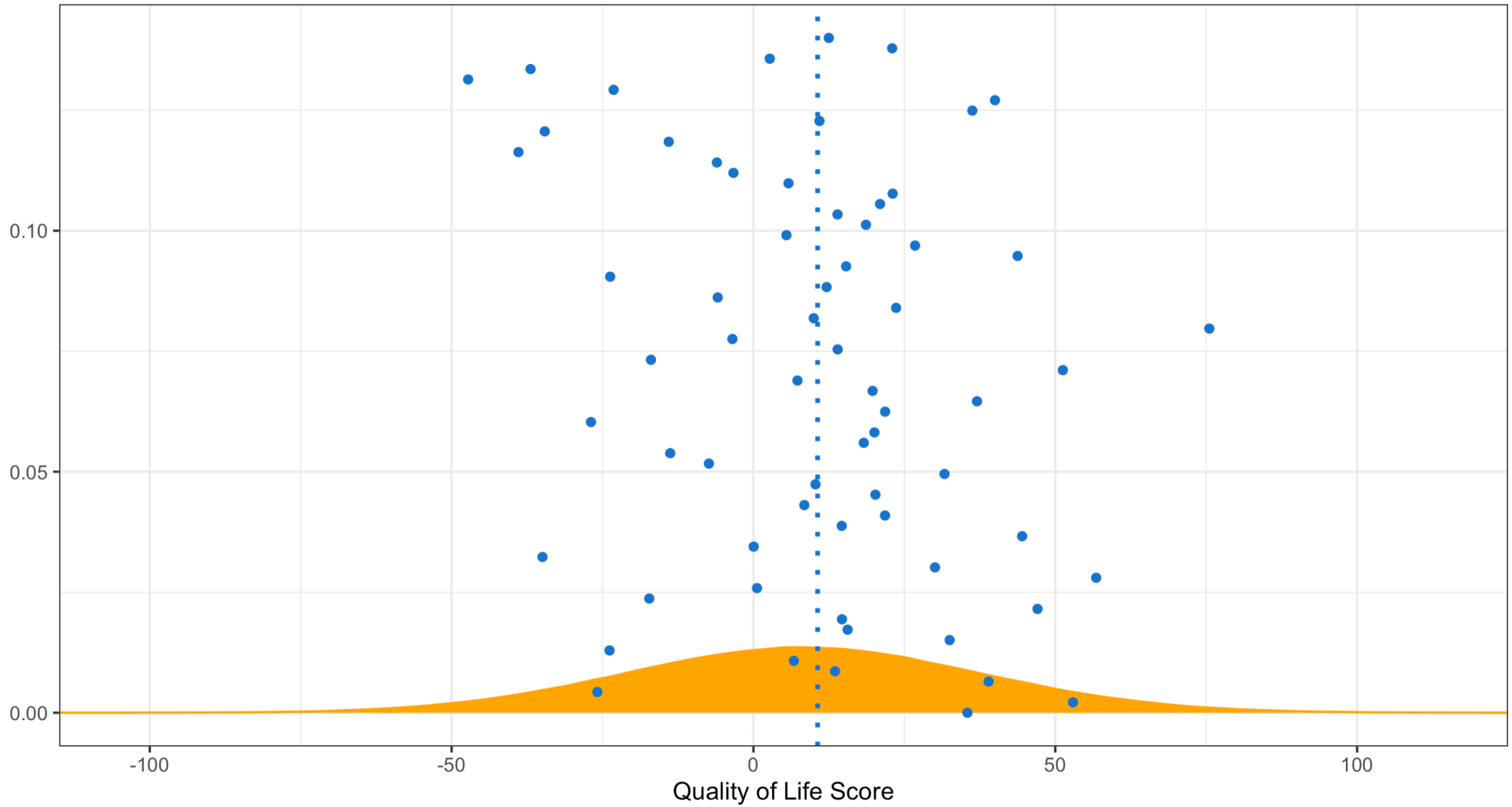


Calculating a Predictive Probability of Success

prior information
- clinical expertise
- previous studies
- purposefully diffuse

+

interim observed data



Calculating a Predictive Probability of Success

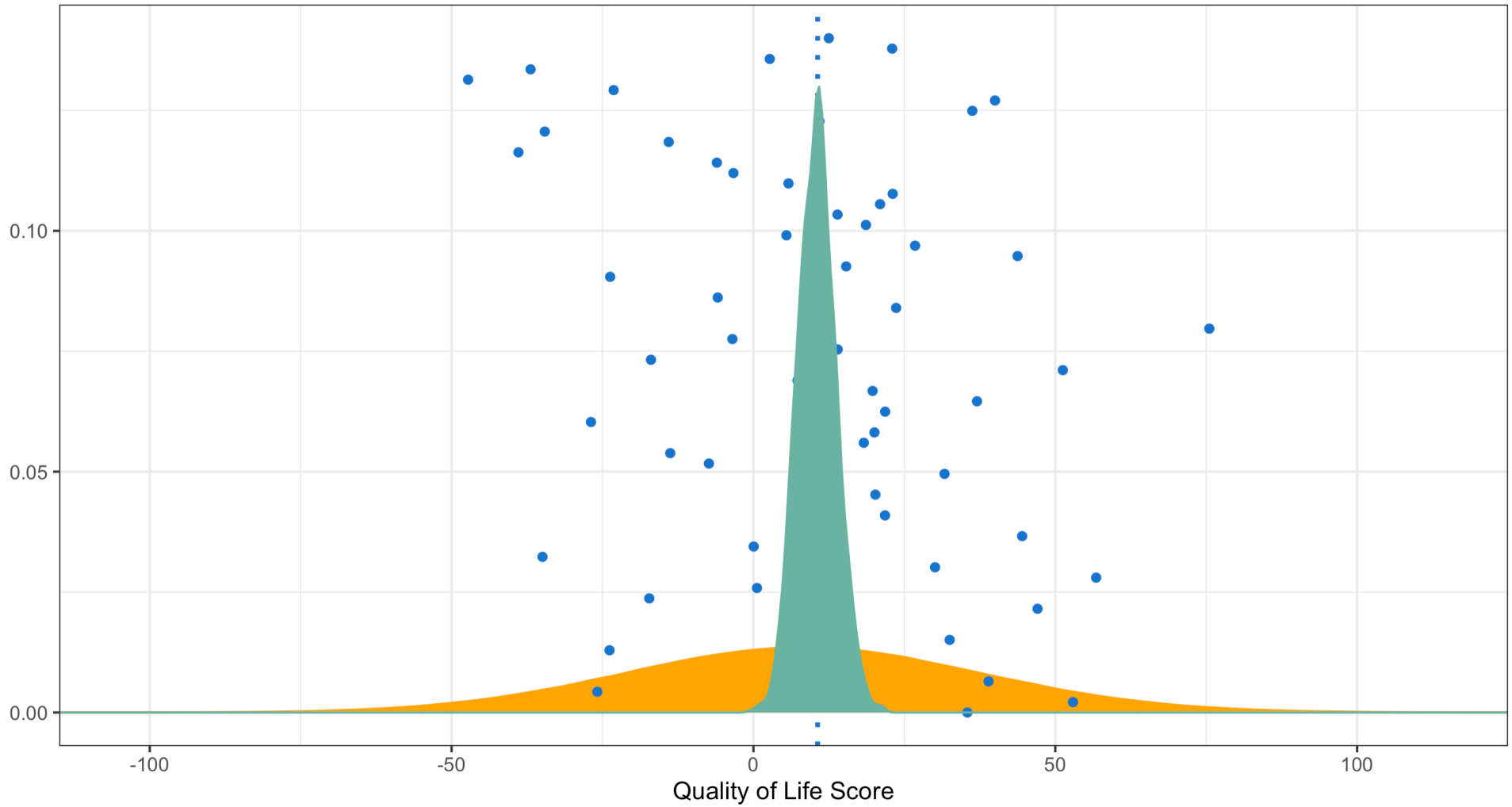
prior information
- clinical expertise
- previous studies
- purposefully diffuse

+

interim observed data

=

posterior distribution of the mean



Calculating a Predictive Probability of Success

prior information
- clinical expertise
- previous studies
- purposefully diffuse

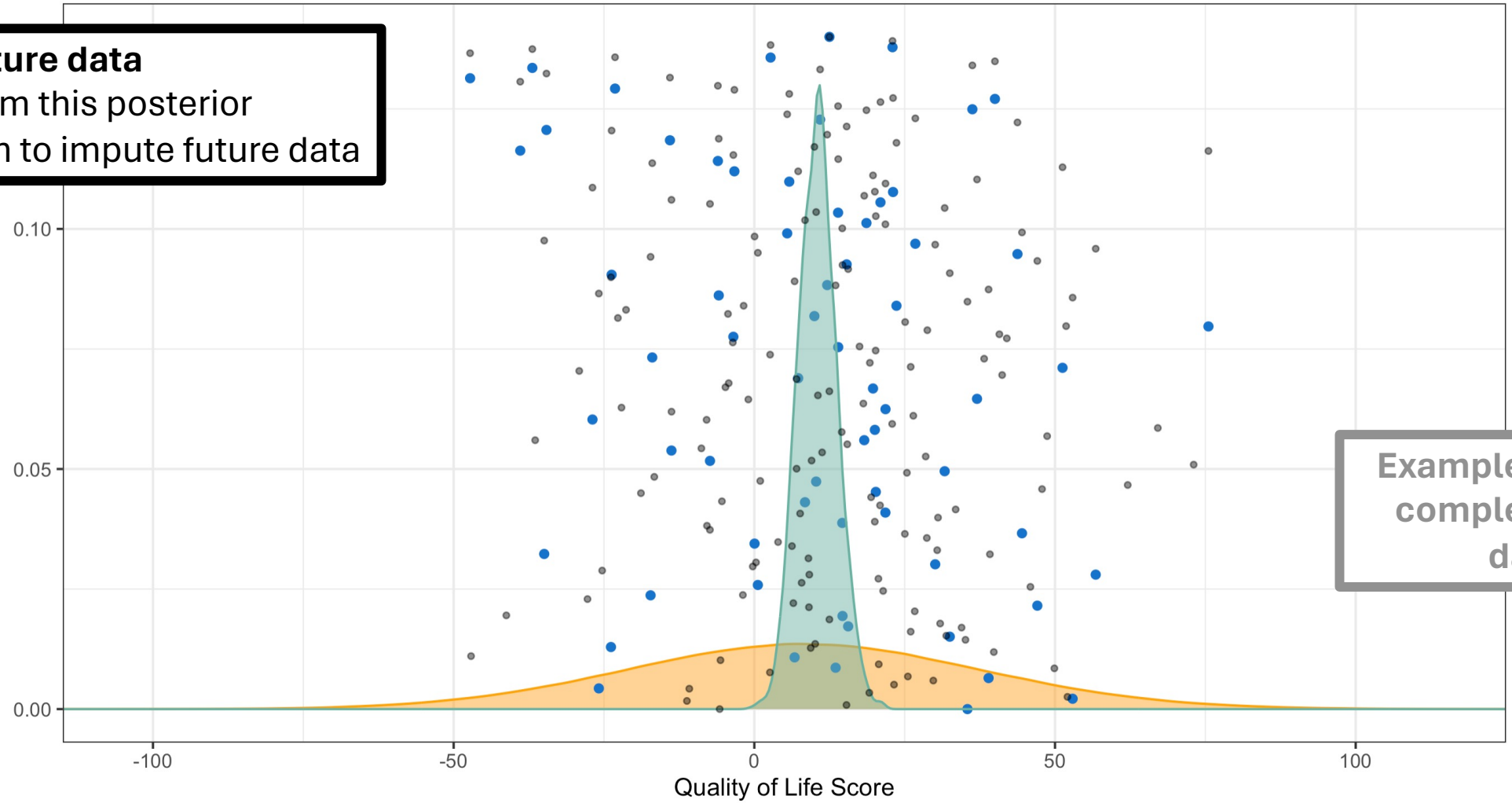
+

interim observed data

=

posterior distribution of the mean

impute future data
sample from this posterior distribution to impute future data



Example imputed completed trial data

Calculating a Predictive Probability of Success

prior information
- clinical expertise
- previous studies
- purposefully diffuse

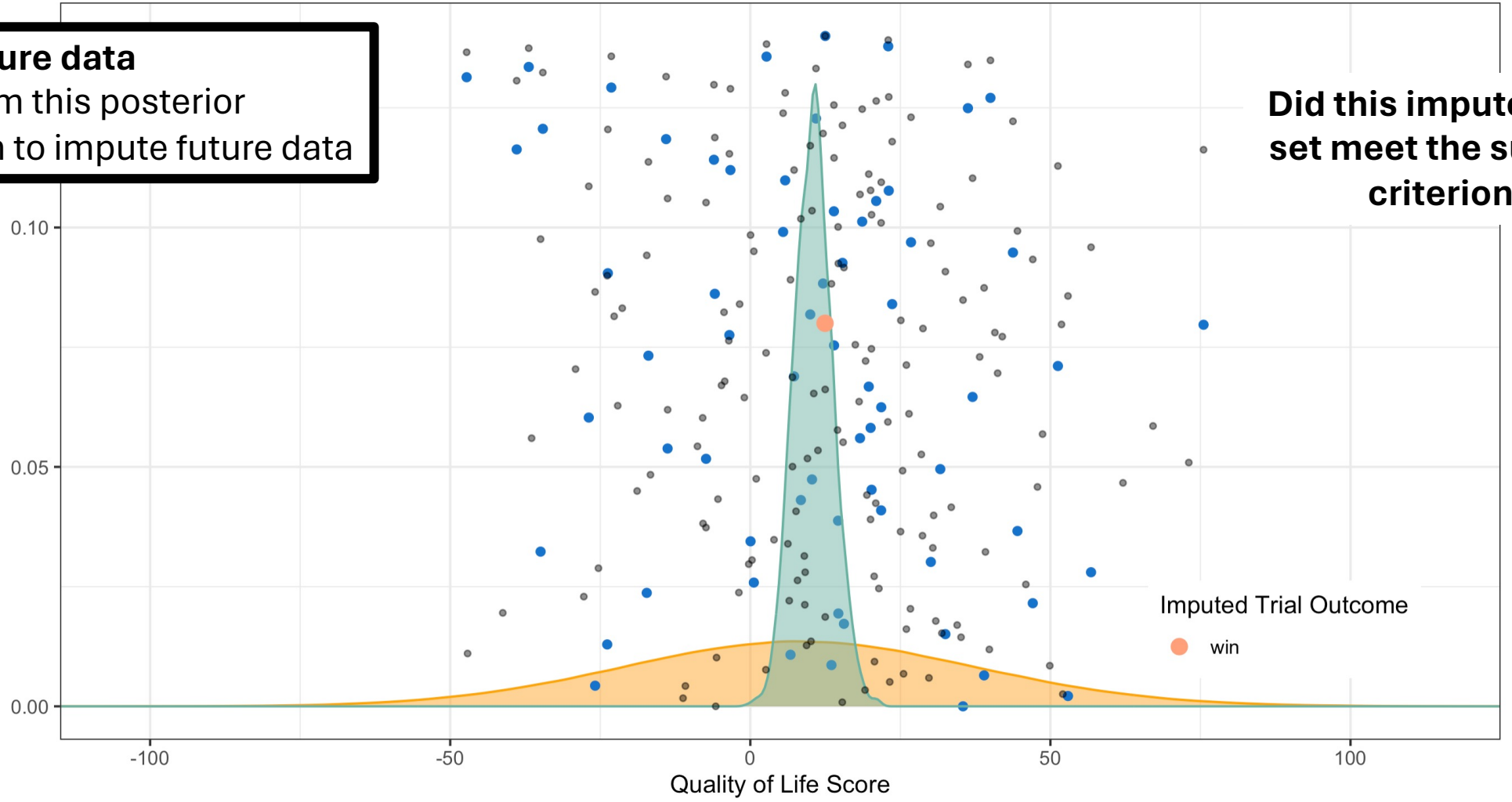
+

interim observed data

=

posterior distribution of the mean

impute future data
sample from this posterior distribution to impute future data



Did this imputed data set meet the success criterion?

