Utrecht Oct 2023

Coping with Information Loss and the Use of Auxiliary Sources of Data

A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions

Silvia Calderazzo, Sergey Tarima, Carissa Reid, Nancy Flournoy, Tim Friede,
Nancy Geller, James L Rosenberger, Nigel Stallard, Moreno Ursino, Marc Vandemeulebroecke,
Kelly Van Lancker, Sarah Zohar

Context

The NISS Special Series: The NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions

The NISS Ingram Olkin Forum on Unplanned Clinical Trial Disruptions

Nancy Flournoy

Pages 92-93 | Received 19 O

Research Article

Coping with Information Loss and the Use of Auxiliary Sources of Data: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions

Silvia Calderazzo, Sergey Tarima, Carissa Reid, Nancy Flournoy, Tim Friede, Nancy Geller, James L Rosenberger, Nigel Stallard, Moreno Ursino, Marc Vandemeulebroecke, Kelly Van Lancker & Sarah Zohar ...show less

Received 23 Jun 2022, Accepted 11 Apr 2023, Published online: 26 Jun 2023

The NISS Special Series: The NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions

Estimands and their Estimators for Clinical Trials Impacted by the COVID-19 Pandemic: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions

Kelly Van Lancker, Sergey Tarima, Jonathan Bartlett, Madeline Bauer, Bharani Bharani-Dharan, Frank Bretz, Nancy Flournoy, Hege Michiels, Camila Olarte Parra, James L. Rosenberger & Suzie Cro ...show less

Pages 94-111 | Received 07 Feb 2022, Accepted 22 Jun 2022, Published online: 14 Sep 2022

Research Article

Using Randomization Tests to Address Disruptions in Clinical Trials: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions

Diane Uschner, Oleksandr Sverdlov, Kerstine Carter, Jonathan Chipman, Olga Kuznetsova, Jone Renteria, Adam Lane, Chris Barker, Nancy Geller, Michael Proschan, Martin Posch, Sergey Tarima, Frank Bretz & William F. Rosenberger ...show less

Received 08 Oct 2022, Accepted 06 Aug 2023, Accepted author version posted online: 18 Oct 2023

Context

COVID-19 pandemic had a disruptive effect on many ongoing clinical trials

- around 80% of **non-**COVID-19 trials have been stopped or interrupted
- not anymore statistical power to yield interpretable results

Context

COVID-19 pandemic had a disruptive effect on many ongoing clinical trials

- around 80% of **non-**COVID-19 trials have been stopped or interrupted
- not anymore statistical power to yield interpretable results

Beyond COVID-19, Fogel et al. 2018

- failure in patients' recruitment in 25% of cancer trials
- 18% of trials closed with less than half of the target sample size
- 22% of the failed phase 3 studies failed due to lack of funding

D. B. Fogel. Contemp Clin Trials Commun, 11:156–164, Sep 2018

Objective

Hypothesis:

- augmenting the trial data with **auxiliary data** will allow the trialists stakeholders to obtain an answer to the primary scientific and medical question

Aim:

- propose how to cope with information loss in the context of interrupted and stopped RCT by using **auxiliary sources**

Auxiliary sources

Internal

External

Auxiliary sources

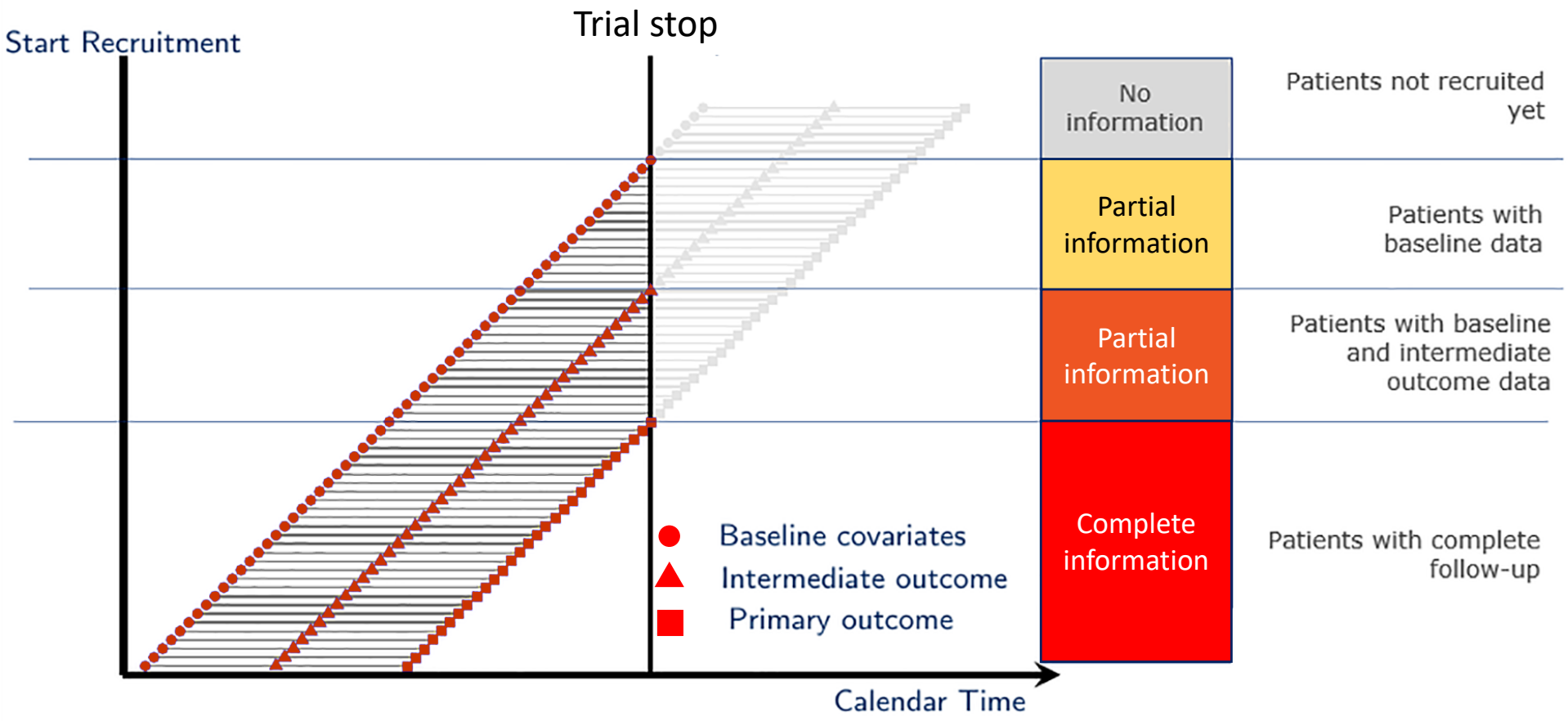
Internal

auxiliary information is available from the patients in the trial itself:

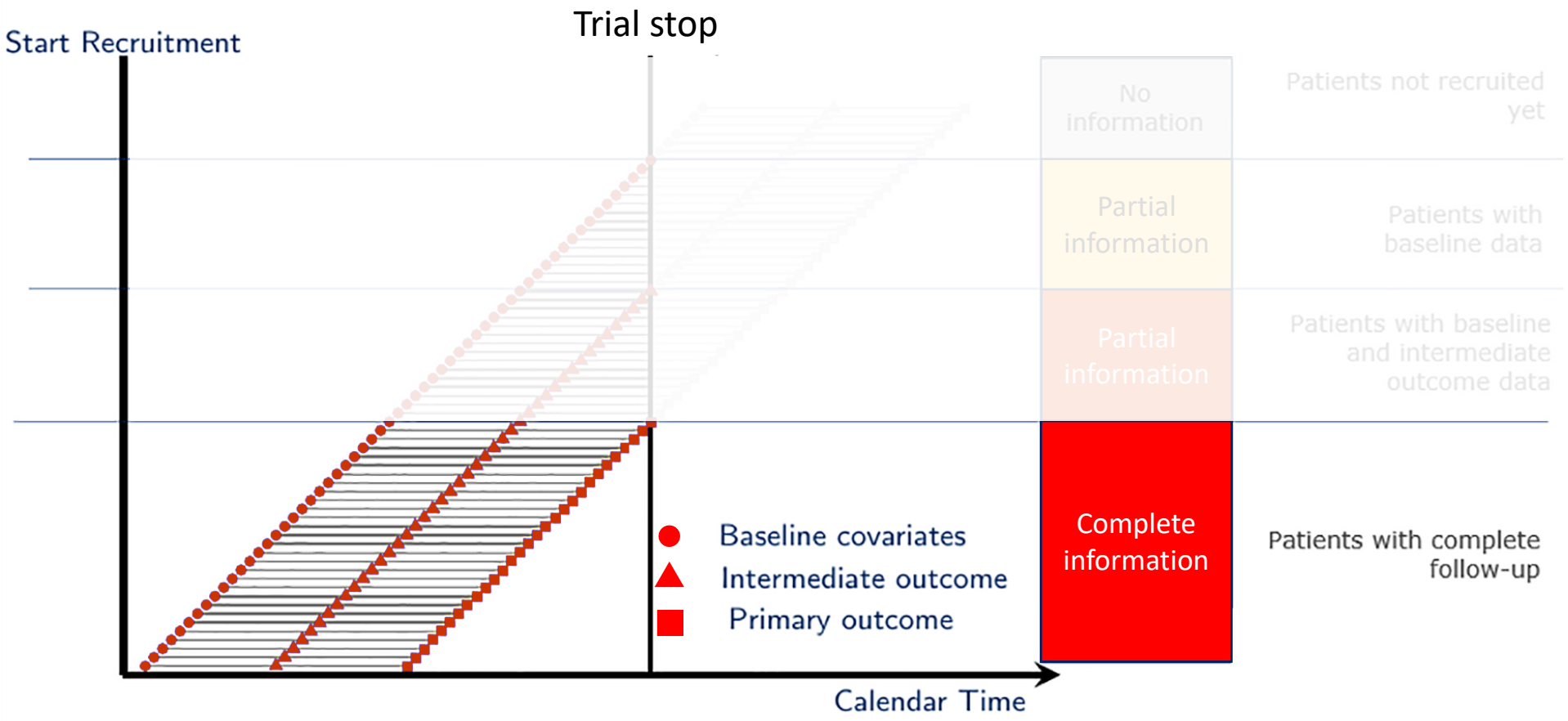
early or baseline data in inference on the primary endpoint of interest.