Calculating a Predictive Probability of Success

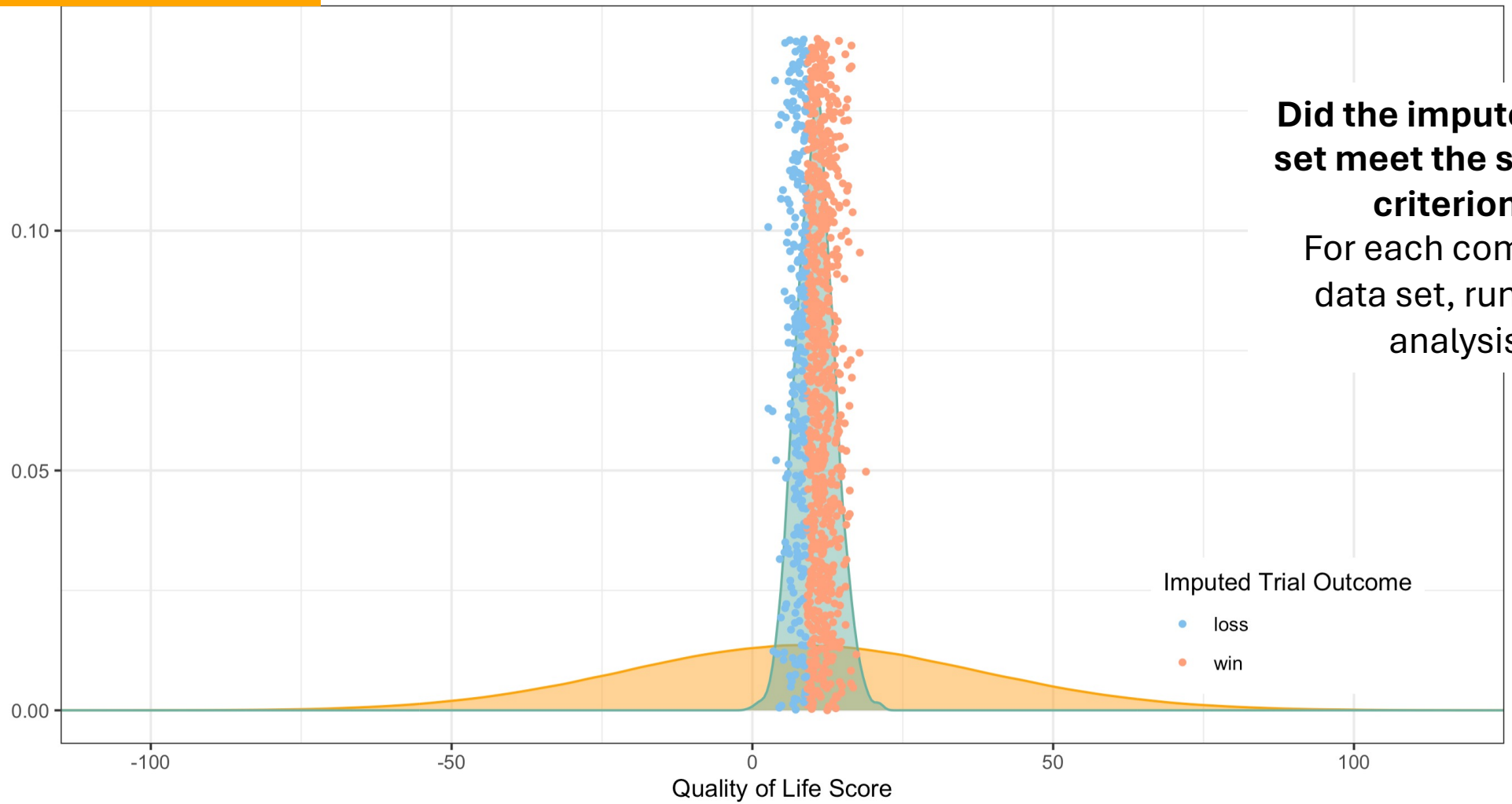
prior information
- clinical expertise
- previous studies
- purposefully diffuse

+

interim observed data

=

posterior distribution of the mean



Did the imputed data set meet the success criterion?

For each complete data set, run final analysis

Calculating a Predictive Probability of Success

prior information
- clinical expertise
- previous studies
- purposefully diffuse

+

interim observed data

=

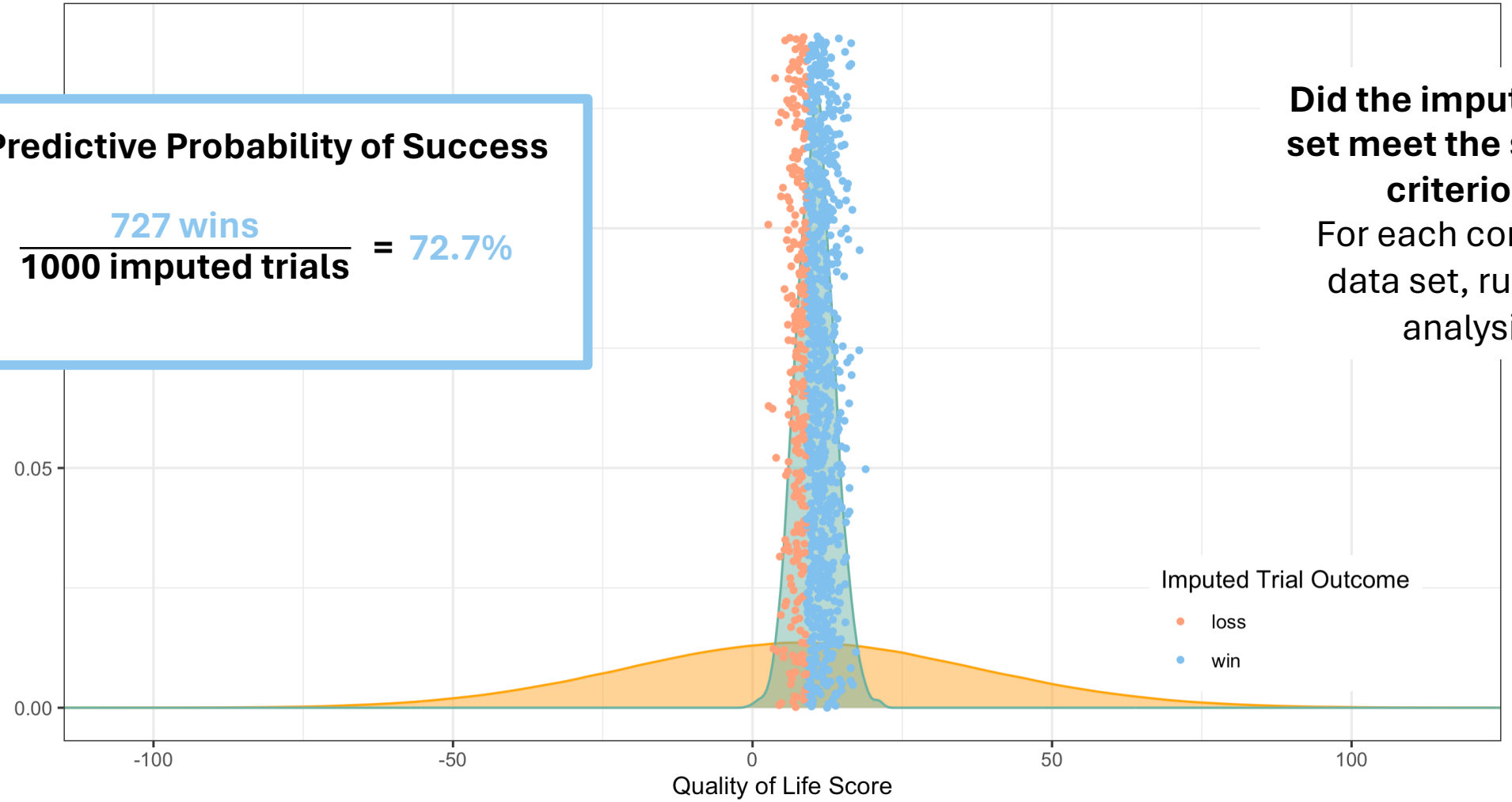
posterior distribution of the mean

Predictive Probability of Success

$$\frac{\text{727 wins}}{\textbf{1000 imputed trials}} = 72.7\%$$

Did the imputed data set meet the success criterion?

For each complete data set, run final analysis



When would we need predictive probabilities?

- To **choose a sample size** at a prespecified interim analysis

The NEW ENGLAND JOURNAL of MEDICINE

SPECIALTIES TOPICS MULTIMEDIA CURRENT ISSUE LEARNING/CME AUTHOR CENTER PUBLICATIONS

ORIGINAL ARTICLE

Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct

Authors: Raul G. Nogueira, M.D., Ashutosh P. Jadhav, M.D., Ph.D., Diogo C. Haussen, M.D., Alain Bonafe, M.D., Ronald F. Budzik, M.D., Parita Bhuva, M.D., Dileep R. Yavagal, M.D., Marc Ribo, M.D., Christophe Cognard, M.D., Ricardo A. Hanel, M.D., Cathy A. Sila, M.D., Ameer E. Hassan, D.O., Monica Millan, M.D., Elad I. Levy, M.D., Peter Mitchell, M.D., Michael Chen, M.D., Joey D. English, M.D., Qaisar A. Shah, M.D., Frank L. Silver, M.D., Vitor M. Pereira, M.D., Brijesh P. Mehta, M.D., Blaise W. Baxter, M.D., Michael G. Abraham, M.D., Pedro Cardona, M.D., Erol Veznedaroglu, M.D., Frank R. Hellinger, M.D., Lei Feng, M.D., Jawad F. Kirmani, M.D., Demetrius K. Lopes, M.D., Brian T. Jankowitz, M.D., Michael R. Frankel, M.D., Vincent Costalat, M.D., Nirav A. Vora, M.D., Albert J. Yoo, M.D., Ph.D., Amer M. Malik, M.D., Anthony J. Furlan, M.D., Marta Rubiera, M.D., Amin Aghaebrahim, M.D., Jean-Marc Olivot, M.D., Wondwossen G. Tekle, M.D., Ryan Shields, M.Sc., Todd Graves, Ph.D., Roger J. Lewis, M.D., Ph.D., Wade S. Smith, M.D., Ph.D., David S. Liebeskind, M.D., Jeffrey L. Saver, M.D., and Tudor G. Jovin, M.D., for the DAWN Trial Investigators* -40 Author Info & Affiliations

Published November 11, 2017 | N Engl J Med 2018;378:11-21 | DOI: 10.1056/NEJMoa1706442 | VOL. 378 NO. 1

STATISTICAL ANALYSIS

The adaptive trial design allowed for a sample size ranging from 150 to 500 patients. During interim analyses, the decision to stop or continue enrollment was based on a prespecified calculation of the probability that thrombectomy plus standard care would be superior to standard care alone with respect to the first primary end point. The enrichment trial design gave us the flexibility to identify whether the benefit of the trial intervention was restricted to a subgroup of patients with relatively small infarct volumes at baseline. The interim analyses, which included patients with available follow-up data at the time of the analysis, were prespecified to test for the futility, enrichment, and success of the trial.

When would we need predictive probabilities?