External

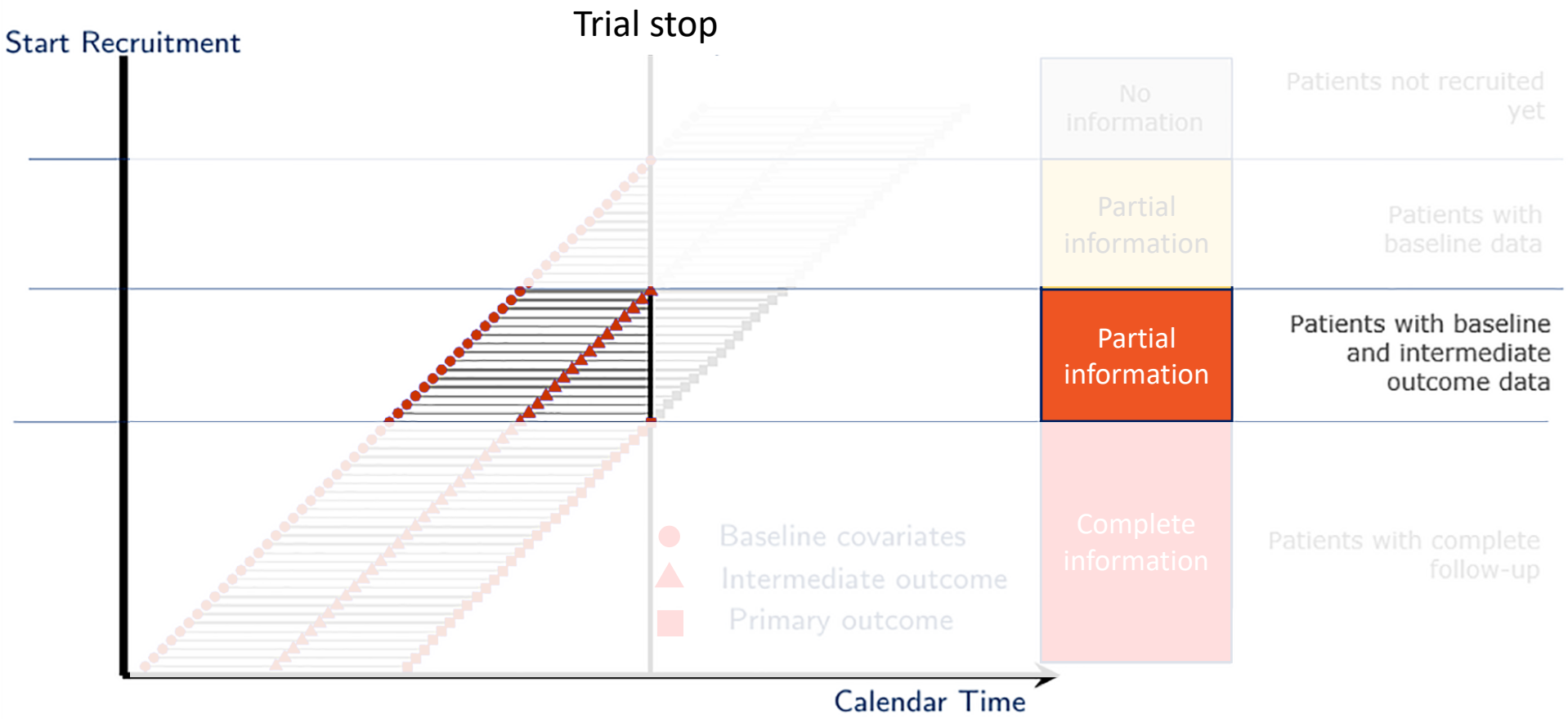
Auxiliary sources



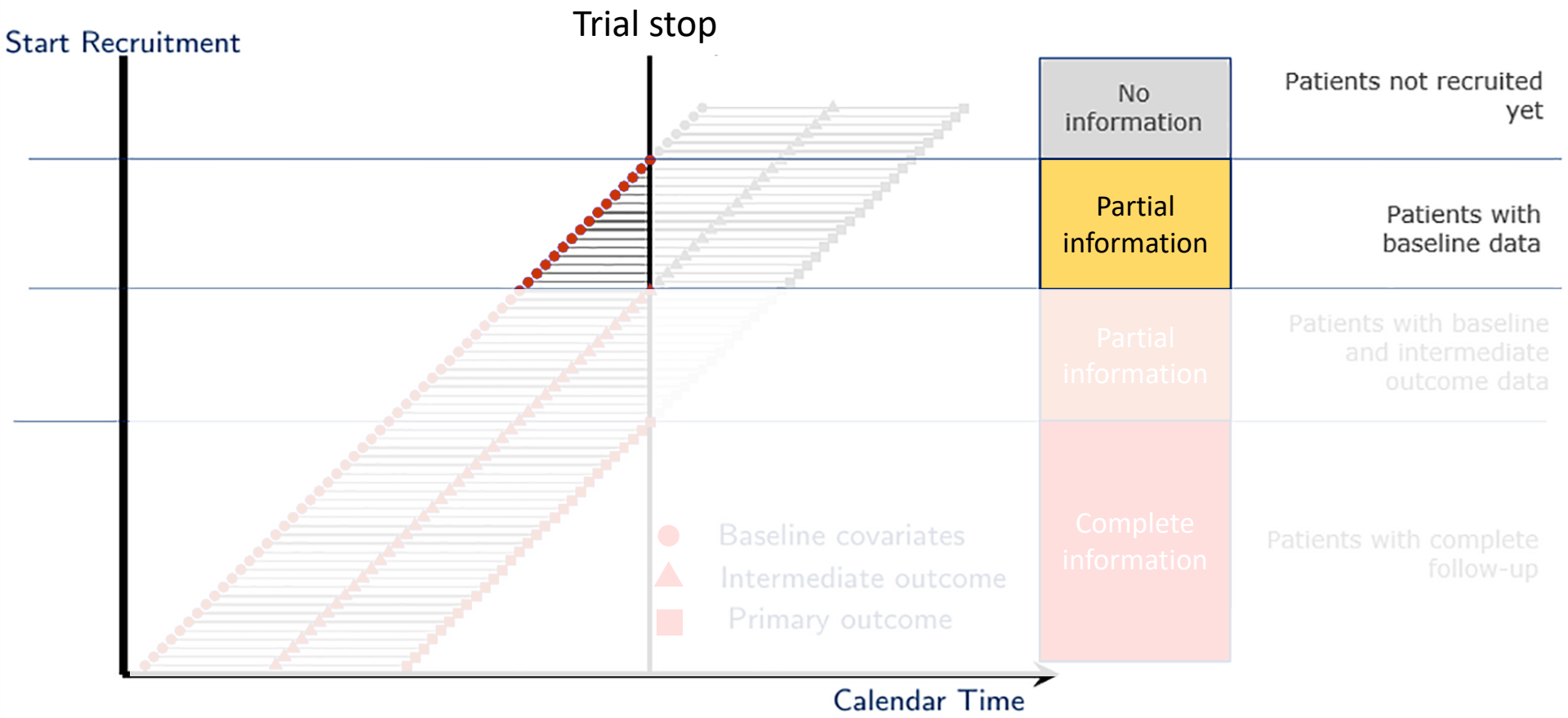
Auxiliary sources



Auxiliary sources



Auxiliary sources



Auxiliary sources

Internal

auxiliary information is available from the patients in the trial itself:

early or baseline data in inference on the primary endpoint of interest.

External

- previously collected (historic) data
- previous reports or publications
- expert knowledge

Auxiliary sources

Internal

auxiliary information is available from the patients in the trial itself:

early or baseline data in inference on the primary endpoint of interest.

methods used on adaptive designs with interim analyses

External

- previously collected (historic) data
- previous reports or publications
- expert knowledge

meta-analysis methods
Bayesian inference (power priors, etc.)

Methods: Bayesian power prior

Let

- D: $x_1, \ldots, x_n$ trial data
- $\boldsymbol{\theta}$ parameter of interest
- D_0: $x_1^0, \ldots, x_m^0$ previous trial data

Bayesian analysis: $\pi_{post}(\boldsymbol{\theta}) \propto \mathcal{L}(\boldsymbol{\theta} \mid x_0, \ldots, x_n)\, \pi_{prior}(\boldsymbol{\theta})$

Power prior $\propto \mathcal{L}(\boldsymbol{\theta}|D_0)^{\alpha}\pi_0(\boldsymbol{\theta})$

$\in [\mathbf{0}, \mathbf{1}]$

How to choose α

Ollier et al. (2020)

$$\alpha = \alpha_0(1-\gamma)$$

α_0: depends on the maximum quantity of information that it is allowed

γ : a similarity criterion (commensurability parameter)

Ollier et al. SMMR, 2020, 29.8: 2282-2294.

How to choose α

Ollier et al. (2020)

$$\alpha = \alpha_0(1-\gamma)$$

ESS
unit-information standard deviation

α_0: depends on the maximum quantity of information that it is allowed

γ : a similarity criterion (commensurability parameter)

How to choose α

$$\alpha = \alpha_0(1-\gamma)$$

Effective Sample Size
unit-information standard deviation

α_0: depends on the maximum quantity of information that it is allowed

γ : a similarity criterion (commensurability parameter)

Commensurability allow to to quantify the degree of similarity between external information and available data.

Ollier et al. (2020) proposed a parameter, using the Hellinger distance between the two normalized likelihoods:

$$\Delta^2(D_0, D_n) = \frac{1}{2}\int\left(\sqrt{\frac{\mathcal{L}(\boldsymbol{\theta}|D)^{\min\left(1,\frac{n_0}{n}\right)}}{\int \mathcal{L}(\boldsymbol{\theta}|D)^{\min\left(1,\frac{n_0}{n}\right)}d\boldsymbol{\theta}}} - \sqrt{\frac{\mathcal{L}(\boldsymbol{\theta}|D_0)^{\min\left(1,\frac{n}{n_0}\right)}}{\int \mathcal{L}(\boldsymbol{\theta}|D_0)^{\min\left(1,\frac{n}{n_0}\right)}d\boldsymbol{\theta}}}\right)^2 d\boldsymbol{\theta}$$

The commensurability parameter can be then defined as Δ^c, with c ∈ R+.

The advantage of this definition is that Δ is bounded between 0 and 1, providing an easy interpretation of the degree of similarity (1- Δ).

add a weakly informative prior to both likelihoods to stabilize the computation

Modifications

Power prior is not tailored to borrow only a subset of $\boldsymbol{\theta}$. Imagine we are interested at borrowing information only on θ_3.

A potential solution:

1. computing posterior of external trial $\mathcal{L}(\boldsymbol{\theta}|D_0)\,\pi_0(\boldsymbol{\theta})$
2. computing Δ on marginal posteriors of θ_3 using the previous Hellinger distance formula (between external trial and the actual trial D)
3. approximating the new marginal prior of θ_3 with a normal distribution
   - mean = posterior mean of step 1
   - sd = $\sqrt{\frac{\left(I_u * n_{missing}\right)^{-1}}{(1-\Delta)^2}}$ with $I_u = \frac{1}{(\text{posterior variance of step 1}) * n_0}$

Modifications

Power prior is not tailored to borrow only a subset of $\boldsymbol{\theta}$. Imagine we are interested at borrowing information only on θ_3.

A potential solution:

1. computing posterior of external trial $\mathcal{L}(\boldsymbol{\theta}|D_0)\,\pi_0(\boldsymbol{\theta})$
2. computing Δ on marginal posteriors of θ_3 using the previous Hellinger distance formula (between external trial and the actual trial D)
3. approximating the new marginal prior of θ_3 with a normal distribution
   - mean = posterior mean of step 1
   - sd = $\sqrt{\dfrac{\left(I_u * n_{missing}\right)^{-1}}{(1-\Delta)^2}}$

As "γ": a function of the commensurability parameter

with $I_u = \dfrac{1}{(\text{posterior variance of step 1}) * n_0}$

Information unit: the information brought in average by 1 individual

Example: PLAN study

PLAN (Primary care pediatrics Learning Activity Nutrition) trial, a diet and exercise intervention for overweight children and one overweight parent compared to usual care.

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

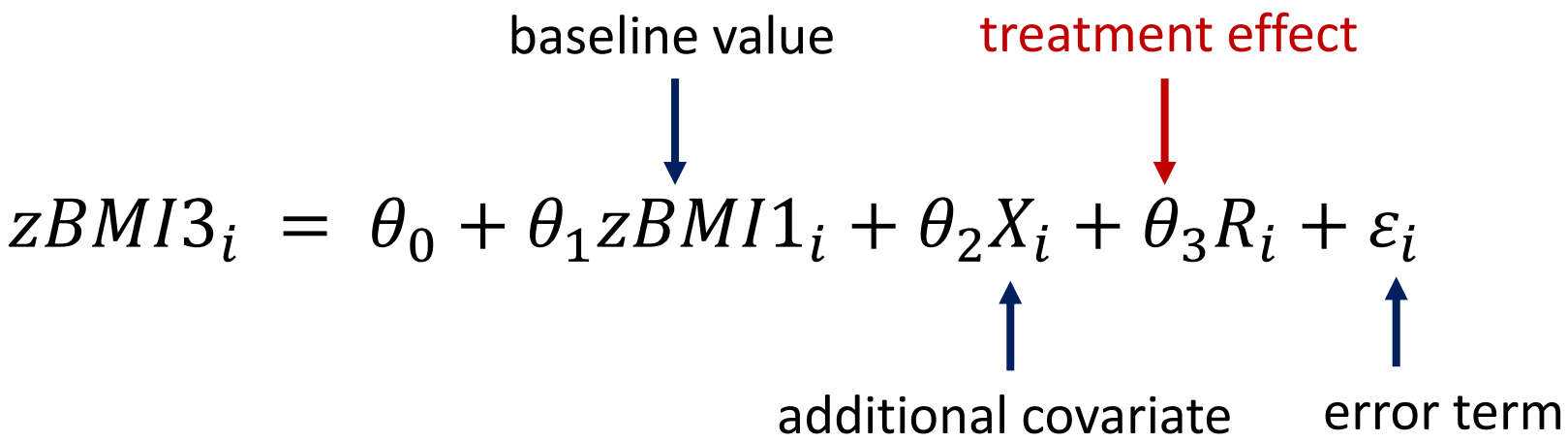
Implementing family-based behavioral treatment in the pediatric primary care setting: Design of the PLAN study

Leonard H. Epstein [a,*], Kenneth B. Schechtman [b], Colleen Kilanowski [a], Melissa Ramel [c], Nasreen A. Moursi [c], Teresa Quattrin [a], Steven R. Cook [d], Ihouma U. Eneli [e], Charlotte Pratt [f], Nancy Geller [f], Rebecca Campo [f], Daphne Lew [b], Denise E. Wilfley [c]

- Pairs of overweight child and parent were randomized to counseling (or usual care.
- Treatment was 26 or more counseling sessions over 24 months.
- The plan was to enroll 528 pairs with age and sex adjusted BMI percentile greater than 85%.
- The recruitment was completed with 452 pairs (n = 452).