- To **identify subgroups** benefiting most from a treatment

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct

Authors: Raul G. Nogueira, M.D., Ashutosh P. Jadhav, M.D., Ph.D., Diogo C. Haussen, M.D., Alain Bonafe, M.D., Ronald F. Budzik, M.D., Parita Bhuva, M.D., Dileep R. Yavagal, M.D., Marc Ribo, M.D., Christophe Cognard, M.D., Ricardo A. Hanel, M.D., Cathy A. Sila, M.D., Ameer E. Hassan, D.O., Monica Millan, M.D., Elad I. Levy, M.D., Peter Mitchell, M.D., Michael Chen, M.D., Joey D. English, M.D., Qaisar A. Shah, M.D., Frank L. Silver, M.D., Vitor M. Pereira, M.D., Brijesh P. Mehta, M.D., Blaise W. Baxter, M.D., Michael G. Abraham, M.D., Pedro Cardona, M.D., Erol Veznedaroglu, M.D., Frank R. Hellinger, M.D., Lei Feng, M.D., Jawad F. Kirmani, M.D., Demetrius K. Lopes, M.D., Brian T. Jankowitz, M.D., Michael R. Frankel, M.D., Vincent Costalat, M.D., Nirav A. Vora, M.D., Albert J. Yoo, M.D., Ph.D., Amer M. Malik, M.D., Anthony J. Furlan, M.D., Marta Rubiera, M.D., Amin Aghaebrahim, M.D., Jean-Marc Olivot, M.D., Wondwossen G. Tekle, M.D., Ryan Shields, M.Sc., Todd Graves, Ph.D., Roger J. Lewis, M.D., Ph.D., Wade S. Smith, M.D., Ph.D., David S. Liebeskind, M.D., Jeffrey L. Saver, M.D., and Tudor G. Jovin, M.D., for the DAWN Trial Investigators* -40 Author Info & Affiliations

Published November 11, 2017 | N Engl J Med 2018;378:11-21 | DOI: 10.1056/NEJMoa1706442 | VOL. 378 NO. 1

STATISTICAL ANALYSIS

The adaptive trial design allowed for a sample size ranging from 150 to 500 patients. During interim analyses, the decision to stop or continue enrollment was based on a prespecified calculation of the probability that thrombectomy plus standard care would be superior to standard care alone with respect to the first primary end point. The enrichment trial design gave us the flexibility to identify whether the benefit of the trial intervention was restricted to a subgroup of patients with relatively small infarct volumes at baseline. The interim analyses, which included patients with available follow-up data at the time of the analysis, were prespecified to test for the futility, enrichment, and success of the trial.

When would we need predictive probabilities?

- To **identify subgroups** benefiting most from a treatment

The NEW ENGLAND JOURNAL of MEDICINE

SPECIALTIES TOPICS

ORIGINAL ARTICLE

Results led to an FDA expansion

STATISTICAL ANALYSIS

The adaptive trial design allowed for a sample size ranging from 150 to 500 patients. During interim ecision to stop or continue enroll- d on a prespecified calculation of

FDA NEWS RELEASE

FDA expands treatment window for use of clot retrieval devices in certain stroke patients

Furlan, M.D., Marta Rubiera, M.D., Amin Aghaebrahim, M.D., Jean-Marc Olivot, M.D., Wondwossen G. Tekle, M.D., Ryan Shields, M.Sc., Todd Graves, Ph.D., Roger J. Lewis, M.D., Ph.D., Wade S. Smith, M.D., Ph.D., David S. Liebeskind, M.D., Jeffrey L. Saver, M.D., and Tudor G. Jovin, M.D., for the DAWN Trial Investigators* -40 Author Info & Affiliations

prespecified to test for the futility, enrichment, and success of the trial.

https://www.fda.gov/news-events/press-announcements/fda-expands-treatment-window-use-clot-retrieval-devices-certain-stroke-patients

When would we want to use a predictive probability?

- To determine if additional data are likely to provide convincing evidence of a treatment effect. In other words, **should the trial stop for futility**?

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Randomized Trial of Three Anticonvulsant Medications for Status Epilepticus

Authors: Jaideep Kapur, M.B., B.S., Ph.D., Jordan Elm, Ph.D., James M. Chamberlain, M.D., William Barsan, M.D., James Cloyd, Pharm.D., Daniel Lowenstein, M.D., Shlomo Shinnar, M.D., Ph.D., +6, for the NETT and PECARN Investigators* Author Info & Affiliations

Published November 27, 2019 | N Engl J Med 2019;381:2103-2113 | DOI: 10.1056/NEJMoa1905795
VOL. 381 NO. 22

When would we want to use a predictive probability?

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Randomized Trial of Three Anticonvulsant Medications for Status Epilepticus

Authors: Jaideep Kapur, M.B., B.S., Ph.D., Jordan Elm, Ph.D., James M. Chamberlain, M.D., William Cloyd, Pharm.D., Daniel Lowenstein, M.D., Shlomo Shinnar, M.D., Ph.D., +6, for the NETT and P Investigators* Author Info & Affiliations

Published November 27, 2019 | N Engl J Med 2019;381:2103-2113 | DOI: 10.1056/NEJMoa1905
VOL. 381 NO. 22

Table S6. Computations of the futility analysis

Look	Predictive probability that an arm is identified as best / worst at maximum sample size*			Predictive probability that any arm Wins**
	Levetiracetam	Fosphenytoin	Valproate	
Analysis after 400th Enrollment (N=384 unique subjects)				

* Maximum sample size was assumed to be 720 unique subjects for calculation of the predictive probabilities.

** This represents the sum of the predictive probabilities arm is best/worst at the maximum sample size for each of the 3 groups. If this sum is < 5%, the trial stops for futility.

When would we want to use a predictive probability?

The NEW ENGLAND JOURNAL of MEDICINE

SPECIALTIES TOPICS MULTIMEDIA CURRENT ISSUE LEARNING/CME AUTHOR CENTER PUBLICATIONS

ORIGINAL ARTICLE

Randomized Trial of Three Anticonvulsant Medications for Status Epilepticus

Authors: Jaideep Kapur, M.B., B.S., Ph.D., Jordan Elm, Ph.D., James M. Chamberlain, M.D., William Cloyd, Pharm.D., Daniel Lowenstein, M.D., Shlomo Shinnar, M.D., Ph.D., +6, for the NETT and P Investigators* Author Info & Affiliations

Published November 27, 2019 | N Engl J Med 2019;381:2103-2113 | DOI: 10.1056/NEJMoa1905

VOL. 381 NO. 22

Table S6. Computations of the futility analysis

Look	Predictive probability that an arm is identified as best / worst at maximum sample size*			Predictive probability that any arm Wins**
	Levetiracetam	Fosphenytoin	Valproate	
Analysis after 400th Enrollment (N=384 unique subjects)	.0013 / .0008	.002 / .0027	.0022 / .0013	0.01

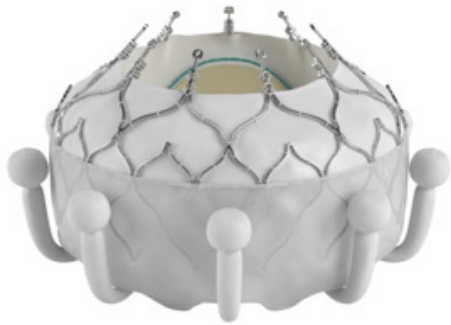
* Maximum sample size was assumed to be 720 unique subjects for calculation of the predictive probabilities.

** This represents the sum of the predictive probabilities arm is best/worst at the maximum sample size for each of the 3 groups. If this sum is < 5%, the trial stops for futility.

In November 2017, enrollment was discontinued at the recommendation of the data and safety monitoring board after the trial met the predefined futility criterion in a planned interim analysis, since there was a 1% chance of showing a most effective or least effective treatment if the trial were to continue to the maximum sample size.

When would we want to use a predictive probability?

- To **support breakthrough device pathway discussions**, as part of a larger submission package



Could we predict success at full follow-up?

Data available:

- ≥ 6-month follow-up in 150 patients
- Some follow-up on the last 250 patients
  - Quality of life measures
  - Tricuspid re-interventions
  - Preliminary HF hospitalization and mortality data

https://www.medtechdive.com/news/edwards-Q4-tricuspid-valve-early-fda-approval/706837/

When would we want to use a predictive probability?

- To **support breakthrough device pathway discussions**, as part of a larger submission package

How Edwards' Evoque valve won an early nod from the FDA

Daveen Chopra, Edwards' vice president for transcatheter mitral and tricuspid therapies (TMTT), said the FDA granted approval for Evoque through the agency's breakthrough device pathway based on the strength of its clinical data in a 150-patient cohort. But the company submitted other data from a larger cohort that saw favorable trends for mortality, heart failure hospitalizations and tricuspid re-interventions.

https://www.medtechdive.com/news/edwards-Q4-tricuspid-valve-early-fda-approval/706837/

SHINE SHADOW

Kristine Broglio

Statistical Science Senior Director

Statistical Innovation, AstraZeneca

SCT | Boston | May 2024

Statistical Paradigms in Clinical Trials

Frequentist

- Common, traditional, and “well-understood”
- Hypothesis testing framework, z-scores, p-values

Bayesian

- More frequently used in early phase
- Integrates data sources together
- Probability statements

ADAPT-IT

- Adaptive Designs Accelerating Promising Trials into Treatments
- FDA and NIH grant
- Objectives
  - Design Bayesian adaptive confirmatory trials for neurologic emergencies
  - Use mixed methods to gain insights into the process

Stroke Hyperglycemia Insulin Network Effect (SHINE)

- First trial in ADAPT-IT
  - Already had a design and protocol
  - Opportunity to create an alternative Bayesian adaptive design
  - Compare the two approaches on their ability to address primary efficacy
- Trial outline
  - Intensive glucose control vs standard of care for stroke patients with hyperglycemia
  - Primary endpoint: Favorable (yes/no) neurologic outcome at 90 days
  - 1400 patients
  - Power for an improvement from 25% (control) to 32% (treatment)

The Designs

Side by Side

SHINE

- 1400 patients + sample size re-estimation
- Frequentist approach
- Early success and futility
- Interim Analyses:
  - Group sequential
  - Number of patients w/ the opportunity to complete
  - 4 interims at 500, 700, 900, and 1100 patients

SHADOW

- 1400 patient maximum
- Bayesian ”Goldilocks” algorithm
- Accrual stopping and futility
- Interim Analyses:
  - Predictive probabilities
  - Accrual milestones
  - 9 interims starting with 500 patients and every subsequent 100 patients

Goldilocks Sample Size Selection

PPn: Predictive probability of success at the current sample size	*PPmax:* Predictive probability with the maximum sample size
• Used for a kind of early success stopping • If this number is high (> 99): Stop accrual -> complete FU -> conduct final analysis	• Used for early futility stopping • If this number is low (<5%): Stop the trial early

Different Stopping Behavior

- Futility is similar:

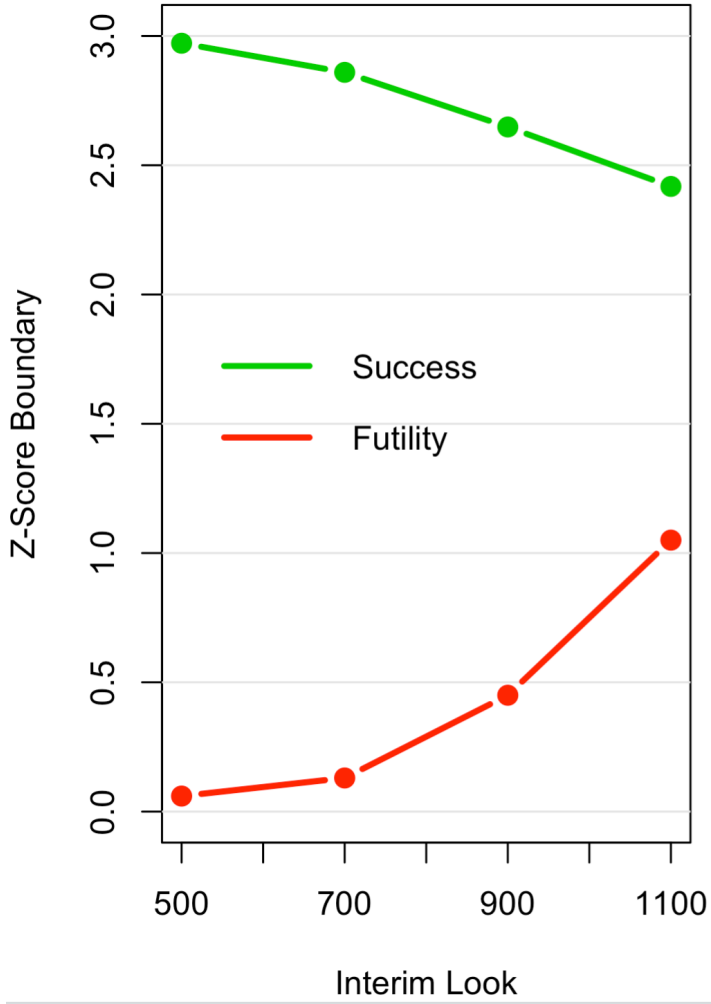
  GSD middle two looks are conservative

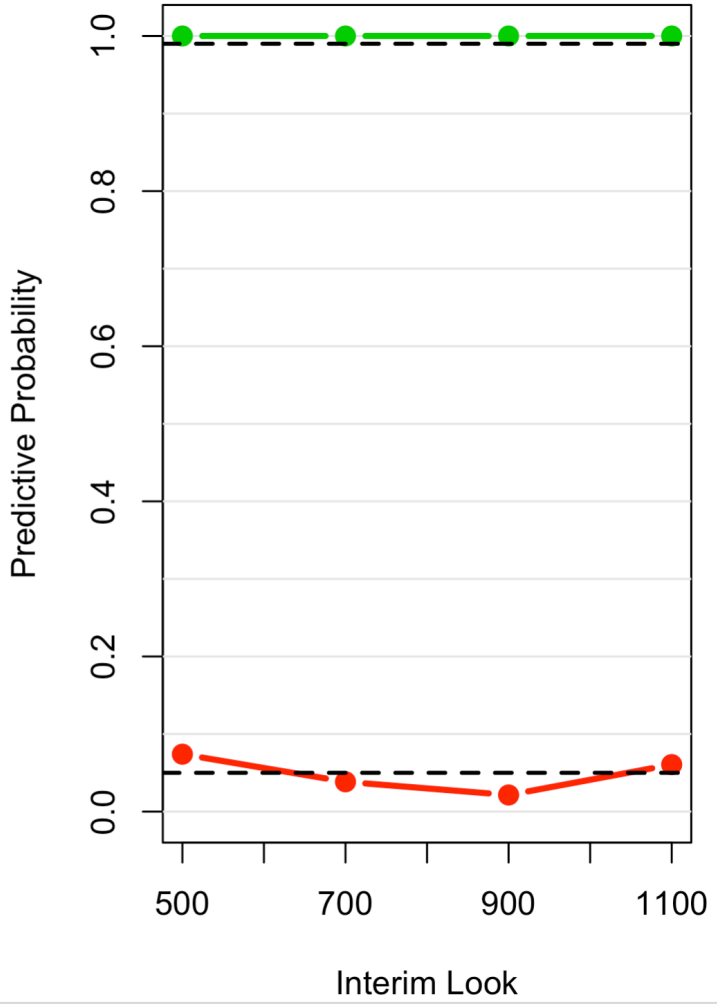
- Success:

  GSD is more conservative

- Final critical values:

4 GSD looks:	0.0215
Goldilocks:	~0.0210
9 GSD looks:	0.0203





A note on alpha: 9 looks for almost the price of 4?