Example: PLAN study

ANCOVA planned for analysis

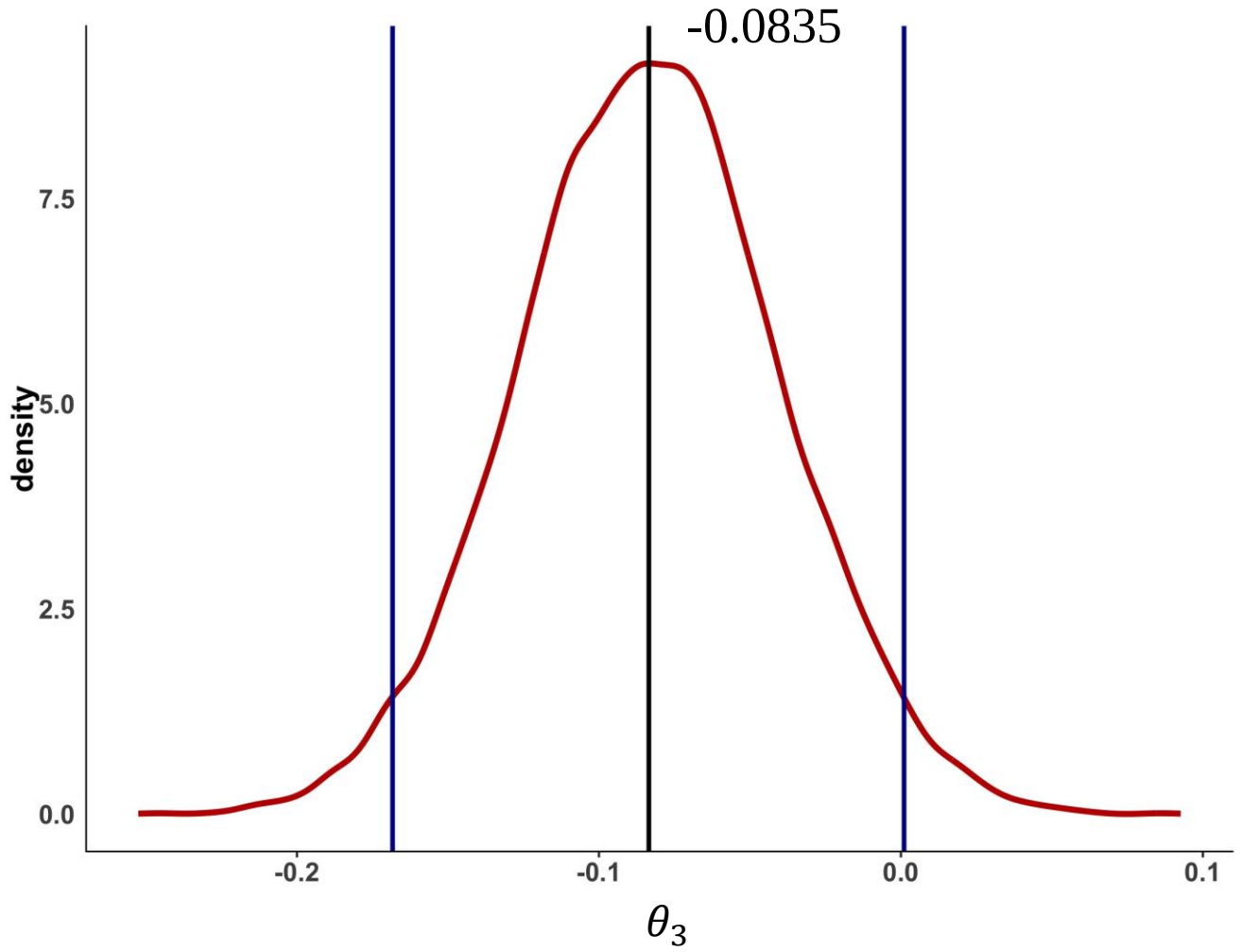


Sample size at trial stopping: 452

250 missing primary endpoint due to COVID19

Example: PLAN study – simulated dataset

Posterior without coping with missing information



$P(\theta_3 > 0) = 0.260$

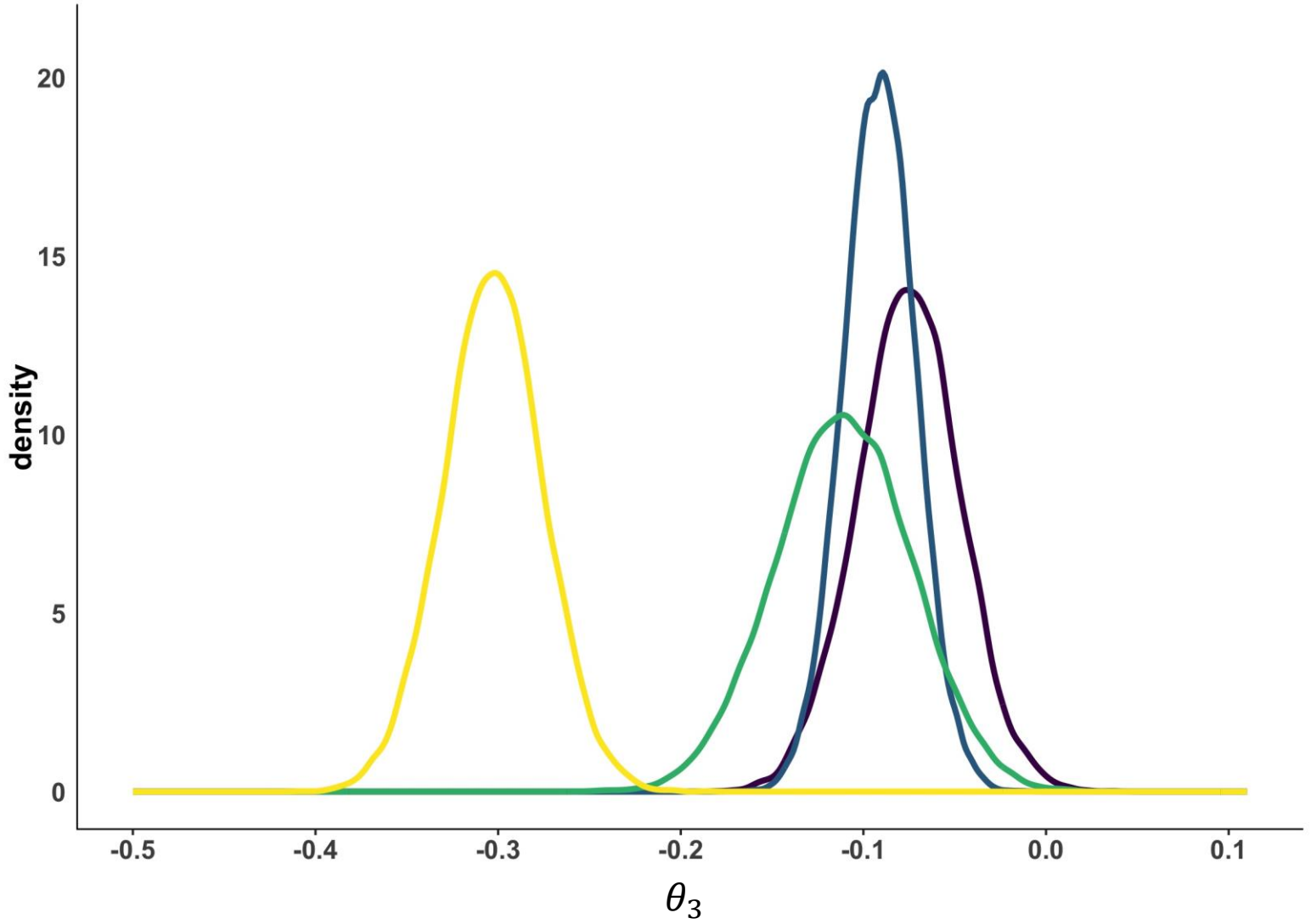
Modifications

Power prior is not tailored to borrow only a subset of $\boldsymbol{\theta}$. Imagine we are interested at borrowing information only on θ_3.

A potential solution:

1. computing posterior of external trial $\mathcal{L}(\boldsymbol{\theta}|D_0)\,\pi_0(\boldsymbol{\theta})$
2. computing Δ on marginal posteriors of θ_3 using the previous Hellinger distance formula (between external trial and the actual trial D)
3. approximating the new marginal prior of θ_3 with a normal distribution
   - mean = posterior mean of step 1
   - sd = $\sqrt{\frac{\left(I_u * n_{missing}\right)^{-1}}{(1-\Delta)^2}}$ with $I_u = \frac{1}{(\text{posterior variance of step 1}) * n_0}$

Example: External data to cope with missing information



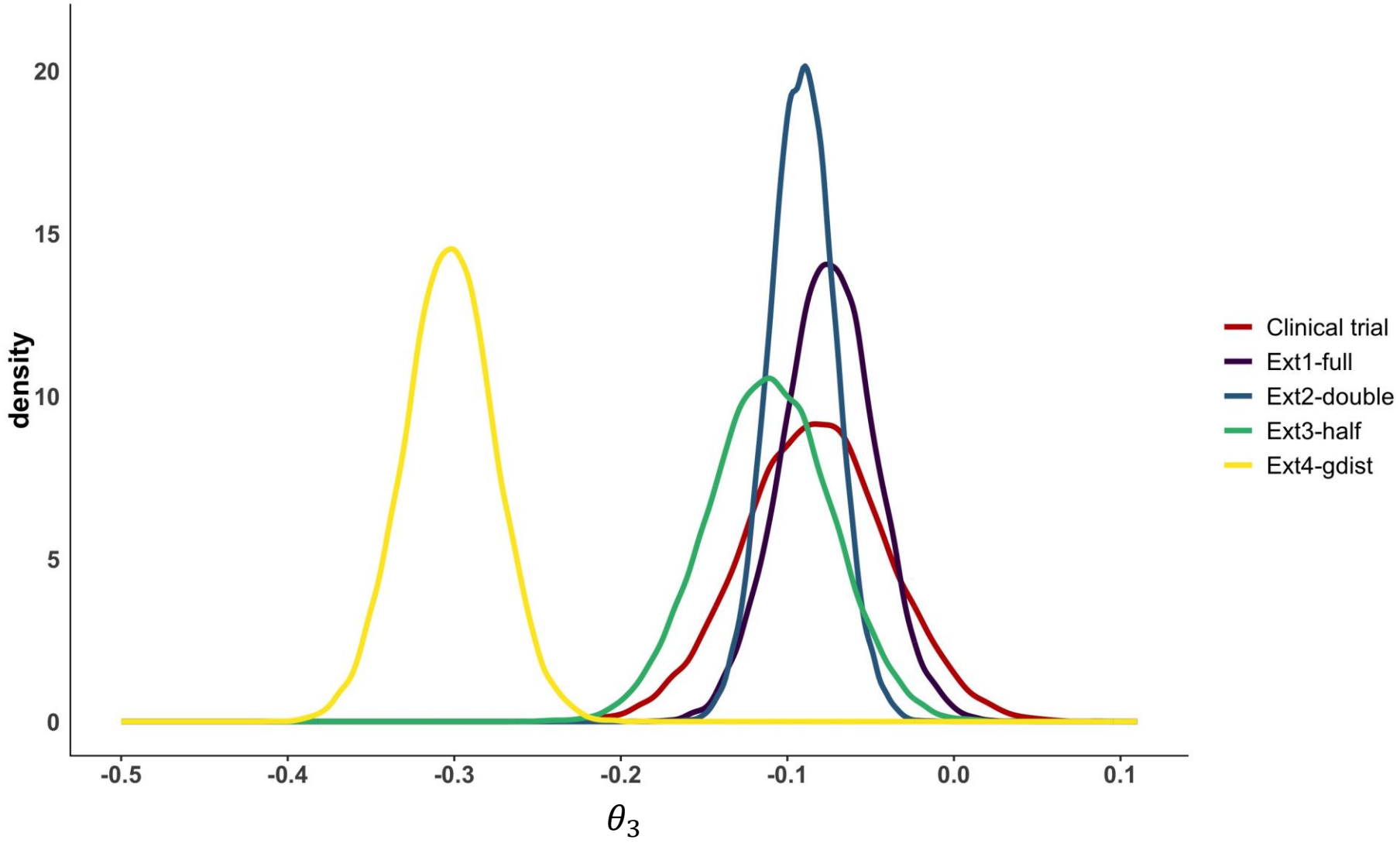
Data generated with the same statistical model and sample size (N=452)

Data generated with the same statistical model and larger sample size (N=904)

Data generated with the same statistical model and half sample size (N=226)

Data generated with a different treatment effect with the same statistical model and sample size (N=452)

Example: External data to cope with missing information



Modifications

Power prior is not tailored to borrow only a subset of $\boldsymbol{\theta}$. Imagine we are interested at borrowing information only on θ_3.