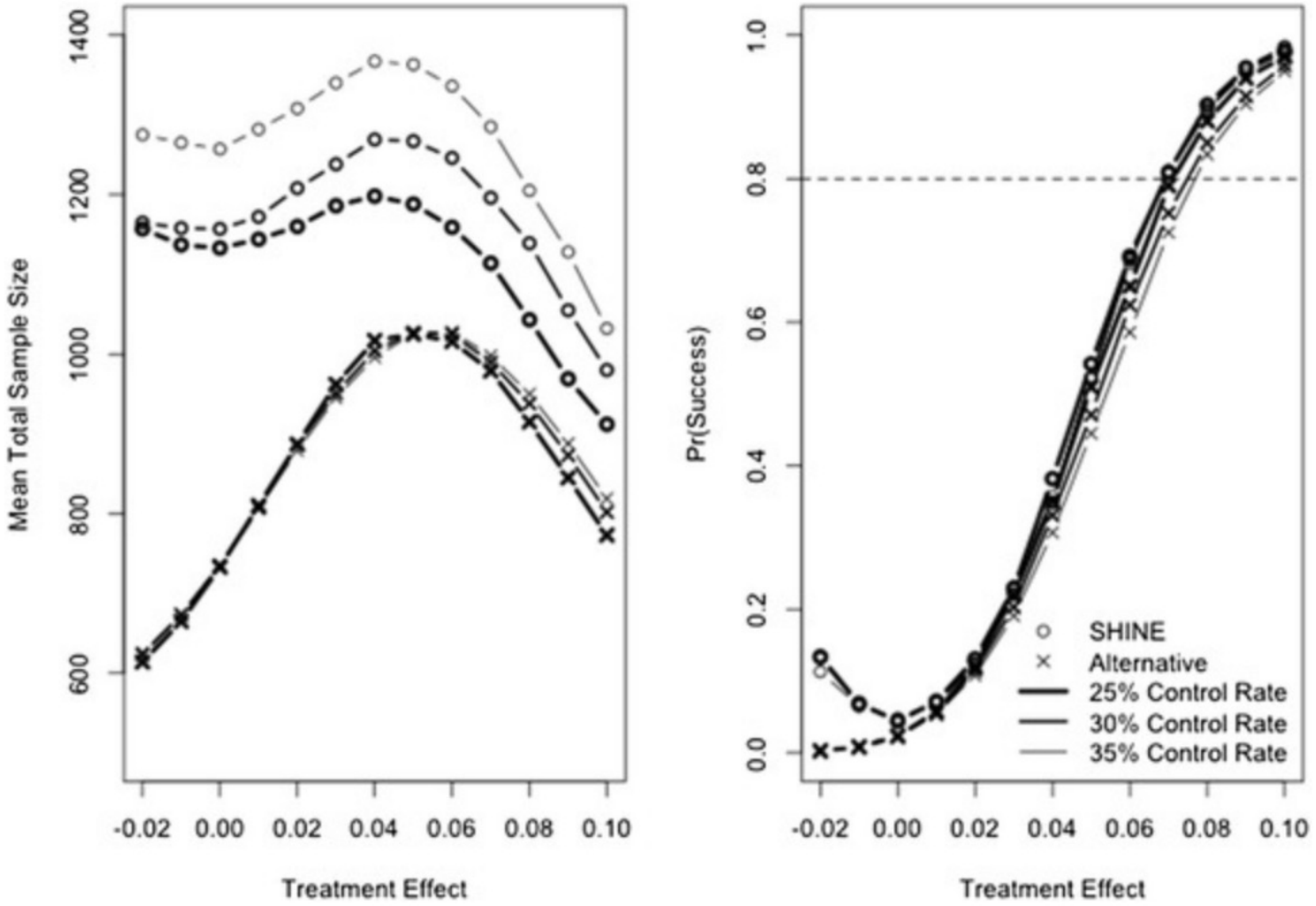
1. Low marginal cost of additional looks

- Adding an additional interim close to an existing one has a small spend
- Little additional data/highly correlated interims

2. The “flip-flop”

- Accrual is stopped, but final analysis conducted with complete follow-up
- Small chance the final analysis misses statistical success
- Reduces the overall Type I error spend
- 99% boundary chosen to minimize this risk

SHADOW maintained Power and Type I error w/ a lower mean sample size



The Results

Because we can’t run both designs

Interim Analysis Results

		SHADOW			SHINE		
N Enrolled	N Complete	Observed Treatment Effect	PPn	PPmax (<5% boundary)	Adjusted Treatment Effect	Observed P-value	Futility Boundary P-value
498	432	2.0%	<1%	21.1%			
579	515	3.4%	<1%	39.7%			
700	621	1.7%	<1%	15.6%			
800	715	0.3%	<1%	2.0%			
936	869	--	--	--	0.1%	0.957	0.652
1137	1061	--	--	--	-0.1%	0.915	0.293

Conclusions

- SHINE cross the futility boundary with 936 patients enrolled while the SHADOW design crossed the futility boundary at 800 patients enrolled

Same answer | More efficient

- Goldilocks with PPs allowed for
  - More frequent interims
  - Preserved power
- Unknown how the DMC would have reacted
  - Seeing predictive probabilities during SHINE to aid interpretation
  - An even earlier stop under SHADOW

Approximating Predictive Probabilities: A fast and accurate method for clinical trial decision making

Liz Lorenzi, PhD

May 21, 2024

Berry Consultants
Statistical Innovation

SCT | 45TH ANNUAL MEETING

Computing Bayesian predictive probabilities

Analytical Calculation

- Mathematical formula to directly calculate predictive probability
- Feasible when integral has closed form
- Very efficient

Monte carlo integration

- In cases when integral does not have closed form (e.g. time to event), requires monte carlo integration
- Can become ***computationally restrictive***

Monte Carlo Integration for Bayesian Predictive Probabilities

1. Fit Bayesian model – estimate posterior distributions of parameters for control and treated groups
2. Impute future data (remaining follow-up and/or additional patients) from each sample of the posteriors (e.g. 1000 imputed dataset)
3. On each imputed dataset, fit final analysis model
4. Summarize proportion of times you meet trial success

Monte Carlo Integration for Bayesian Predictive Probabilities

1. Fit Bayesian model – estimate posterior distributions of parameters for control and treated groups
2. Impute future data (remaining follow-up and/or additional patients) from each sample of the posteriors (e.g. 1000 imputed dataset)
3. On each imputed dataset, **fit final analysis model**
4. Summarize proportion of times you meet trial success

What if?

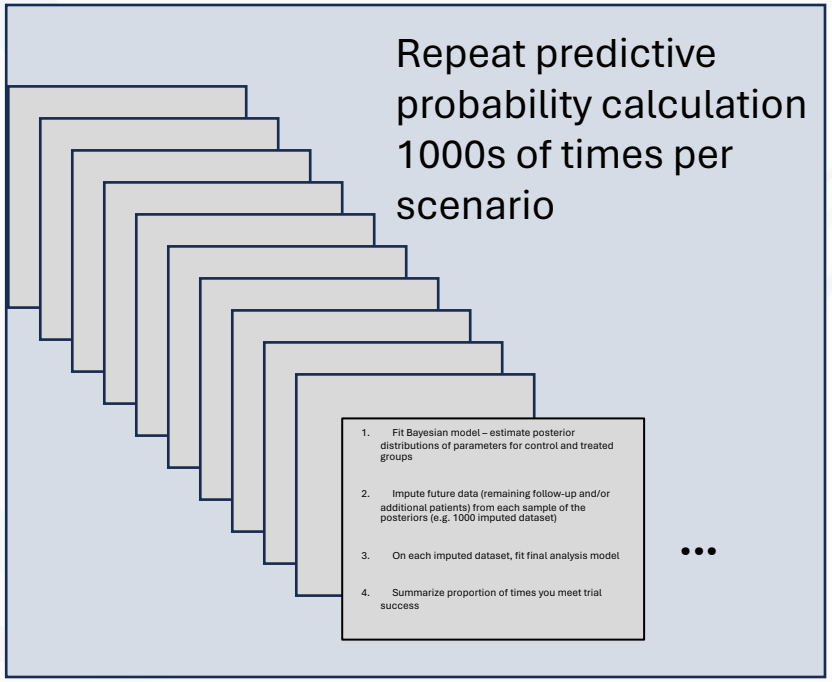
- **Complicated final analyses** – Bayesian model, ranking test (e.g. win ratio)

Monte Carlo Integration for Bayesian Predictive Probabilities

1. Fit Bayesian model – estimate posterior distributions of parameters for control and treated groups
2. Impute future data (remaining follow-up and/or additional patients) from each sample of the posteriors (e.g. 1000 imputed dataset)
3. On each imputed dataset, fit final analysis model
4. Summarize proportion of times you meet trial success

What if?

- Clinical trial simulations for operating characteristics



Computing Bayesian predictive probabilities

Analytical Calculation

- Mathematical formula to directly calculate predictive probability
- Feasible when integral has closed form
- Very efficient

Monte carlo integration

- In cases when integral does not have closed form (e.g. time to event), requires monte carlo integration
- Can become ***computationally restrictive***

Approximate

- Will introduce an alternative method that is simple and fast to calculate
- Approximates the predictive probability

Approximation of predictive probability

Interim analysis

Z_n

- n patients enrolled
- Information level is I_n
- Associated *p*-value p_n

Predictive probability PP_N is probability null hypothesis is rejected if analysis performed at N patients

Rewrite final test statistic as weighted sum of Z_n and Z_{N-n}

$$Z_N\sqrt{I_N} = Z_n\sqrt{I_n} + Z_{N-n}\sqrt{I_N - I_n}$$

Assume uninformative prior distribution $\theta \propto 1$ which yields posterior:

$$\theta|(Z_n = z_n) \sim N\left(z_n/\sqrt{I_n}, 1/I_n\right)$$

Results in predictive distribution for Z_{N-n}

$$Z_{N-n}\,|(Z_n = z_n) \sim N\left(z_n\sqrt{\frac{I_N - I_n}{I_n}}, \frac{I_N}{I_n}\right)$$

$$PP(p_n, r, \alpha) = \Phi\left(\frac{\Phi^{-1}(1-p_n) - \Phi^{-1}(1-\alpha)\sqrt{r}}{\sqrt{1-r}}\right)$$

Final analysis

Z_N

- N patients enrolled
- Information level is I_N
- Associated *p*-value p_N

Berry Consultants

Approximation of predictive probability

- Take your interim analysis run
  - Use p-value (p_n) or posterior probability of treatment effect
  - Summarize current information I_n and total information I_N
    - Information fraction $r = I_n/I_N$
  - Compute following equation

$$PP(p_n, r, \alpha) = \Phi\left(\frac{\Phi^{-1}(1-p_n) - \Phi^{-1}(1-\alpha)\sqrt{r}}{\sqrt{1-r}}\right)$$

Applying the approximate PP

- Key assumption is that the test statistic is approximately Gaussian and the information fraction $r = I_n/I_N$ is known

Endpoint	Example analyses	I_n	I_N
Continuous	T-tests ANOVA/ANCOVA	Interim sample size	Final sample size
Discrete/categorical	z-tests Chi-squared tests	Interim sample size	Final sample size
Time-to-event	Log-rank test Proportional hazards models	Events at interim	Events at final
Ordinal/Non-parametric	Ordinal regression Wilcoxon rank-sum	Interim sample size	Final sample size
Count data	Generalized linear regressions (e.g. Poisson regression)	Interim exposure	Final exposure

Applying the approximate PP

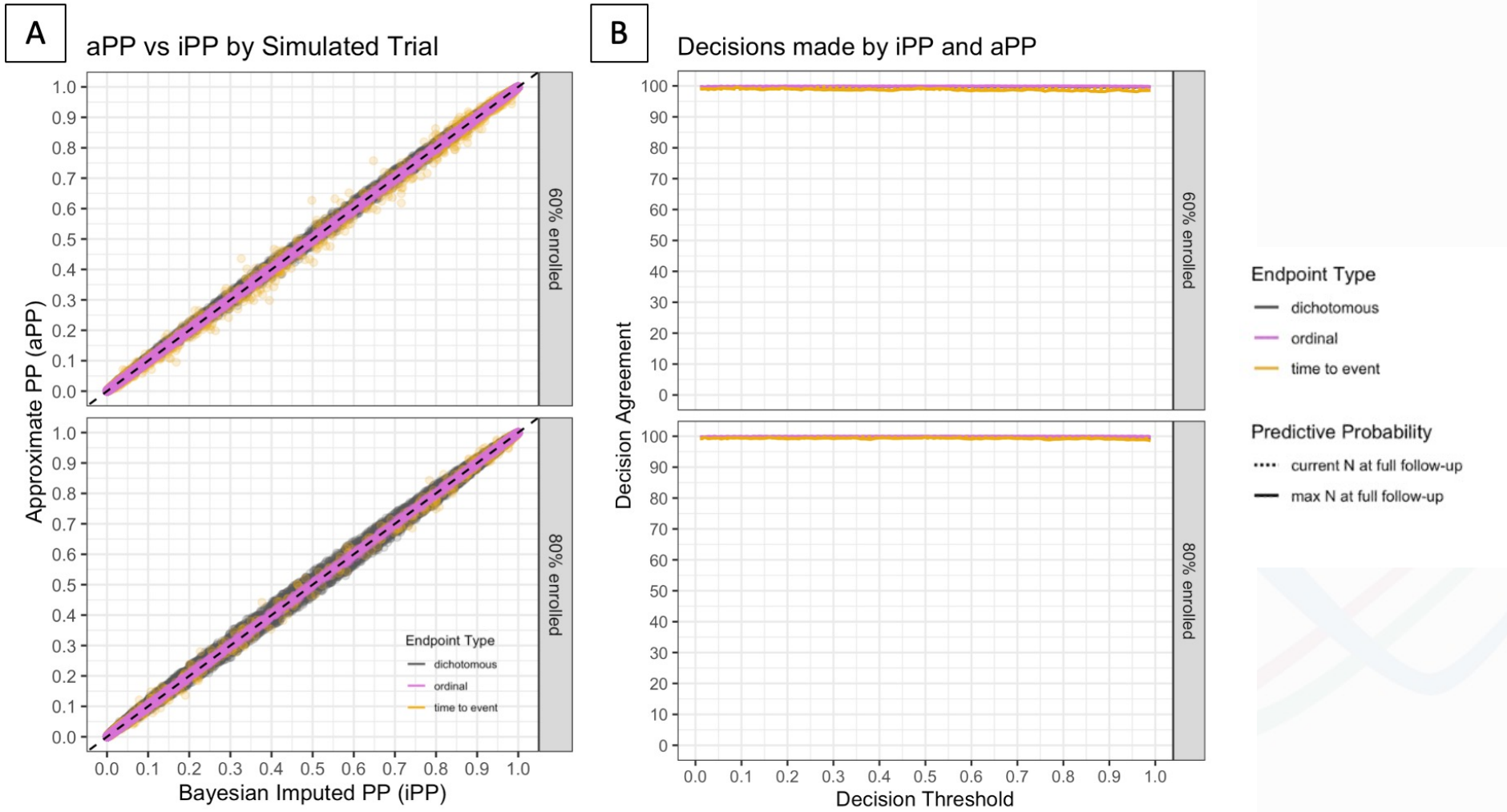
- Key assumption is that the test statistic is approximately Gaussian and the information fraction $r = I_n/I_N$ is known