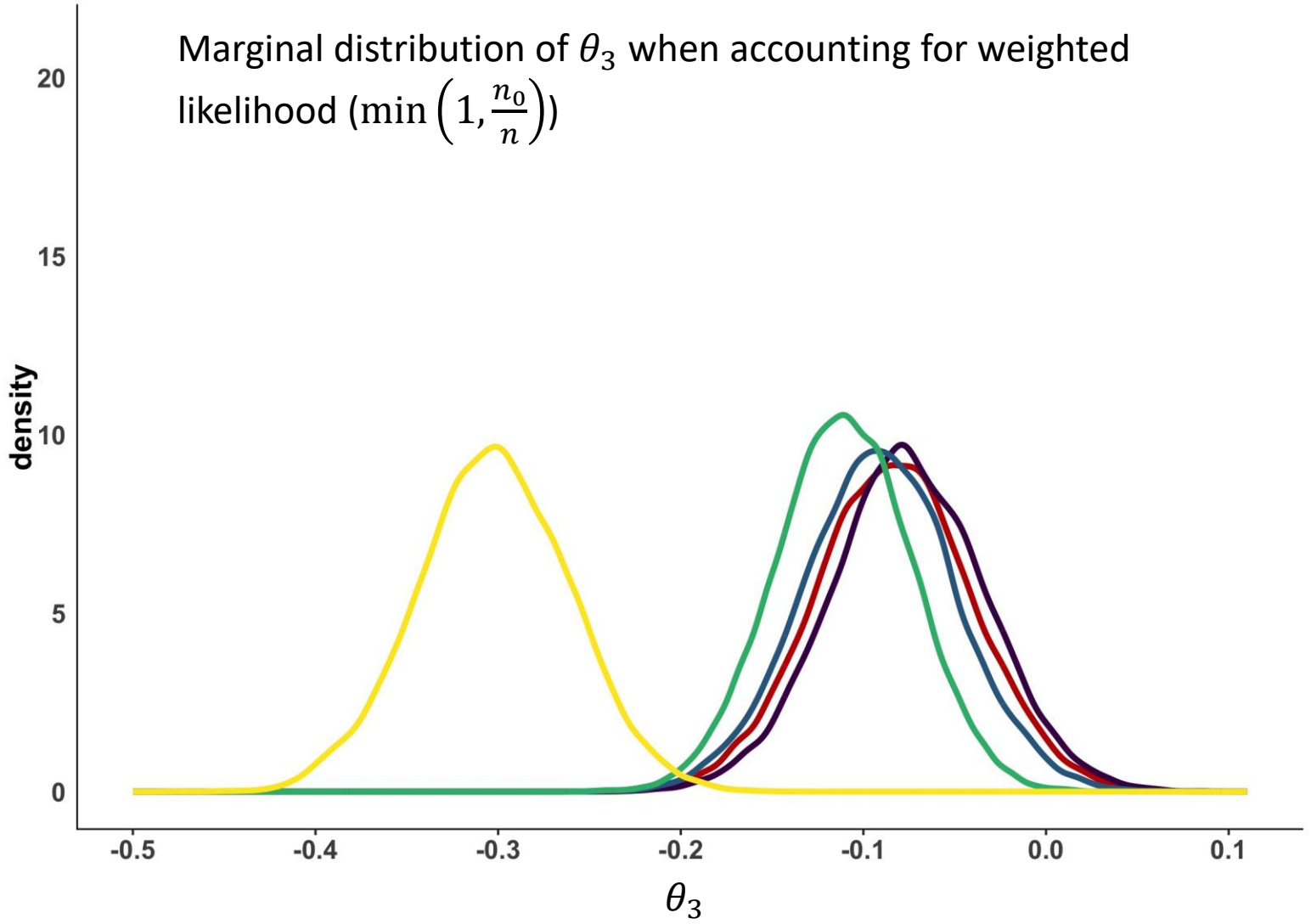
A potential solution:

1. computing posterior of external trial $\mathcal{L}(\boldsymbol{\theta}|D_0)\,\pi_0(\boldsymbol{\theta})$
2. computing Δ on marginal posteriors of θ_3 using the previous Hellinger distance formula (between external trial and the actual trial D)

$$\Delta^2(D_0, D_n) = \frac{1}{2}\int\left(\sqrt{\frac{\mathcal{L}(\boldsymbol{\theta}|D)^{\min\left(1,\frac{n_0}{n}\right)}}{\int\mathcal{L}(\boldsymbol{\theta}|D)^{\min\left(1,\frac{n_0}{n}\right)}d\boldsymbol{\theta}}} - \sqrt{\frac{\mathcal{L}(\boldsymbol{\theta}|D_0)^{\min\left(1,\frac{n}{n_0}\right)}}{\int\mathcal{L}(\boldsymbol{\theta}|D_0)^{\min\left(1,\frac{n}{n_0}\right)}d\boldsymbol{\theta}}}\right)^2 d\boldsymbol{\theta}$$

Example: Δ

Marginal distribution of θ_3 when accounting for weighted likelihood ($\min\left(1, \frac{n_0}{n}\right)$)



Modifications

Power prior is not tailored to borrow only a subset of $\boldsymbol{\theta}$. Imagine we are interested at borrowing information only on θ_3.

A potential solution:

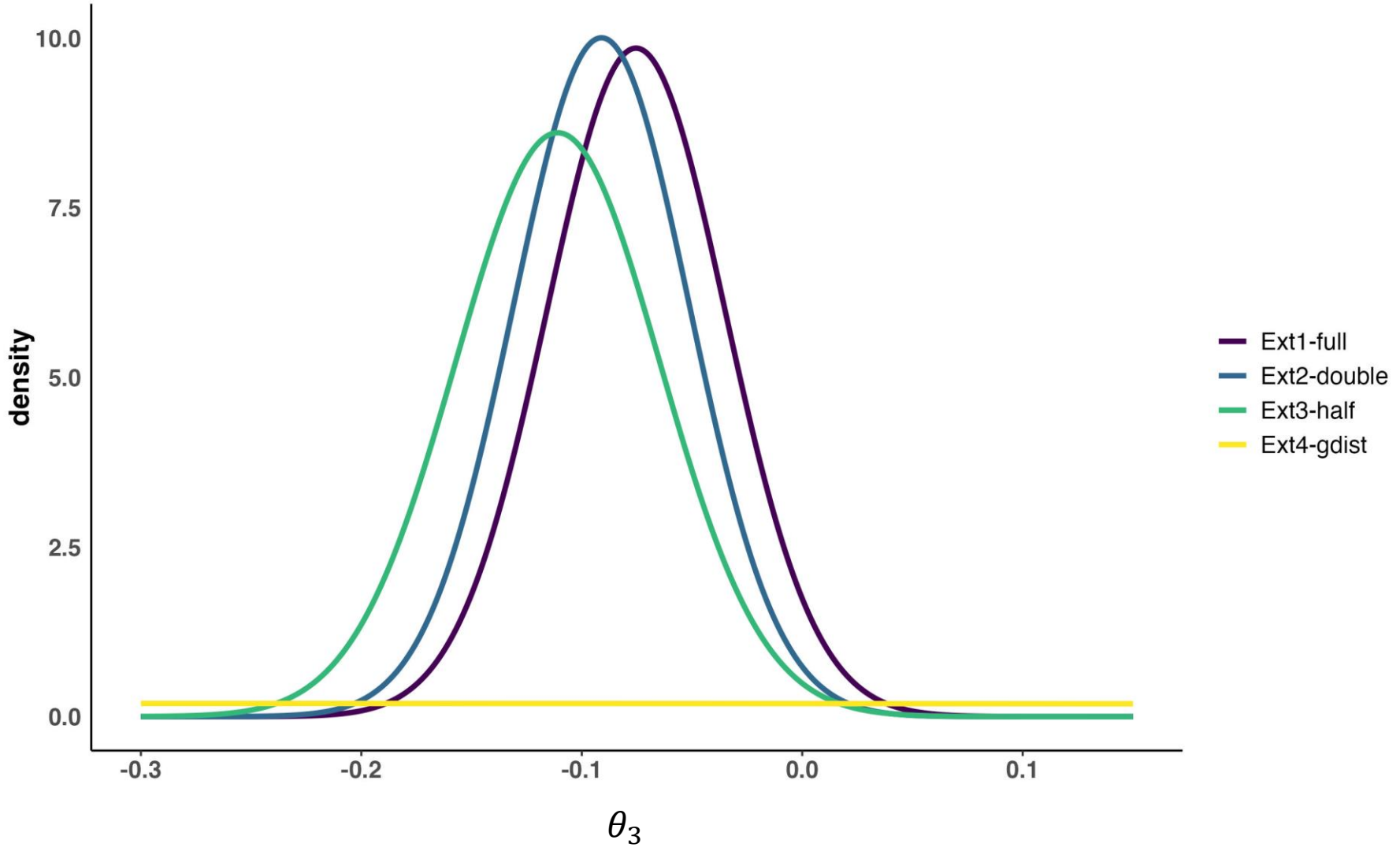
1. computing posterior of external trial $\mathcal{L}(\boldsymbol{\theta}|D_0)\,\pi_0(\boldsymbol{\theta})$
2. computing Δ on marginal posteriors of θ_3 using the previous Hellinger distance formula (between external trial and the actual trial D)
3. approximating the new marginal prior of θ_3 with a normal distribution
   - mean = posterior mean of step 1
   - sd = $\sqrt{\dfrac{\left(I_u * n_{missing}\right)^{-1}}{(1-\Delta)^2}}$