Endpoint	Example analyses	I_n	
Continuous	T-tests ANOVA/ANCOVA	Interim sample size	Final
Discrete/categorical	z-tests Chi-squared tests	Interim sample size	Final
Time-to-event	Log-rank test Proportional hazards models	Events at interim	Event
Ordinal/Non-parametric	Ordinal regression Wilcoxon rank-sum	Interim sample size	Final
Count data	Generalized linear regressions (e.g. Poisson regression)	Interim exposure	Final

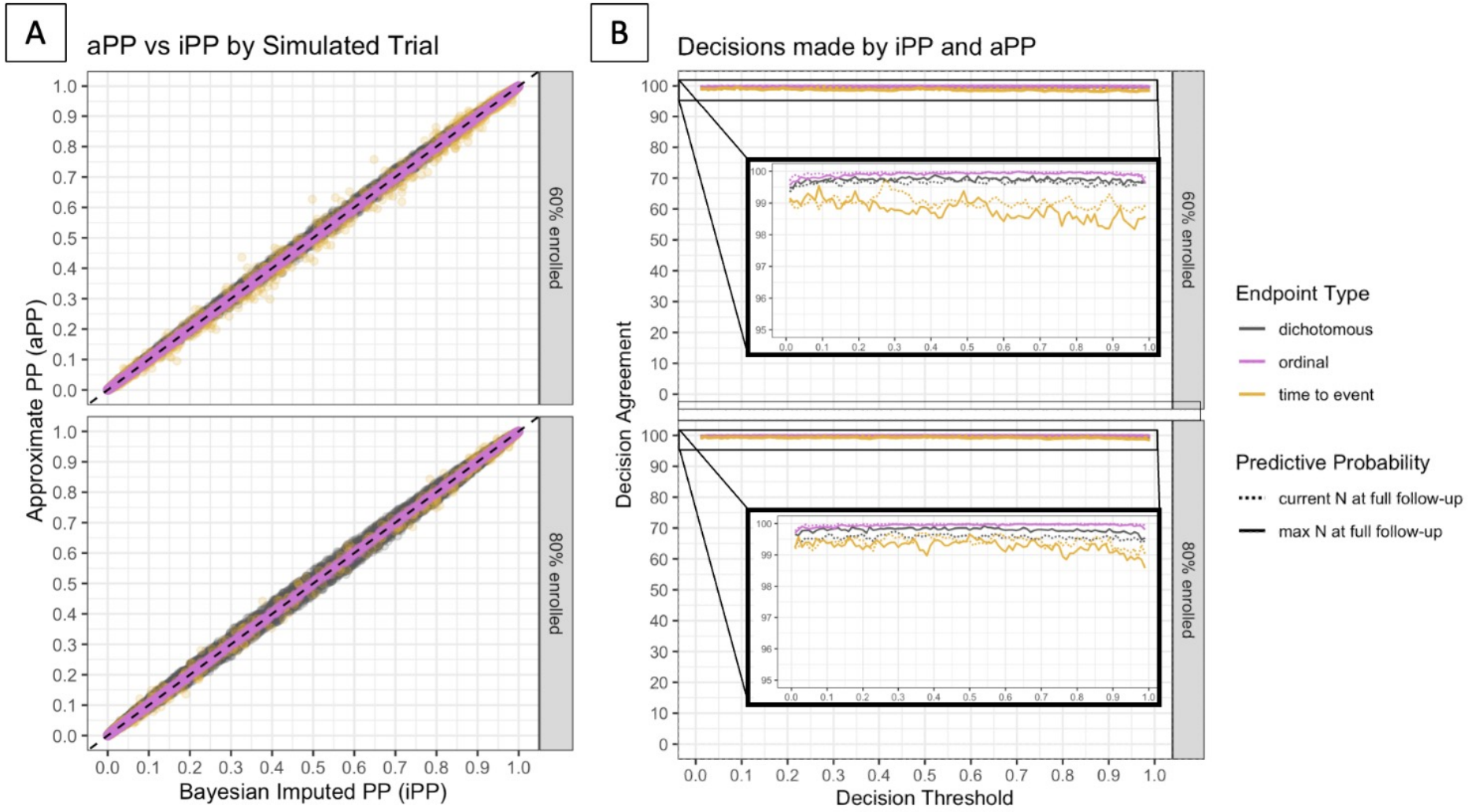
Simulations

- Assume Interims occur at 60% information and 80% information
- Calculate Bayesian imputed predictive probabilities (iPP) and approximate predictive probabilities (aPP)
- Compare similarity and whether they lead to similar trial decisions

Simulations – comparing iPP and aPP across endpoint types



Simulations – comparing iPP and aPP across endpoint types



Example – stroke trial using historical cohort

- Primary endpoint: 90 day mRS
- Primary analysis: Bayesian proportional odds model with borrowing on treatment effect
- Max sample size 1500
- Interims at 500, 750, 1000 and 1250 for
  - stop trial enrollment for expected success if $PP_n > 90\%$
  - stop trial enrollment for futility if $PP_N < 10\%$
- Borrowing from historical cohort of 500 total subjects (250 control, 250 treated)
  - Dynamic borrowing through hierarchical prior (Viele, et al 2014)

Computational challenge of computing pred probability in example

- Example requires Monte Carlo integration
  - At final step when running final analysis on every imputed dataset, would need to run Bayesian model
  - Extremely computationally expensive
  - Combining calculation with clinical trial simulations could be unrealistic

- Perfect case for our aPP
  - EXCEPT…. How do we compute information?

Approximating information

- Dynamic borrowing from external cohort means we don't know how much current information we have
  - Dynamic borrowing uses a variance parameter that is estimated to determine the degree of borrowing
  - When external and current data more similar, borrows more from external. When more different, borrows less

- Solution – approximating information using *effective sample size*

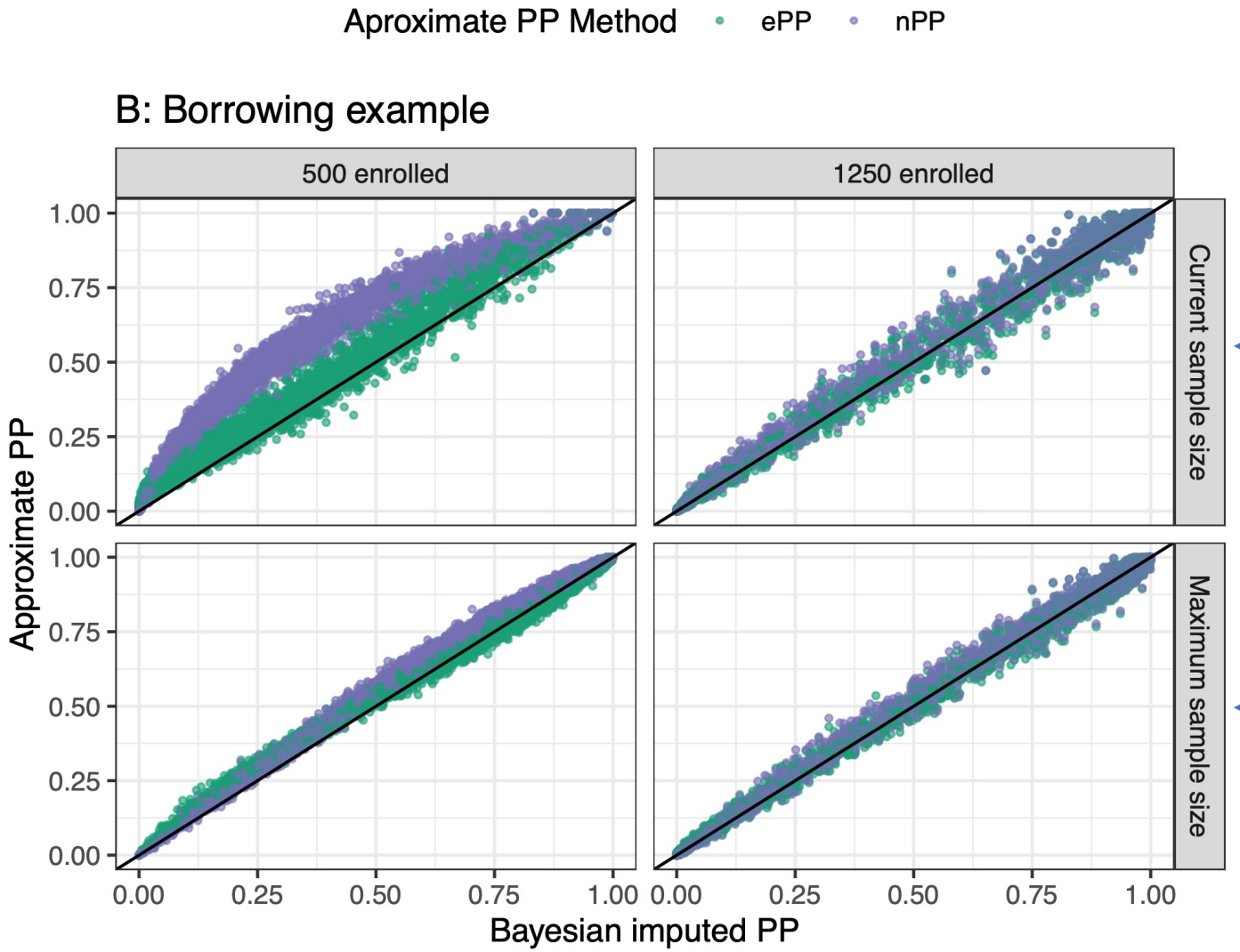
$$\hat{I} = \frac{\mathrm{Var}(\tilde{\theta})}{\mathrm{Var}(\hat{\theta})}\tilde{I}$$

$\tilde{\theta}$: estimate of treatment effect without borrowing
$\hat{\theta}$: estimate of treatment effect with borrowing

How do aPP and iPP compare?

ePP – predictive probability using estimate of information

nPP – predictive probability using nominal information (ignoring information from external cohort)



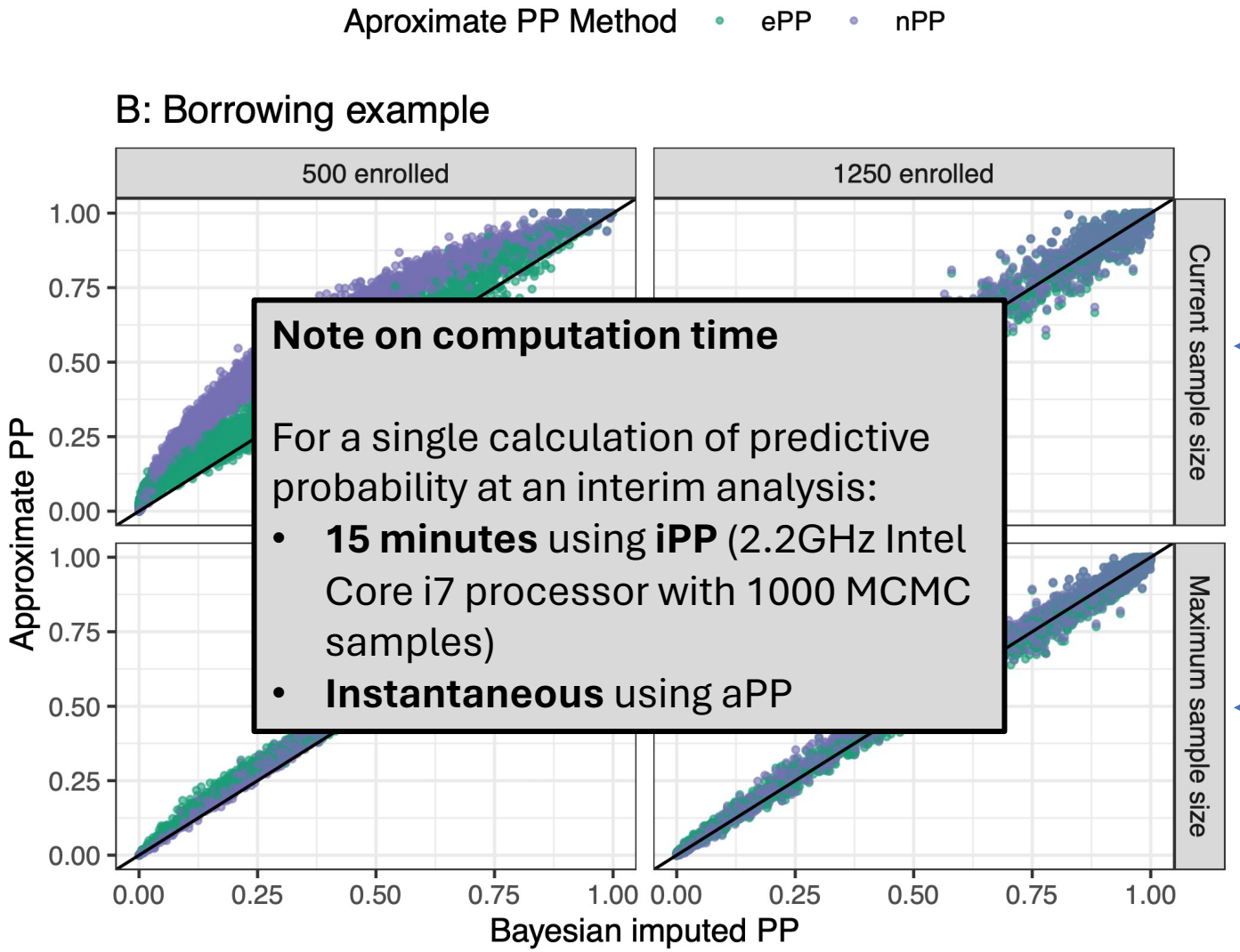
Predictive probability of success at current sample size

Predictive probability of success at maximum sample size

How do aPP and iPP compare?

ePP – predictive probability using estimate of information

nPP – predictive probability using nominal information (ignoring information from external cohort)



Note on computation time

For a single calculation of predictive probability at an interim analysis:

- **15 minutes** using **iPP** (2.2GHz Intel Core i7 processor with 1000 MCMC samples)
- **Instantaneous** using aPP

Predictive probability of success at current sample size

Predictive probability of success at maximum sample size

Final notes

- aPP can make predictive probabilities more accessible to more researchers
- Impressed with similarity between iPP and aPP
  - Though there are cases where they disagree
- Impactful in design phase – reduces computational burden of pred prob with clinical trial simulations
  - If used in design stage, should also use in implementation stage
  - Want the thresholds used/operating characteristics of design to reflect the design you are running

Acknowledgements

- Work is currently under review
- Joint work with Joe Marion, Cora Allen-Savietta, Kert Viele, and Scott Berry

Berry Consultants

Statistical Innovation

CONTACT INFORMATION

elizabeth@berryconsultants.com

WEBSITE

berryconsultants.com

YOUTUBE

youtube.com/berryconsultants

TWITTER

@berryconsultant

A Case Study for Predictive Probabilities:
Biosense Webster inspIRE Study

Giorgio Paulon, PhD
May 21, 2024

2024
BOSTON

SCT | 45TH ANNUAL MEETING

Disclosures

- Dr. Paulon is an employee of Berry Consultants, LLC, a statistical consulting firm that specializes in the design, conduct, oversight, and analysis of adaptive and platform clinical trials.
- Berry Consultants designed a Goldilocks study combined with early success for Biosense Webster. Slight variations of this initial design have been used for several J&J trials.
- The inspIRE trials was one of these design iterations, and a Statistical Analysis Committee (SAC) at Berry Consultants implemented this trial.

Study Design

Key Eligibility Criteria

- Adult patients (aged ≤75 years)
- Drug-refractory symptomatic PAF
- First-time PVI

Institutions and Timing

- 13 institutions across Canada & Europe
- March 2021 to May 2023

Ablation Procedure

- PVI performed and confirmed via entrance block (no waiting period)
- Subjects followed up for 12 months

Rhythm Monitoring	
	Remote rhythm monitoring: Months 3-6 (weekly), Months 6-12 (monthly), and after any symptomatic episodes
	24-hour Holter monitoring: Months 3, 6, and 12
	ECG: preprocedure, predischarge, and Months 1, 3, 6, and 12

Endpoints

- **Safety:** primary AE rate and device- and procedure-related serious AEs for patients completing the study period
- **Primary effectiveness endpoint (PEE):** acute PVI plus freedom from all atrial arrhythmia (AF/AT/AFL) at days 91–365
- **Sub-analyses:** predictors of PEE success, PV reconnection at repeat procedures

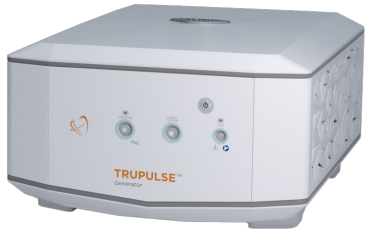
PAF: paroxymal atrial fibrillation; PVI: pulmonary vein isolation; AF: atrial fibrillation; AFL: atrial flutter; AT: atrial tachycardia

Berry Consultants

The VARIPULSE™ Platform

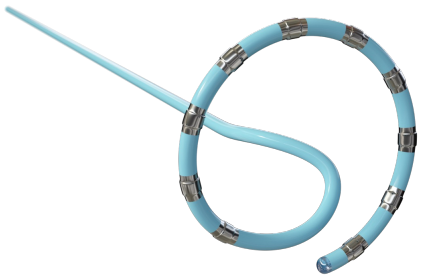
VARIPULSE™ Module

- CARTO™ 3 system integration enabling anatomical mapping, ultrasound catheter mapping, and real-time electrogram feedback



TRUPULSE™ Generator

- Creates pulsed electrical fields by delivering bipolar and biphasic energy through electrode pairs



VARIPULSE™ Catheter

- Circular tip section of the catheter can be expanded and contracted to accommodate the anatomy of the LA
- Facilitates cardiac electrophysiological mapping (stimulating and recording) and ablation

Berry Consultants

Trial Characteristics

- Prospective, multicenter, single-arm clinical trial (comparison to PG)
  - PG for effectiveness: 50%
  - PG for safety: 14%
- Registration trial (pre-market evaluation)
- Treatment: PVI was performed with the VARIPULSE™ Platform to obtain Irreversible Electroporation (IRE)

Final Analysis:

$$H_0: p_E \leq 0.5 \qquad H_a: p_E > 0.5$$

$$P(p_E > 0.5) > 0.9775$$

Goals

Demonstrate safety and effectiveness of the irreversible electroporation (IRE) system when used in treatment of participants' drug refractory, symptomatic paroxysmal atrial fibrillation (PAF) for isolation of the atrial pulmonary veins

Uncertainty in Trial Designs

- Most clinical trials have substantial uncertainty in design stage
  - How best to treat subjects? What is the best measure of benefit?
  - Event rates, optimal dose, best duration, target population

Traditional approach

- We make our best guesses and run the trial, all key trial parameters are defined upfront and held constant throughout the trial
- If assumptions are wrong, greater risk of a failed trial
- Does not take advantage of accruing data

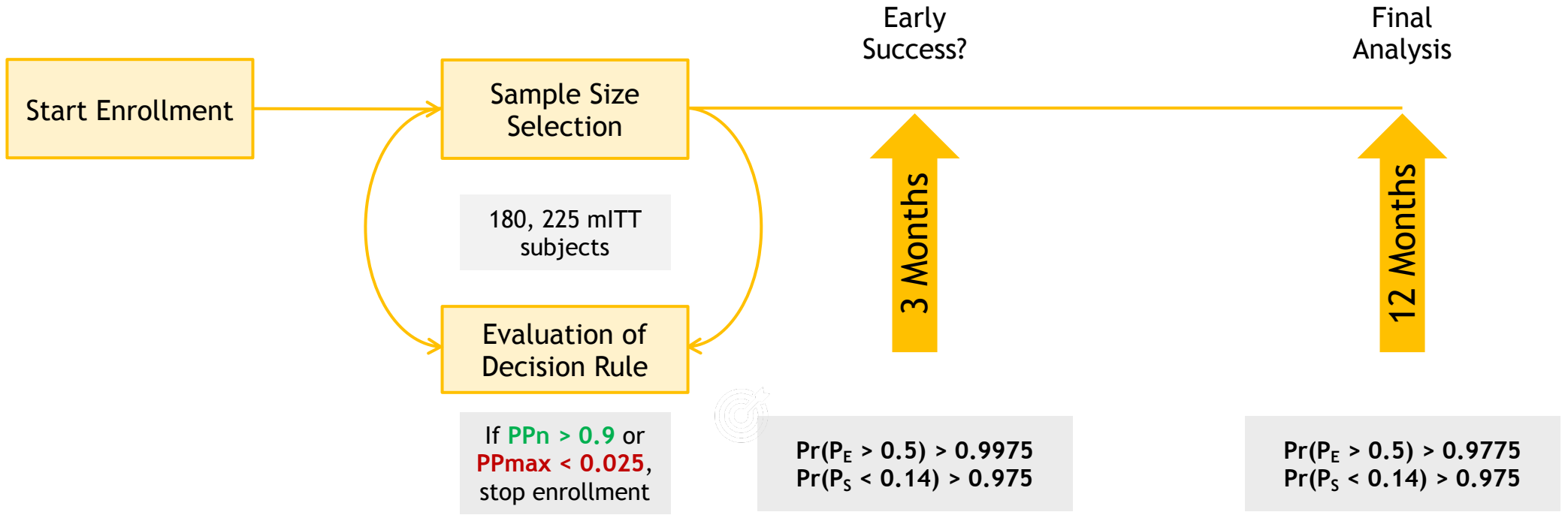
Adaptive design

- After enrollment begins and some outcomes are known, there is less uncertainty about many of these key parameters
- Adaptive trials allow modification to key trial parameters in response to accumulating data and according to *prespecified* rules

- Learning the 'optimal' design
  - Can increase the probability of getting the right answer at the end of the trial
  - Can treat patients more effectively within the trial

Trial Schema

Maximum sample size: 330 main study subjects



PP_n: predictive probability of trial success (effectiveness) at the current sample size
PP_{max}: predictive probability of trial success (effectiveness) at the maximum sample size
P_E: rate for the effectiveness endpoint
P_S: rate for the safety endpoint

Key Features of Predictive Probabilities

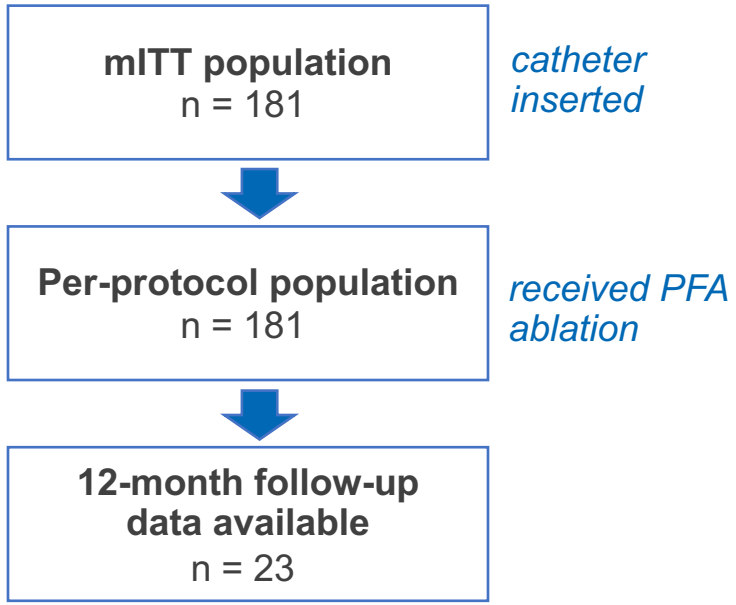
Saying “Predictive Probability” is not enough. Details include:

- Dropout assumption:
  - Currently observed
  - Assumed in design
- Future accrual:
  - Based on observed across trial
  - Based on observed in last X weeks
  - Assumed in design
- Algorithm:
  - Simulation
  - Exact formula
  - Approximation

Consistency between design and implementation of the trial is key

Interim 1: Patient Population

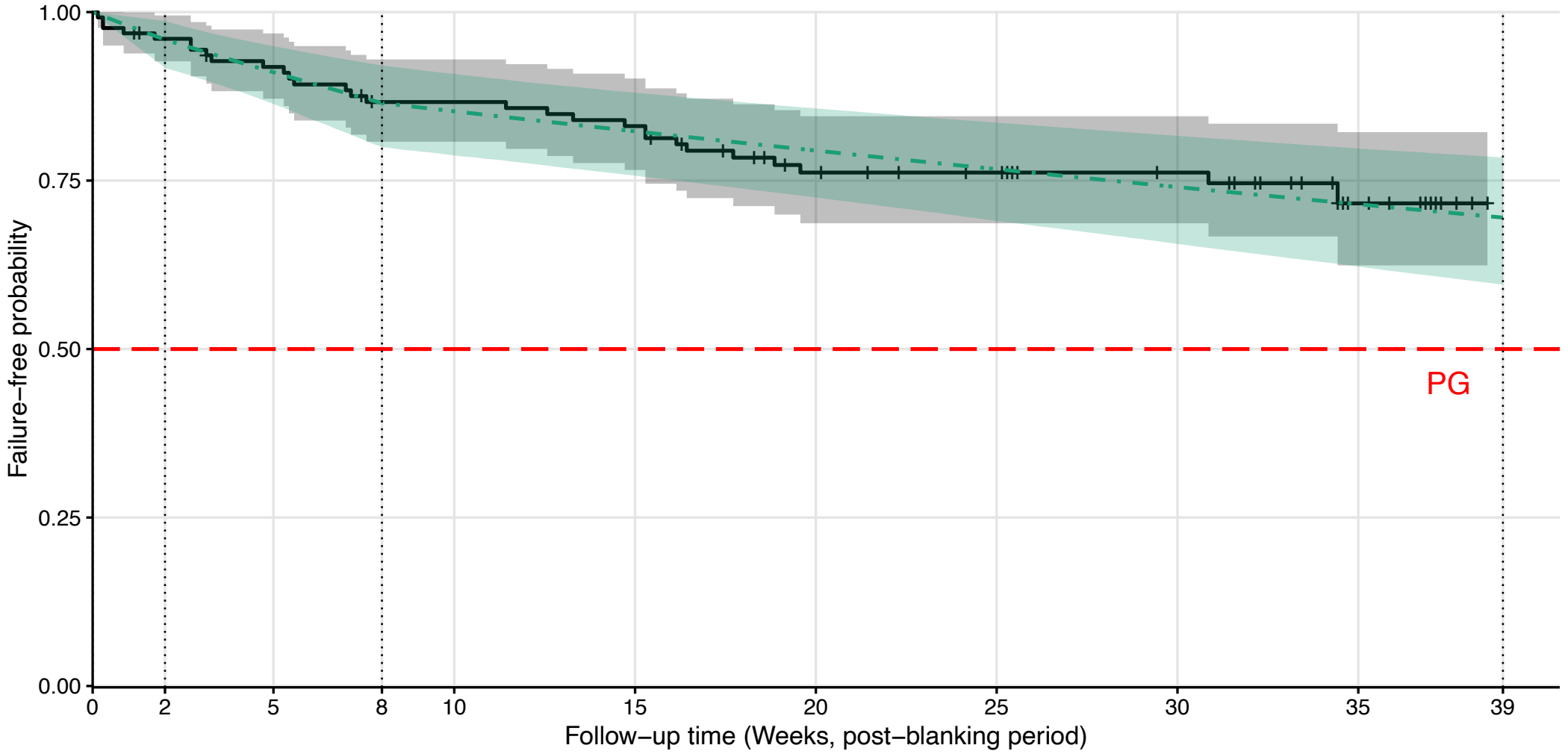
PG for effectiveness: 50%



	Treatment ($N = 181$) **n (%)**
Completed follow-up[a]	23 (13)
Number of responders	13 (57)
Number of non-responders	10 (43)
Not completed follow-up	158 (87)
Ongoing subjects	158 (100)
Not failure for effectiveness endpoint	139 (88)
Failure for effectiveness endpoint	19 (12)
Withdrawn / LTFU / Deceased[b]	0 (0)
Exposure time (weeks)[c]	
Median (IQR)	15.4 (0.0, 31.4)
≥ 2 weeks	118
≥ 8 weeks	97

[a] Follow-up refers to the 12-month follow-up visit, i.e. 9-month post-blanking visit (protocol visit window day 335 – day 379).
[b] Withdrawals occurring after the failure events will not be counted as such.
[c] Time from end of blanking period to failure event date, early withdrawal date, end of follow-up, or data freeze date, whichever comes first.

Interim Analysis 1 (Sample Size Selection)



Berry Consultants

Predictive Probability Algorithm

Steps (repeated ~10,000 times):

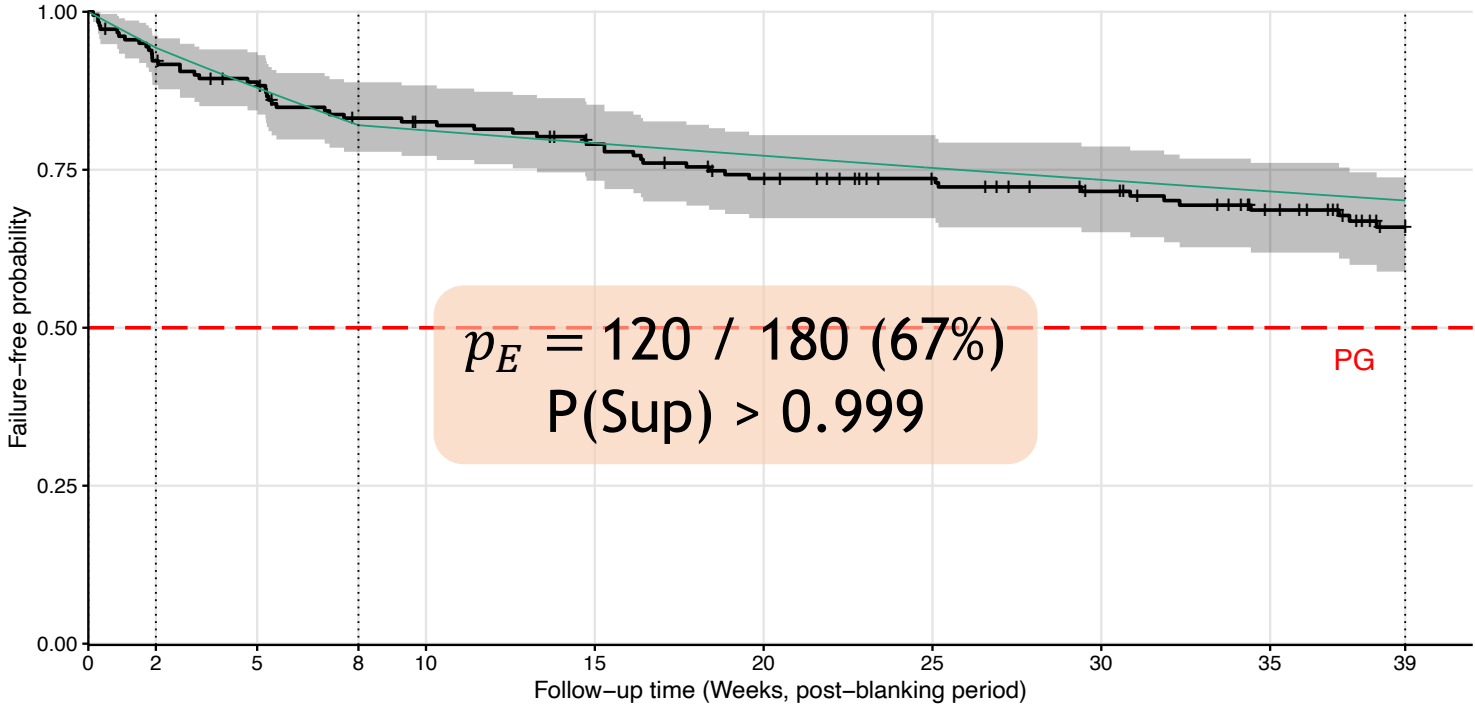
1. One sample, denoted $\lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_3^{(k)}$ is drawn from the posterior distribution.
2. Future subjects' enrollment times are simulated (*which rate?*).
3. A "complete" dataset is created: Unknown event times are imputed for all future subjects and all currently enrolled subjects still in follow-up (without an event and not dropped out) from a piecewise exponential distribution conditional on the current censoring time using the sample $\lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_3^{(k)}$. A censoring time is also drawn for each subject from an exponential distribution (*which rate?*).
4. The final analysis is performed on the "complete" dataset.

Predictive probability of success: proportion of "complete" datasets leading to statistical significance.

Predictive Probability Algorithm

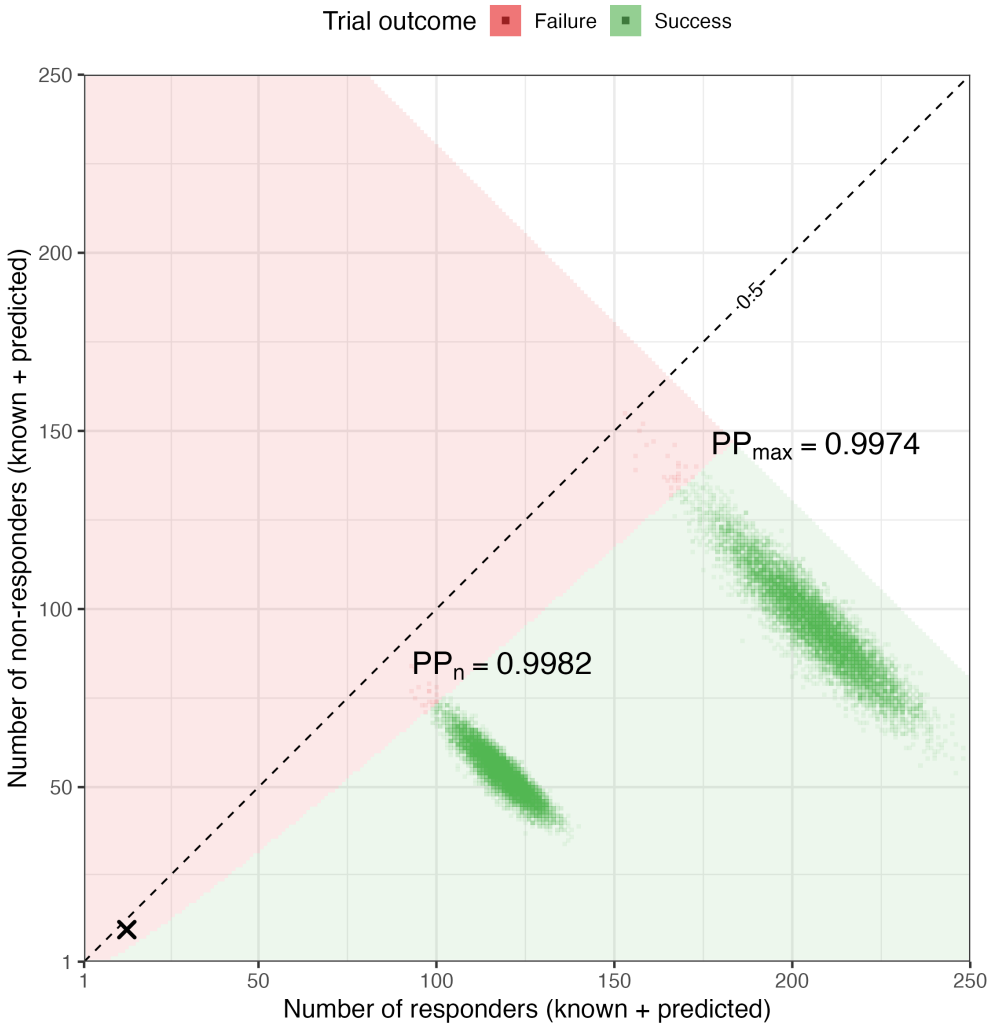
Steps (repeated ~10,000 times):

4. The final analysis is performed on the "complete" dataset.d all currently enrolled subjects still in follow-up (without an event and not dropped out) from a piecewise exponential distribution conditional on the current censoring time. A censoring time is also drawn for each subject from an exponential distribution (*which rate?*).



Predictive probability of success: proportion of "complete" datasets leading to statistical significance.

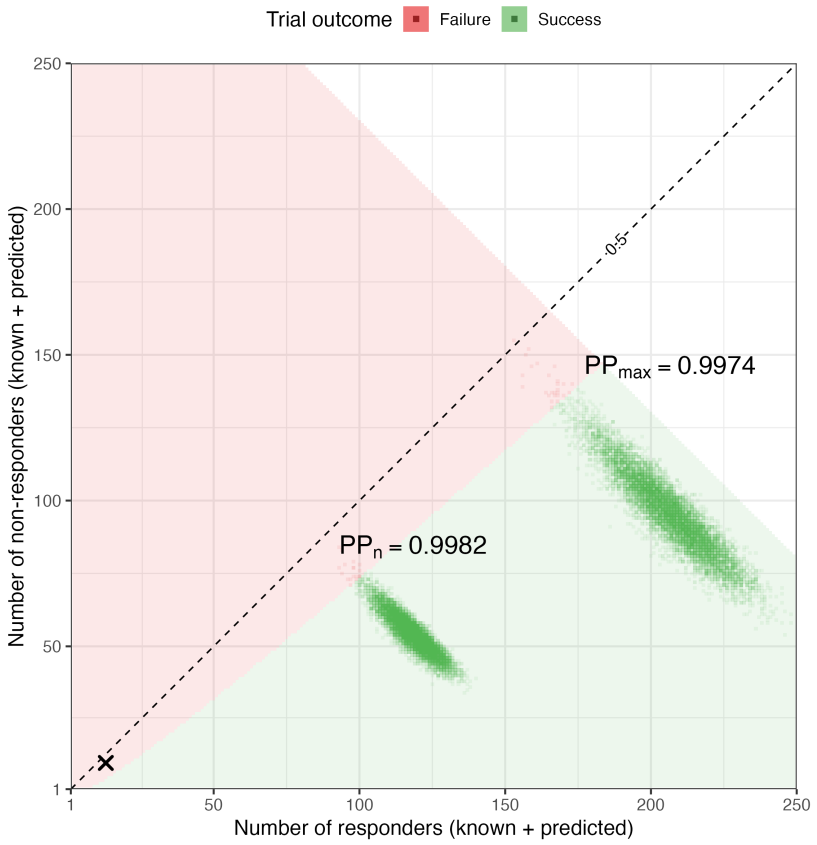
Predictive Probabilities



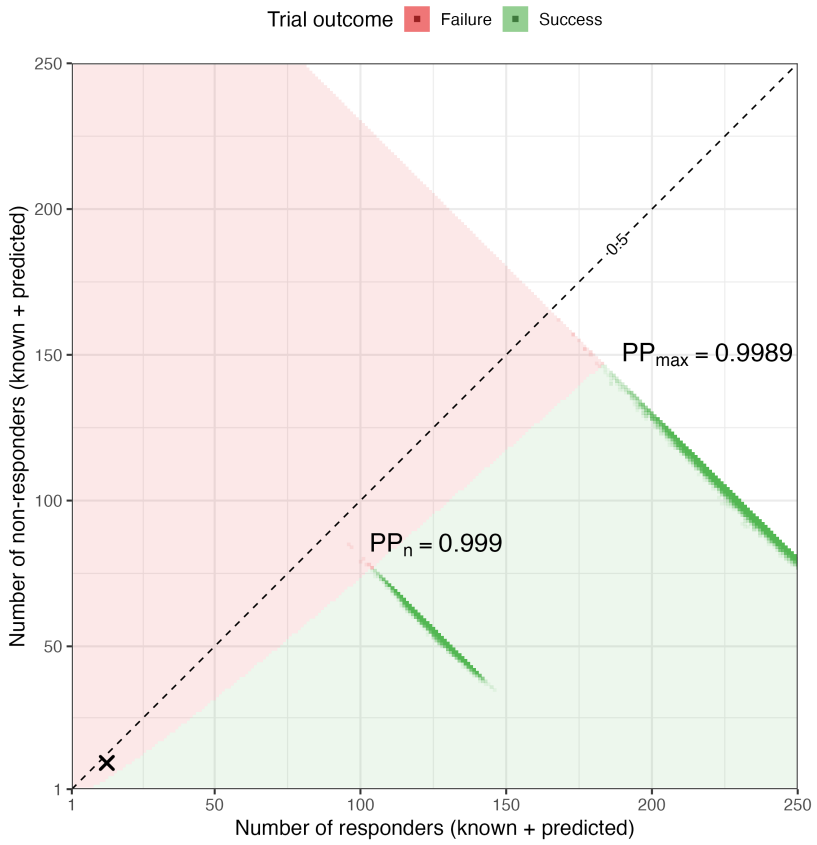
Berry Consultants

Predictive Probabilities - Sensitivity

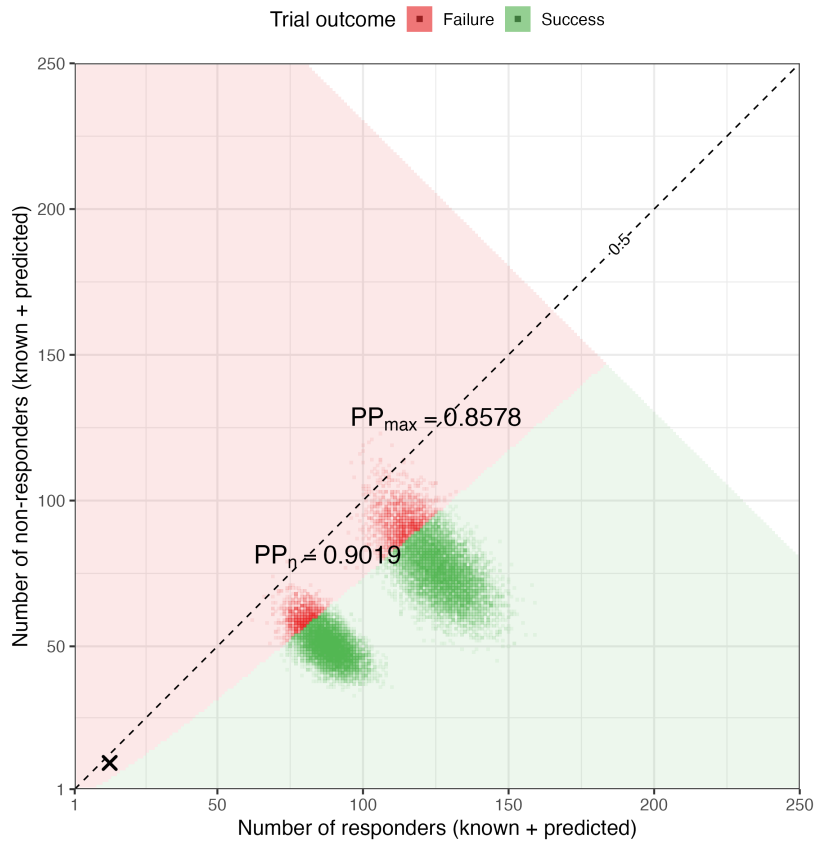
15% design assumed dropout rate



0.5% observed dropout rate

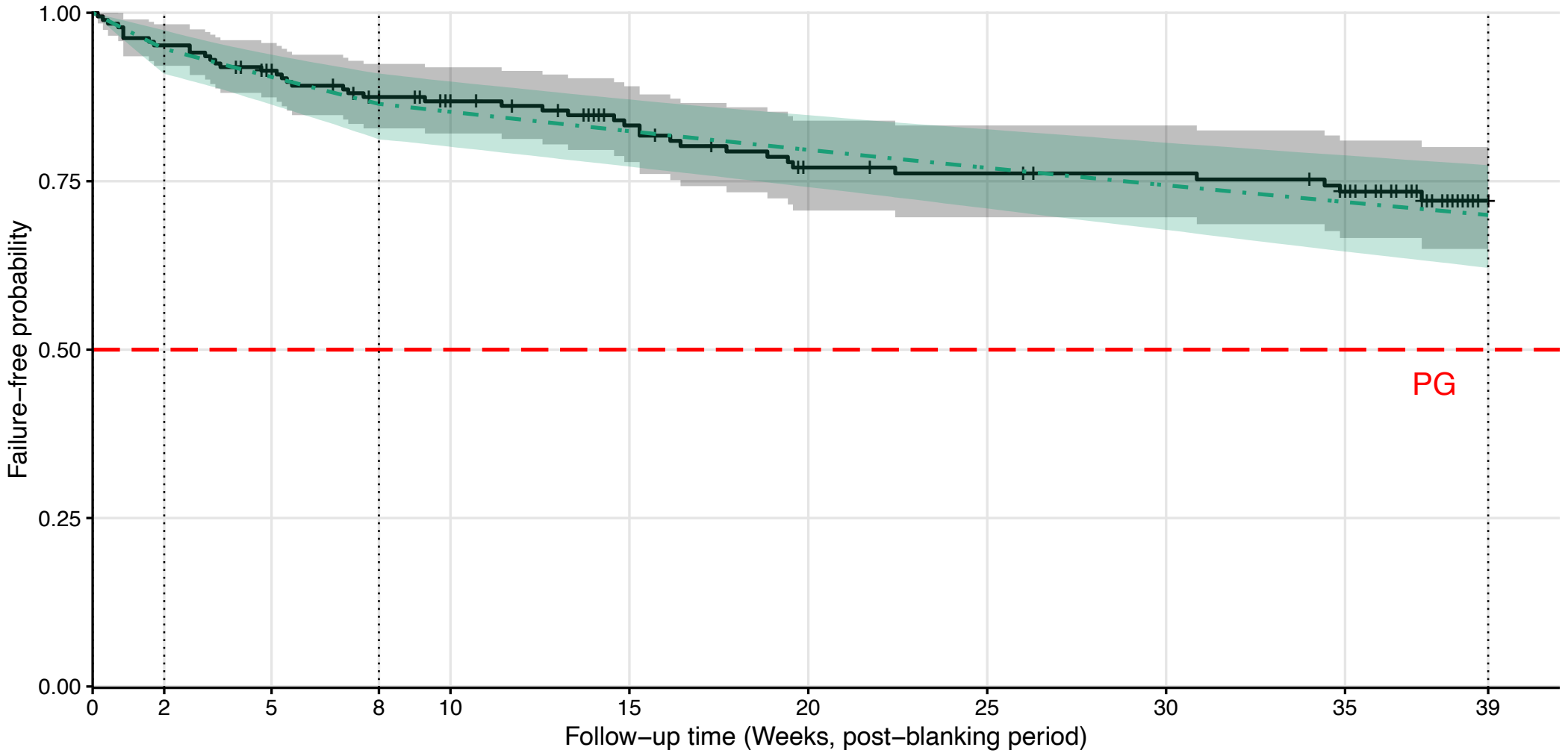


64% assumed dropout rate (tipping point)



Berry Consultants

Interim Analysis 2 (Early Success)

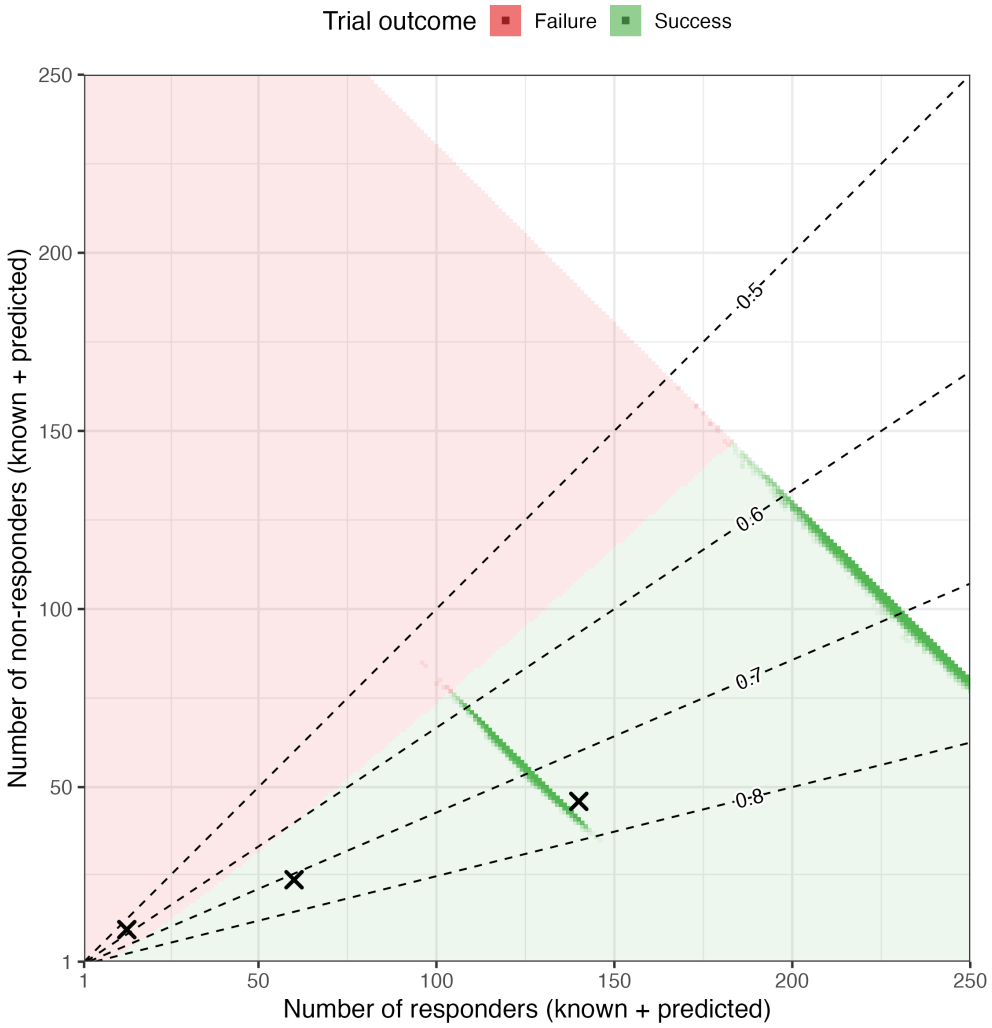


Berry Consultants

Trial History

Observed effectiveness:

- Sample Size Interim: 56.5% (13 / 23)
- Early Success: 71.4% (60 / 84)
- Final Data: 75.7% (140 / 185)



Conclusions

- Predictive Probabilities can make interim results more interpretable, especially when not all subjects have the same follow-up
- Predictive Probabilities can be calculated in different ways
- It is important that there is consistency between design and implementation of the trial
- It is also crucial that events are collected at the same rate across sites

Berry Consultants

Statistical Innovation

CONTACT INFORMATION

info@berryconsultants.com | (512) 213-6428

WEBSITE
berryconsultants.com

YOUTUBE
youtube.com/berryconsultants

TWITTER
@berryconsultant

Predictive Probabilities - A clinical PI perspective

Will Meurer
University of Michigan, Ann Arbor

Disclosures

Twitter: @goalsofhair @willmeurer

insta: @goals.of.hair

- Consulting fees for medicolegal cases / Berry Consultants (various commercial clients – I am not on any projects related to cooling)
- Funding from NINDS
  - -PI of clinical trials methodology course
  - Funded co-investigator on NETT/SIREN-CCC and BOOST trial
- NIH-NIDCD (cluster RCT studying dizziness intervention)
- NIH-NIMHD - PI of trial of hypertension
  - Co-investigator to improve stroke care in Flint
- NIH-NHLBI (ICECAP PI and P-ICECAP PI/ also co-investigator in cardiac arrest expedited transport/ECMO trial)
- AHRQ (prehospital stroke care / dizziness self management)
- Massey Foundation (studying pupillary response in TBI) - past
- FDA and NIH (reviewer)
- Meth Stats Reviewer
  - Annals Of Emergency Medicine
  - Academic Emergency Medicine
- Senior Editor (paid)
  - BMC Journal - Trials
- PCORI (co-I Kidney Stone Trial)

Disclosure

Imagery created using prompts to ChatGPT4o

Imagery designed to maximize alertness after 4:30 pm on 2nd day of conference…

I wear a lot of hats

PI

Co-Investigator

DSMB / DMC member

Trial Designer / consultant

Teacher of trials

SIREN Clinical Trials Network

4 Ongoing Trials (Multicenter)

Max N: 1800, 900, 200

1(and a half) designed with ADAPT-IT

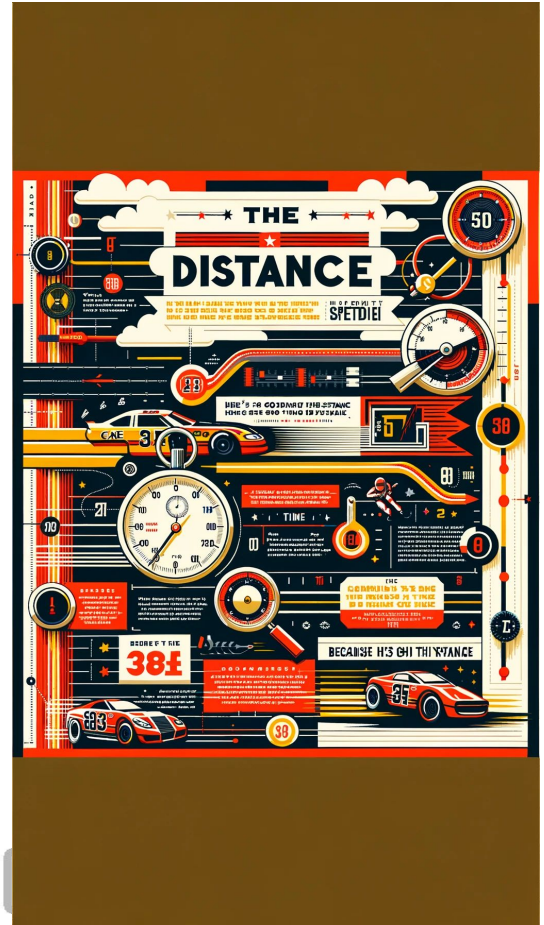
Needs as a clinical PI

-Confidence

-Is the question answered?

-Is the question answerable?

-Should it go the distance?



Needs on a DMC

-Quantitative guidelines

-Think about patients

(Inside, outside, future)

-Predict limitations

Needs as teacher

-Accessible method

-Can convince academics to use it

-Previous publications

Discussion:

Audience!!!!

Discussion:

If audience is quiet

-How to get more uptake?

-Barriers?

-What new methods could help?