As "γ": a function of the commensurability parameter

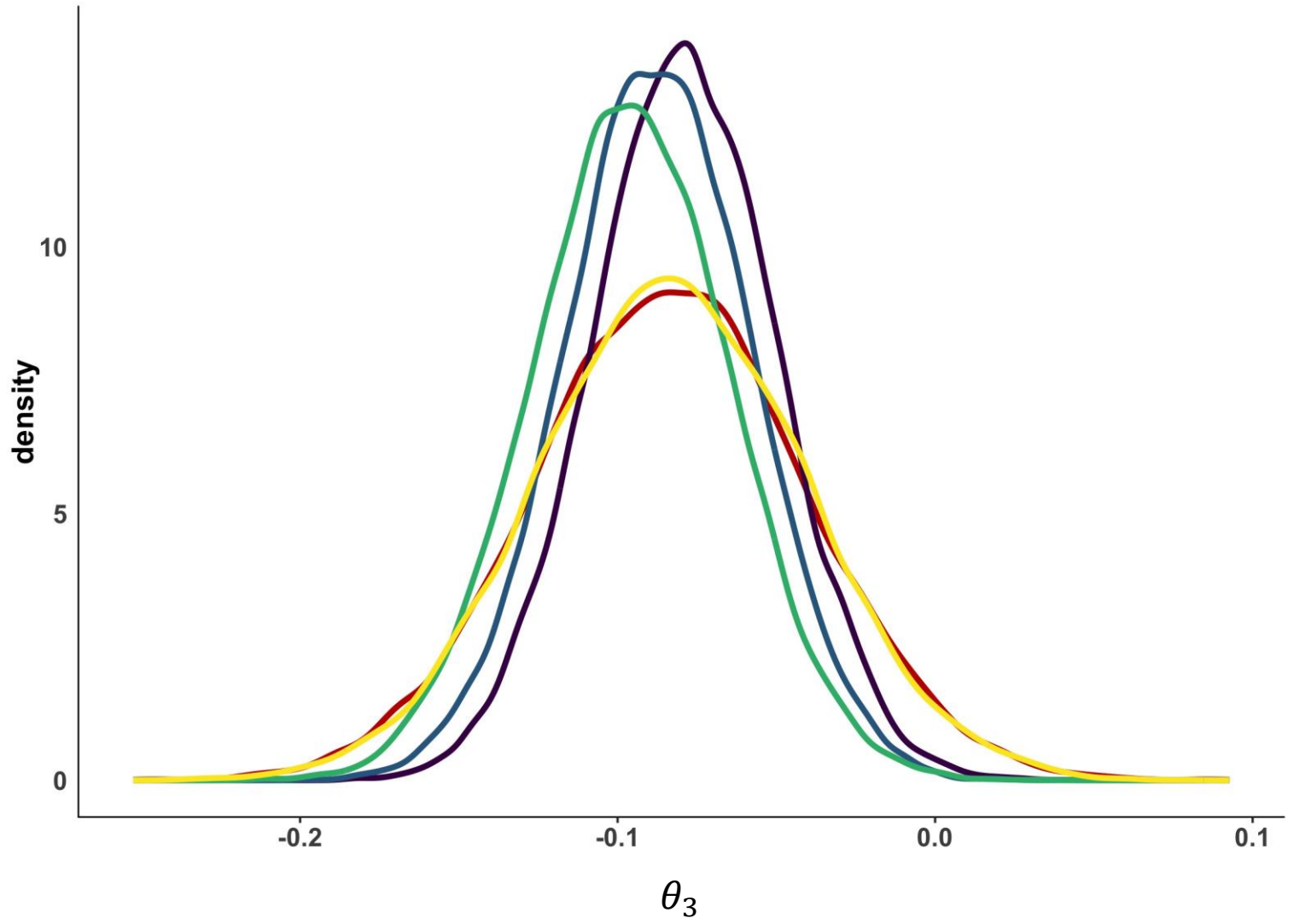
with $I_u = \dfrac{1}{(\text{posterior variance of step 1}) * n_0}$

Information unit: the information brought in average by 1 individual

Example: priors based on external data



Example: final results



$P(\theta_3 > 0) = 0.260$

$P(\theta_3 > 0) = 0.033$

$P(\theta_3 > 0) = 0.015$

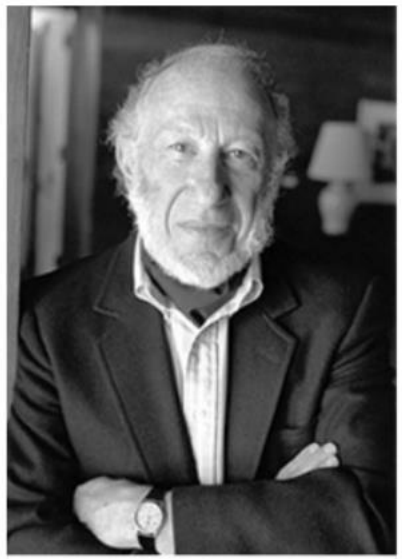
$P(\theta_3 > 0) = 0.011$

$P(\theta_3 > 0) = 0.252$

Conclusion and remark

- Adding external information can lead to “more” conclusive results
- The Bayesian method uses the trial data twice: simulations can be set to verify operational characteristics
- Normal approximation can be avoided and we can work with non-parametric density estimation
- Always checking inclusion/exclusion criteria and trial populations

Acknowledgments



The authors thank the National Institute of Statistical Sciences for facilitating this work on Coping with Information Loss and the Use of Auxiliary Sources of Data, which is part of the Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions.