Analysis of longitudinal studies with treatment by indication

Reagan Mozer
Department of Mathematical Sciences
Bentley University

Joint work with Mark E. Glickman
Support from VA HSR&D IIR 15-115

Motivating application

- Study of veterans diagnosed with Pulmonary Hypertension (PH) from 2005-2012
- One treatment: *phosphodiesterase-5-inhibitors* (PDE5Is)
  - Useful for some rare forms of PH such as type 1 (arterial)
  - Contraindicated for patients with PH types 2 and 3
- Over 2,000 veterans inappropriately prescribed PDE5Is
- **Objective:** Measure causal effects of prescribing contraindicated PDE5Is for treatment of types 2 and 3 PH on time-to-event outcomes
- Several challenges in this setting
  - How should the control group be defined?
  - How can we conceptualize a time-to-event outcome for patients who were not prescribed a PDE5I?

Thought experiment

- Treatment is initiating PDE5I therapy; control is withholding PDE5Is.
- Patients are randomized only when their health status indicates that clinical intervention may be beneficial. This is called the "indication time".
- Outcomes are defined as time from indication to a specified adverse event (e.g., death).

For this idealized experiment

- Randomization $\rightarrow$ balance (in expectation) between treatment and control
- Can use standard inferential techniques to estimate causal effects.

Hypothetical Randomized Experiment

Unit i	**Treatment Indicator** $Z_i \in \{0,1\}$	**Time of Indication** $T_i \in \{1,2,\ldots\}$	**Pre-indication Covariates** X_i	**Potential Outcomes**	
				$Y_{iT_i}(0) - T_i$	$Y_{iT_i}(1) - T_i$
1	1	t_1	x_1	?	$y_1(1) - t_1$
2	1	t_2	x_2	?	$y_2(1) - t_2$
3	0	t_3	x_3	$y_3(0) - t_3$	?
4	0	t_4	x_4	$y_4(0) - t_4$	?

In the real world

- Assignment is applied at time of indication, which is observed for treated units, but missing otherwise.
- All units who are not observed to receive treatment
  - may receive treatment after study period, or
  - may be controls with unobserved indication times.
- Controls may receive alternative therapy (which may not be identifiable from available data) or nothing at all.
- In addition to missing indication times, potential confounders pose a challenge for causal inference.

In the real world

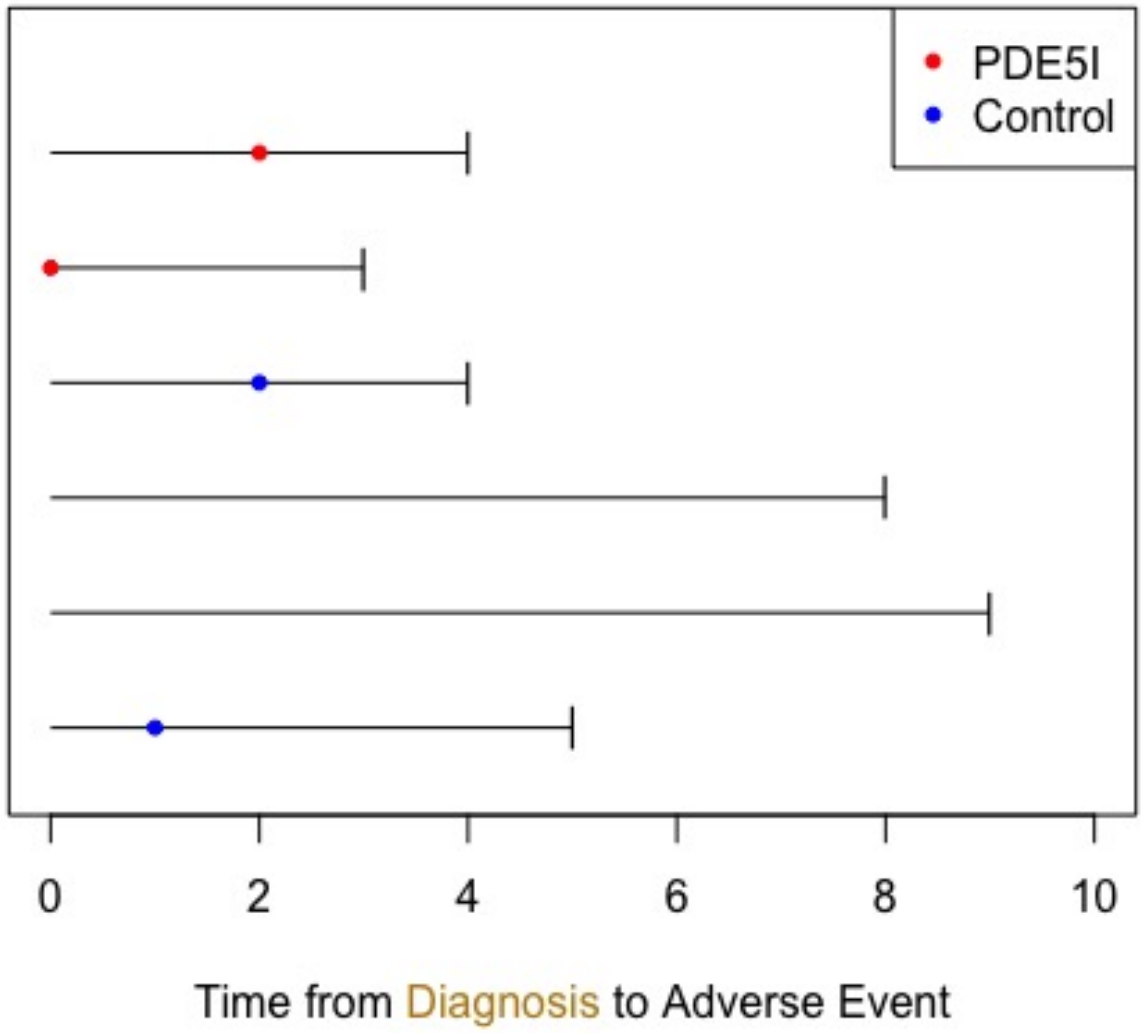
Observed Data

Unit i	**Time of PDE5I** $T_i \in \{1, 2, \ldots\}$	**Covariates** X_{i1}	X_{i2}	$\ldots$	**Time of Adverse Event** Y_i
1	t_1	x_{11}	x_{12}	$\ldots$	y_1
2	t_2	x_{21}	x_{22}	$\ldots$	y_2
3	NA	x_{31}	x_{32}	$\ldots$	y_3
4	NA	x_{41}	x_{42}	$\ldots$	y_4

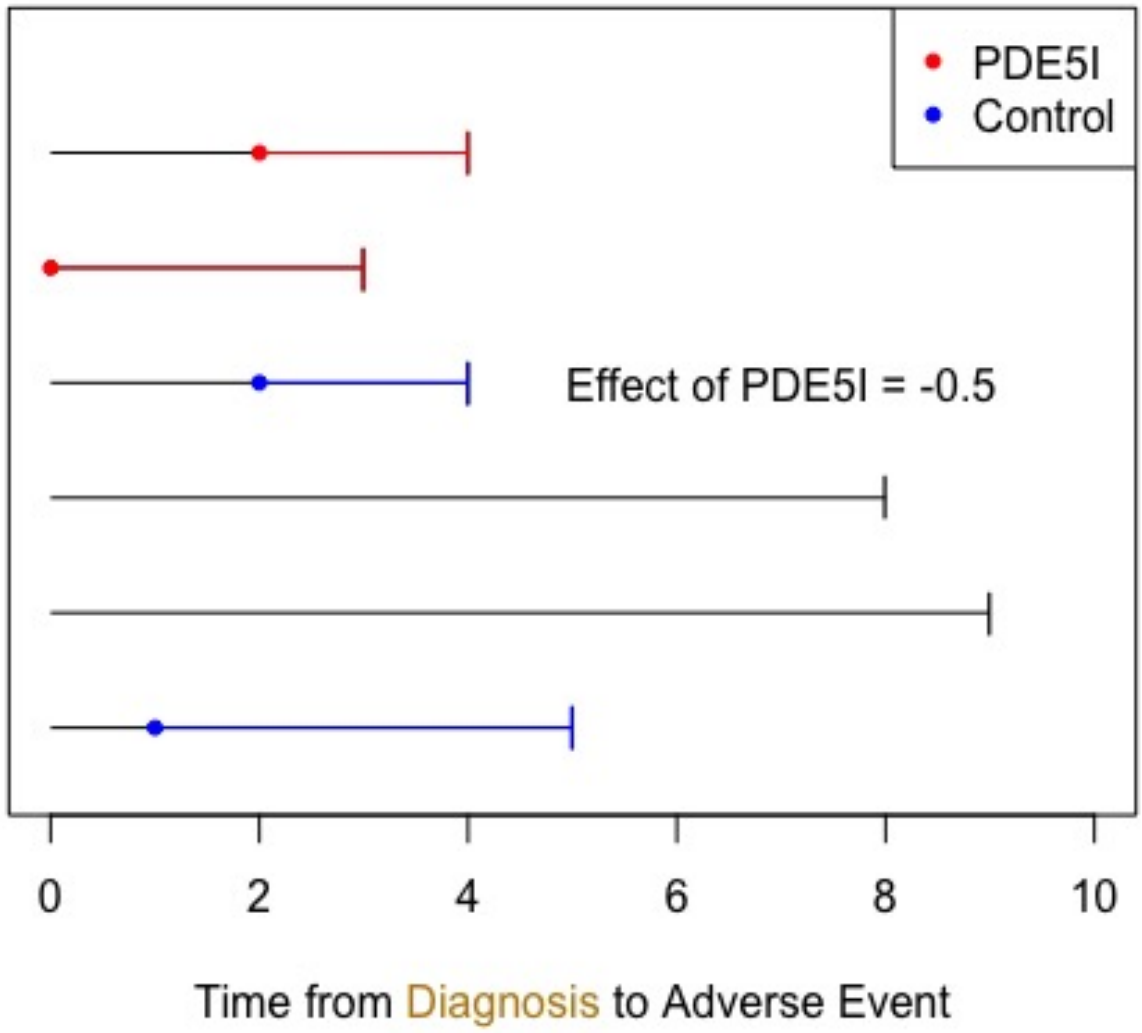
"Messy" Observational Study

Unit i	**Treatment Indicator** $Z_i \in \{0, 1\}$	**Time of Indication** $T_i \in \{1, 2, \ldots\}$	**Pre-treatment Covariates** $X_i = (X_{i1}, ..., X_{iT_i})$	**Potential Outcomes** $Y_i(0) - T_i$	$Y_i(1) - T_i$
1	1	t_1	x_1	?	$y_1 - t_1$
2	1	t_2	x_2	?	$y_2 - t_2$
3	??	??	??	y_3—??	?
4	??	??	??	y_4—??	?

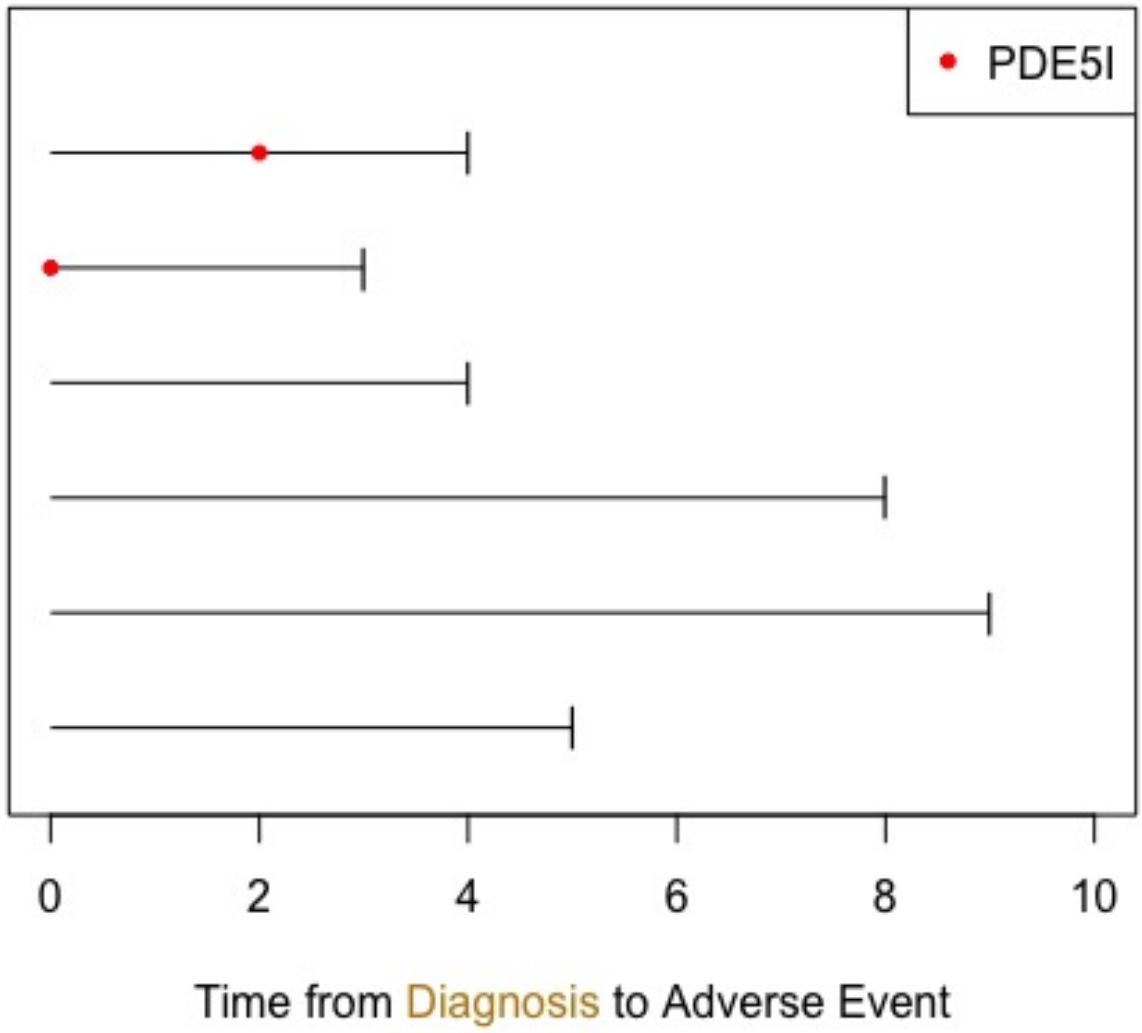
The full picture - idealized experiment



The full picture - idealized experiment



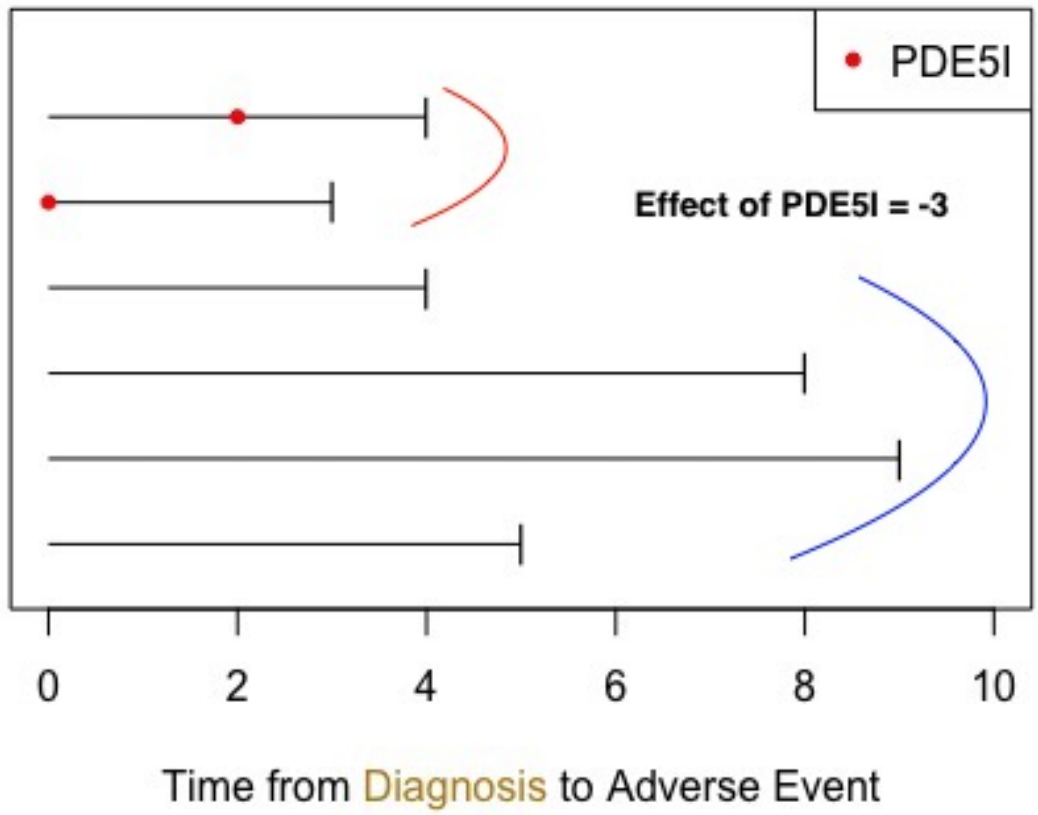
Real-world observed data



What Do I Do Now

Option 1: Naive comparisons

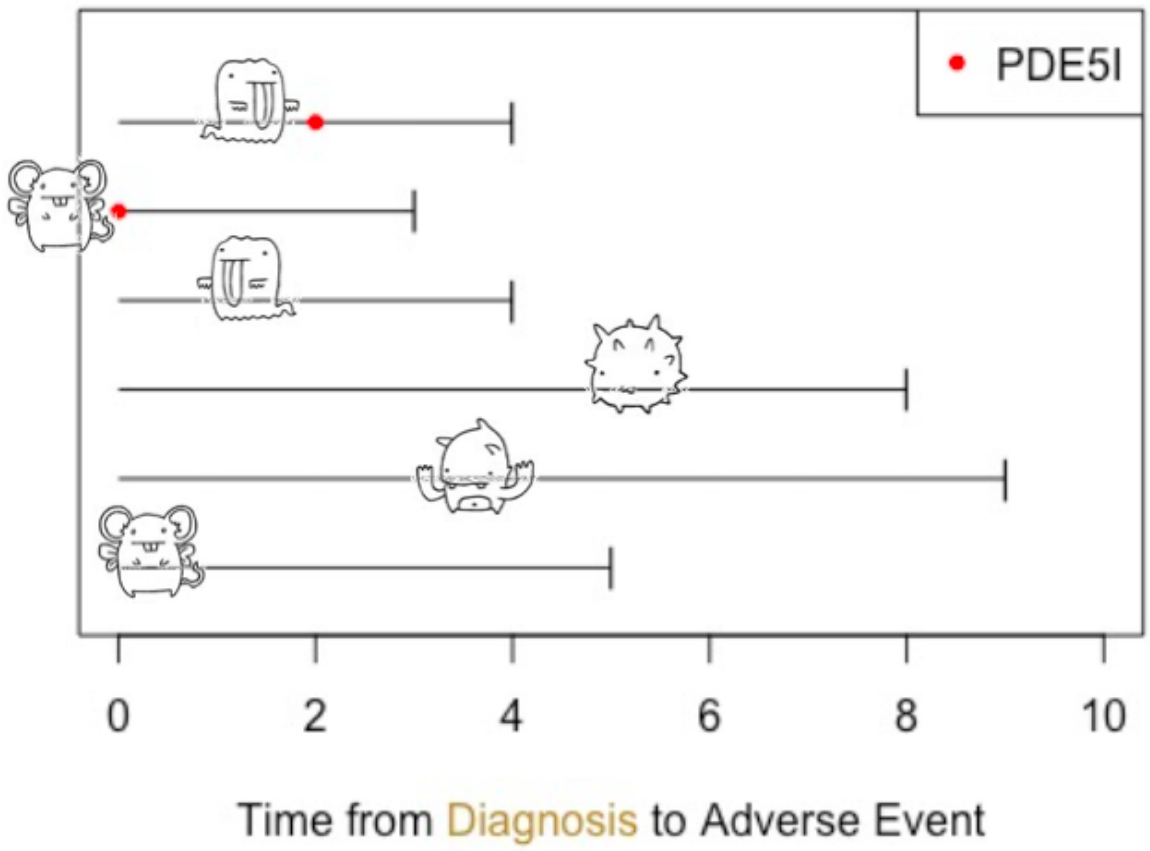
- **Strategy**: Compare times from *diagnosis* to adverse event for patients observed to receive a PDE5I during the study versus those not observed to receive a PDE5I?



- **Not good**: Wrong estimand, wrong control population

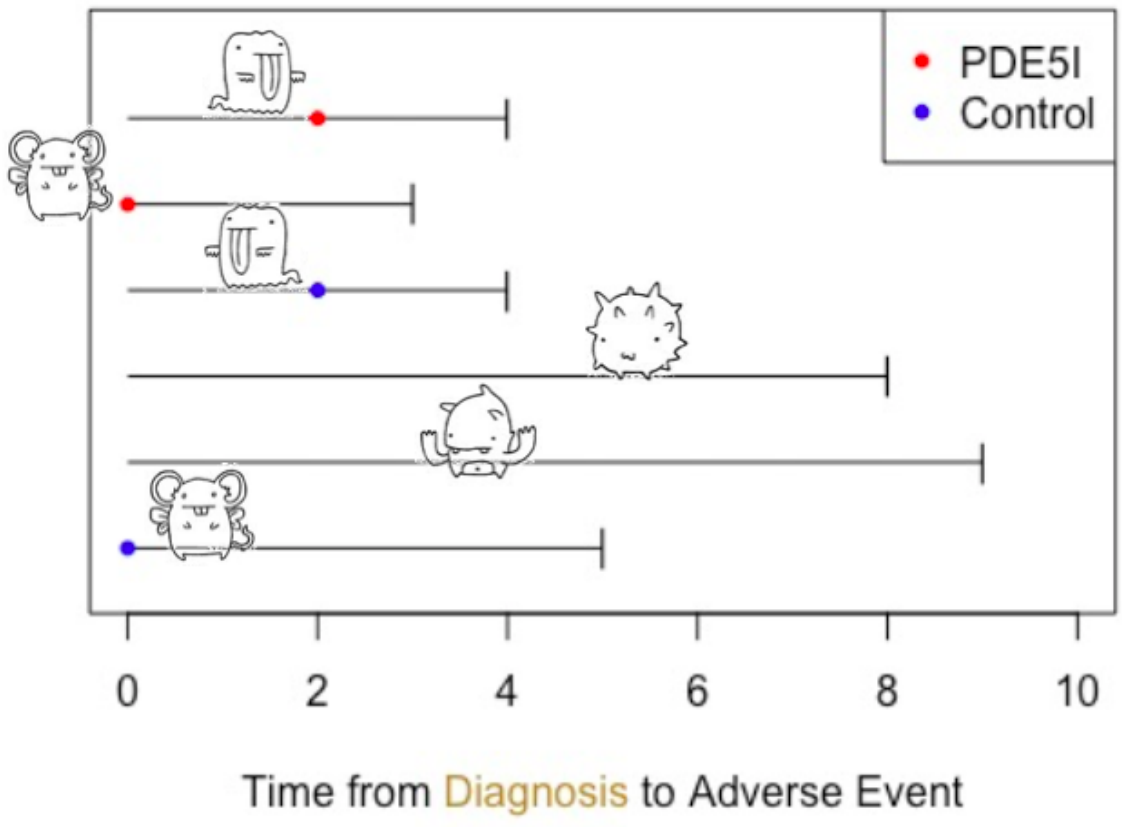
Option 2: Risk Set Matching (Li et al., 2001)

- For each treated unit, find not-yet-treated unit most similar before assignment



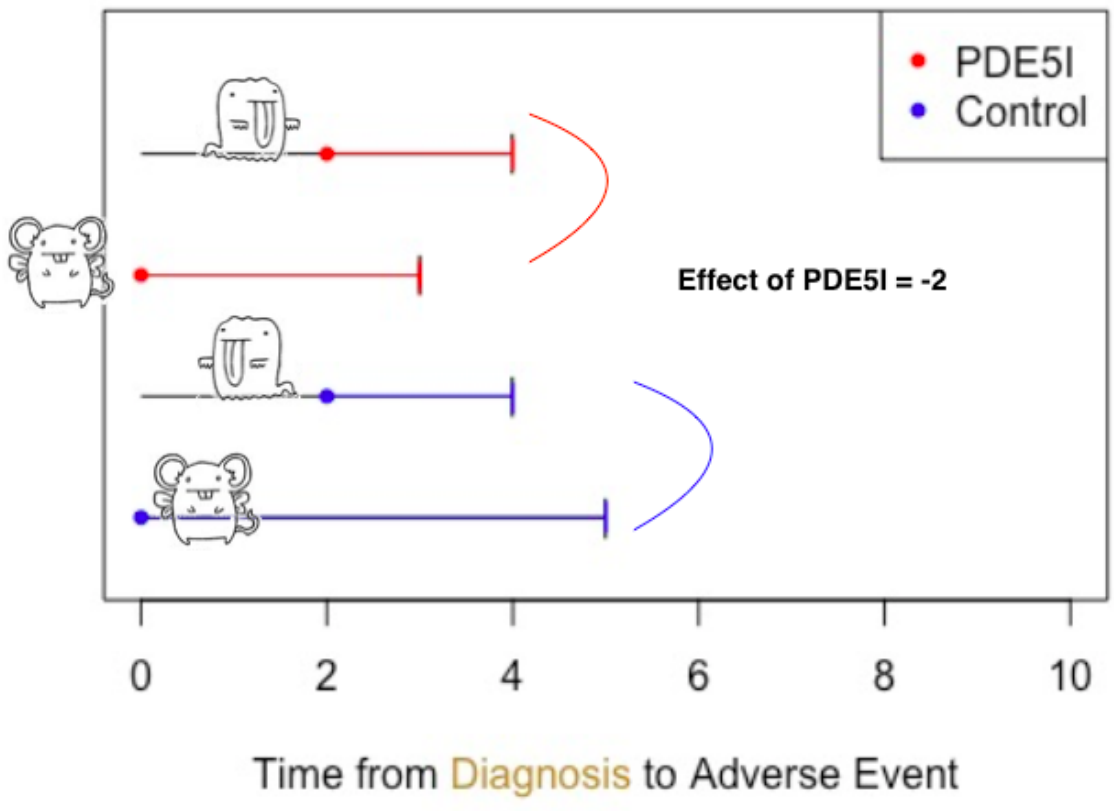
Option 2: Risk Set Matching (Li et al., 2001)

- For each treated unit, find not-yet-treated unit most similar before assignment
- Assume they would have been treated at the same time



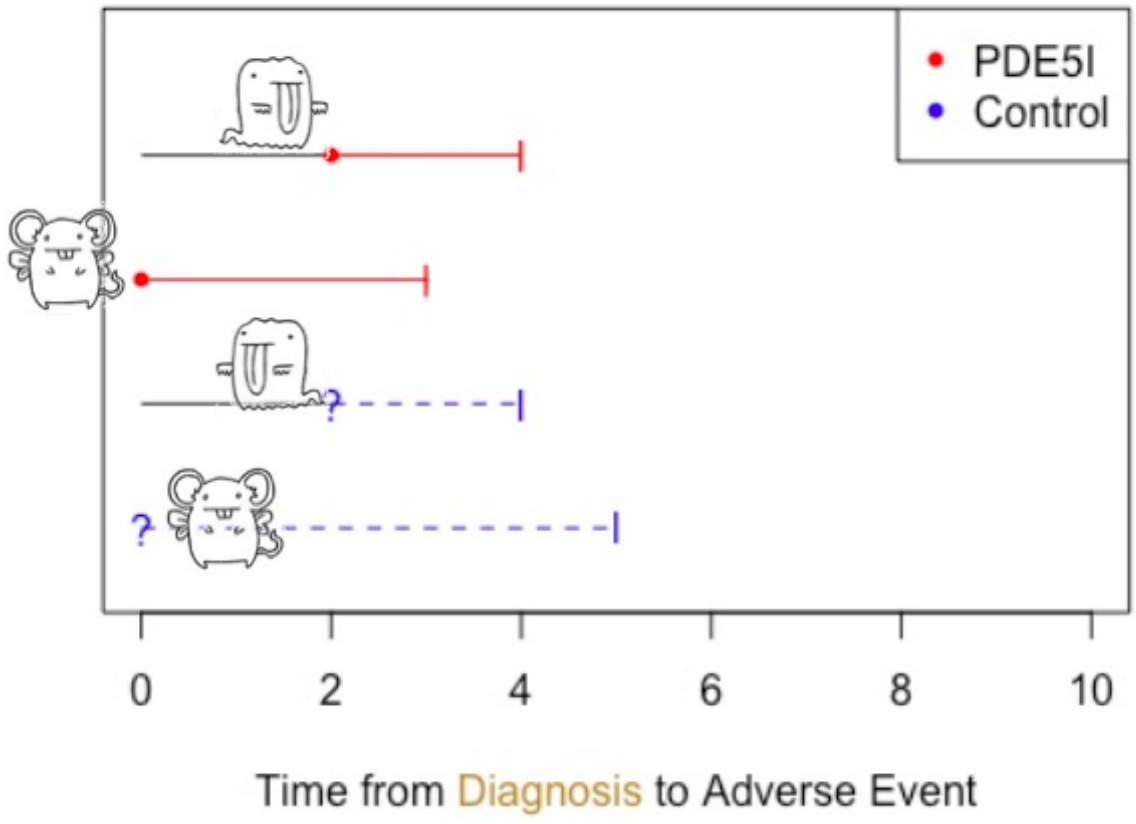
Option 2: Risk Set Matching (Li et al., 2001)

- For each treated unit, find not-yet-treated unit most similar before assignment
- Assume they would have been treated at the same time
- Compare outcomes for these units



Option 2: Risk Set Matching (Li et al., 2001)

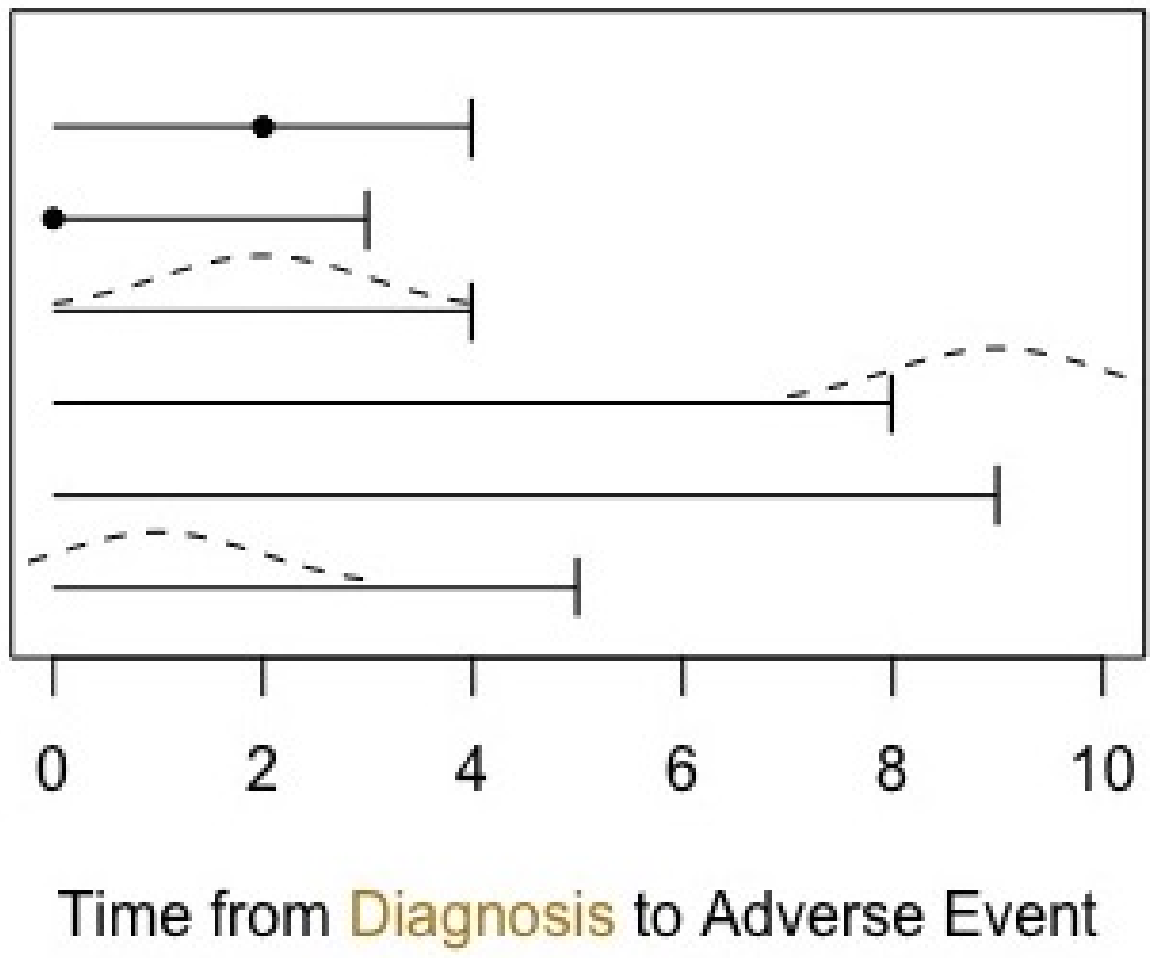
- This is better, but...



- Uncertainty in assignment time → uncertainty in outcomes. We should account for this when estimating treatment effects!

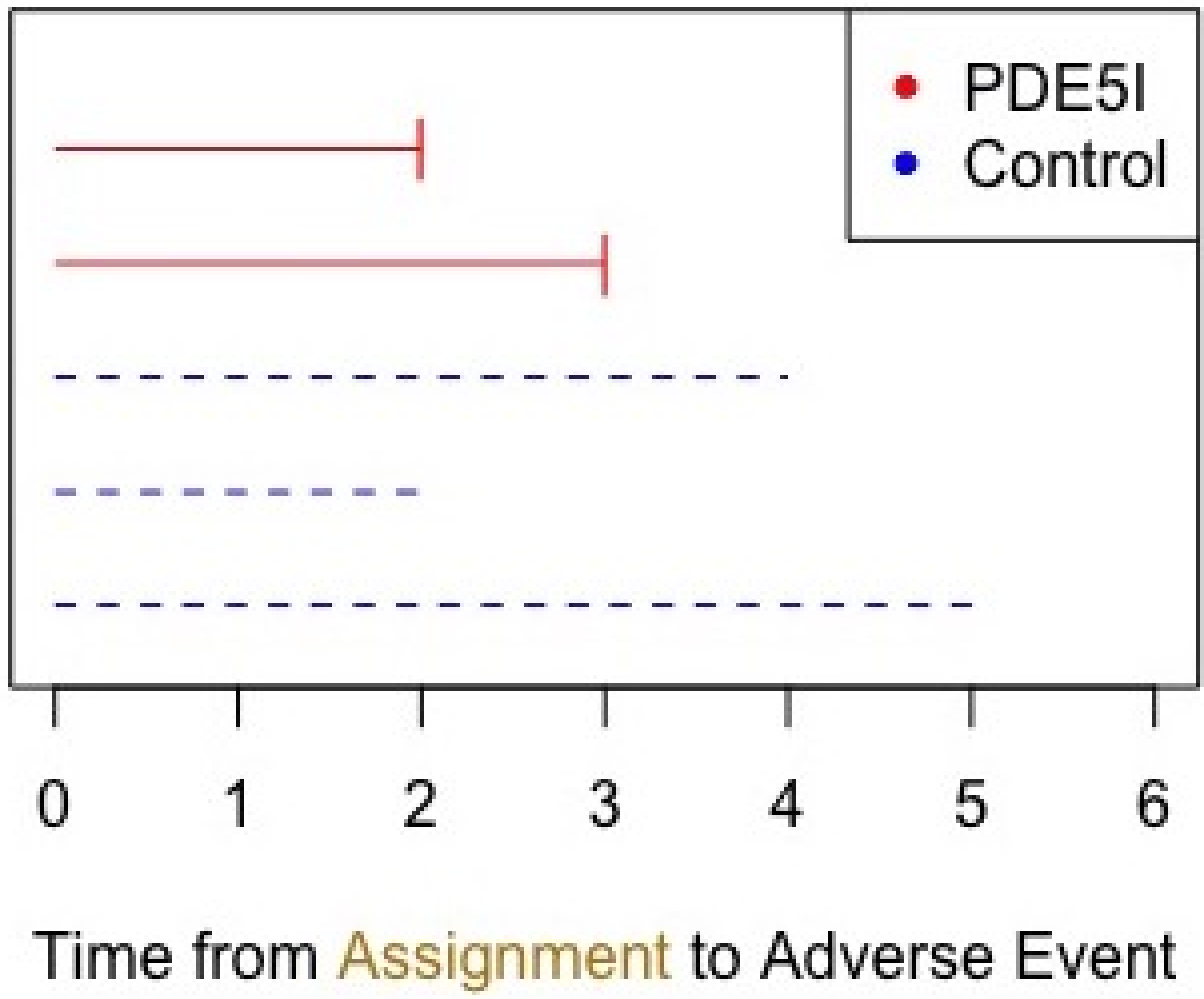
Our approach

Step 1: Obtain the posterior distribution of missing indication times, ignoring outcome time data in the analysis.

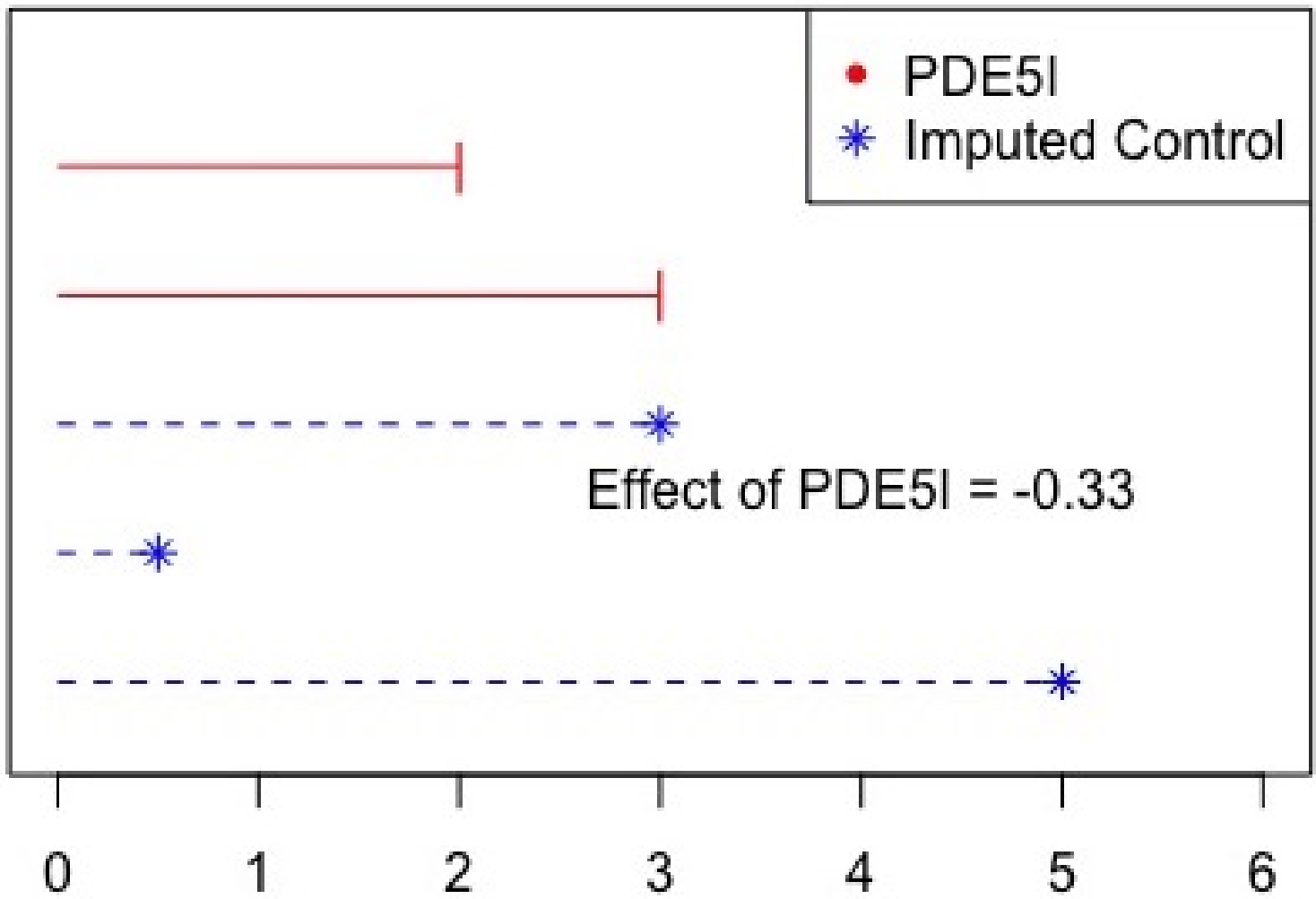


Our approach

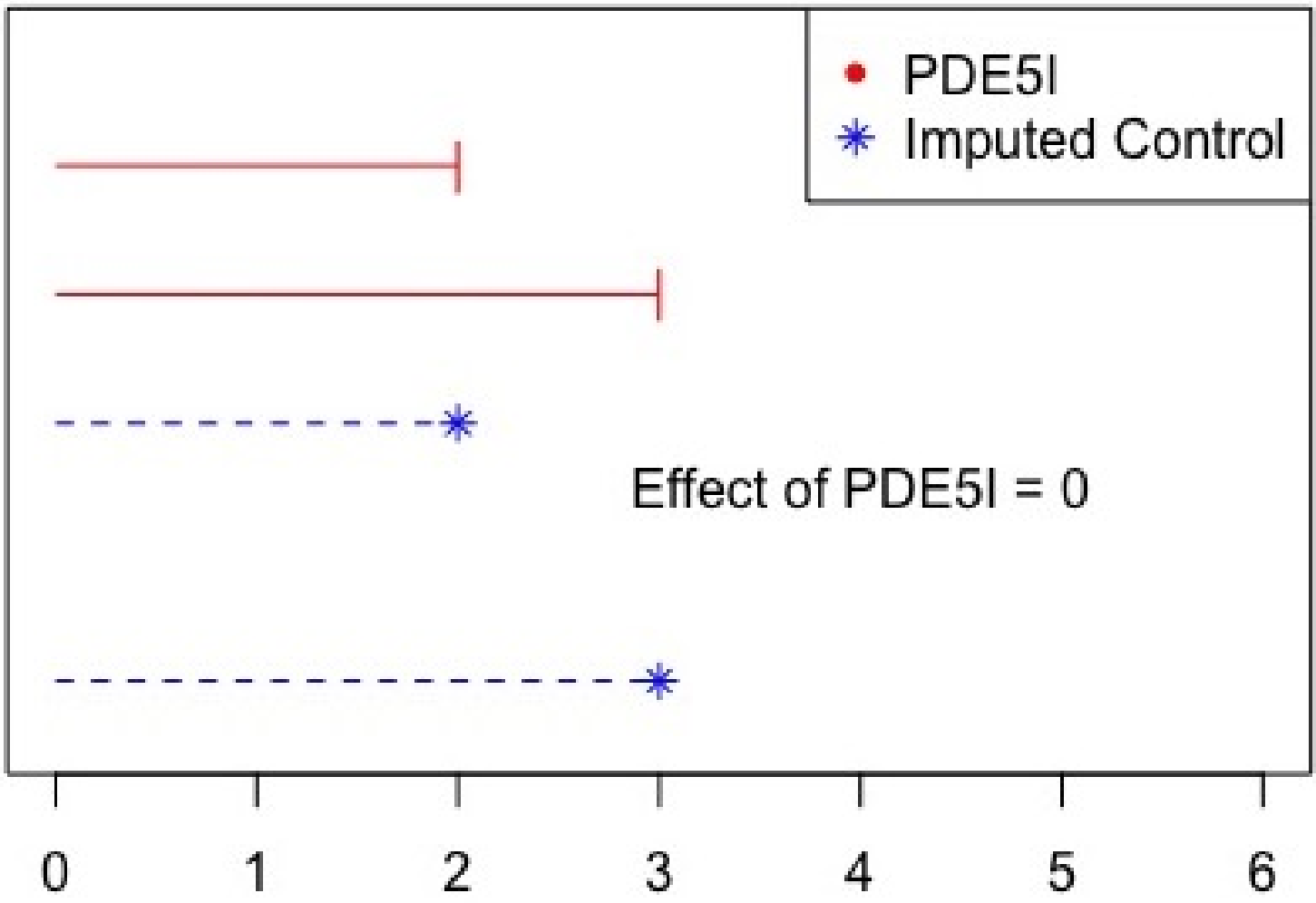
Step 2: Summarize treatment effects based on the posterior distribution of indication times.



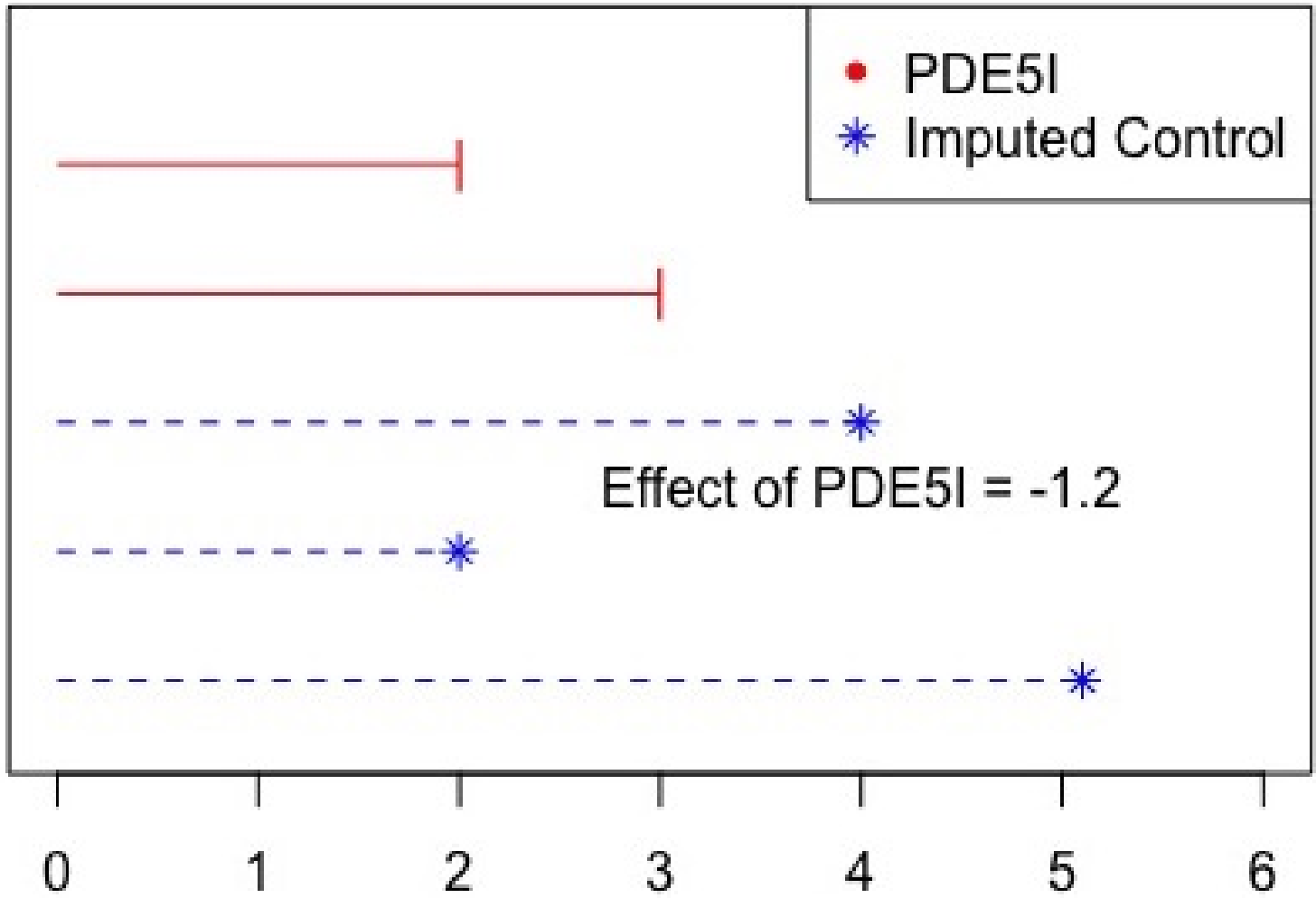
Posterior simulation



Posterior simulation



Posterior simulation



Quick recap of proposed framework

- Study period over discrete time points $0, \ldots, K$
- **Time of assignment**: indication time T_i
- **Treatment**: single medical intervention of interest, which is either *initiated* ($Z_i = 1$) or *withheld* ($Z_i = 0$) upon indication
- **Outcomes**: $Y_{iT_i}(Z_i)$ defined relative to indication time
- **Estimands**: $E[Y_{it}(1) - Y_{it}(0)]$, $0 \leqslant t \leqslant K$
- **Study population**: subset of patients i with $T_i \in [0, K]$
- **Patient sample**: consists of
  - *treated* units ($T_i \in [0, K], Z_i = 1$)
  - *"true" controls* ($T_i \in [0, K], Z_i = 0$)
  - *ineligible* units ($T_i > K$)

Modeling approach

Model two underlying processes conditional on observed covariates:

1. Patients' health determines *time* of indication.
   - Patients enter the study at varying levels of overall health.
   - Patients become eligible for treatment only when some such medical intervention is deemed necessary.
2. Given time of indication, external factors can influence assignment to treatment versus control.
   - Clinicians' knowledge about effectiveness of therapies, adherence to protocol, etc., varies over time.

1. A model for indication time

Suppose patient health follows the discrete-time state space model for $t = 1, \ldots, K$:

$$\Psi_{it} = \theta_{it} + X_{it}\beta + \nu_{it}, \quad \nu_{it} \sim^{iid} \mathcal{N}(0, 1)$$
$$\theta_{it} = \rho\theta_{it-1} + \epsilon_{it}, \quad \epsilon_{it} \sim^{iid} \mathcal{N}(0, 1)$$

Indication times can be expressed as as the first-hitting time:

$$T_i = \inf\{t : \Psi_{it} > 0\}$$

with probabilities corresponding to the probit regression model

$$P(\Psi_{it} > 0 | \theta_{it}, X_{it}) = \Phi(\theta_{it} + X_{it}\beta).$$

2. A model for the assignment mechanism

For $t = 1, \ldots, K$, the probability of receiving treatment versus control upon indication is:

$$P(Z_i = 1 | T_i, D_i) = f(T_i, D_i, \delta)$$

where $0 < f(T_i, D_i, \delta) < 1$ is a function of exogenous factors D_i (e.g., calendar date) and parameters δ.

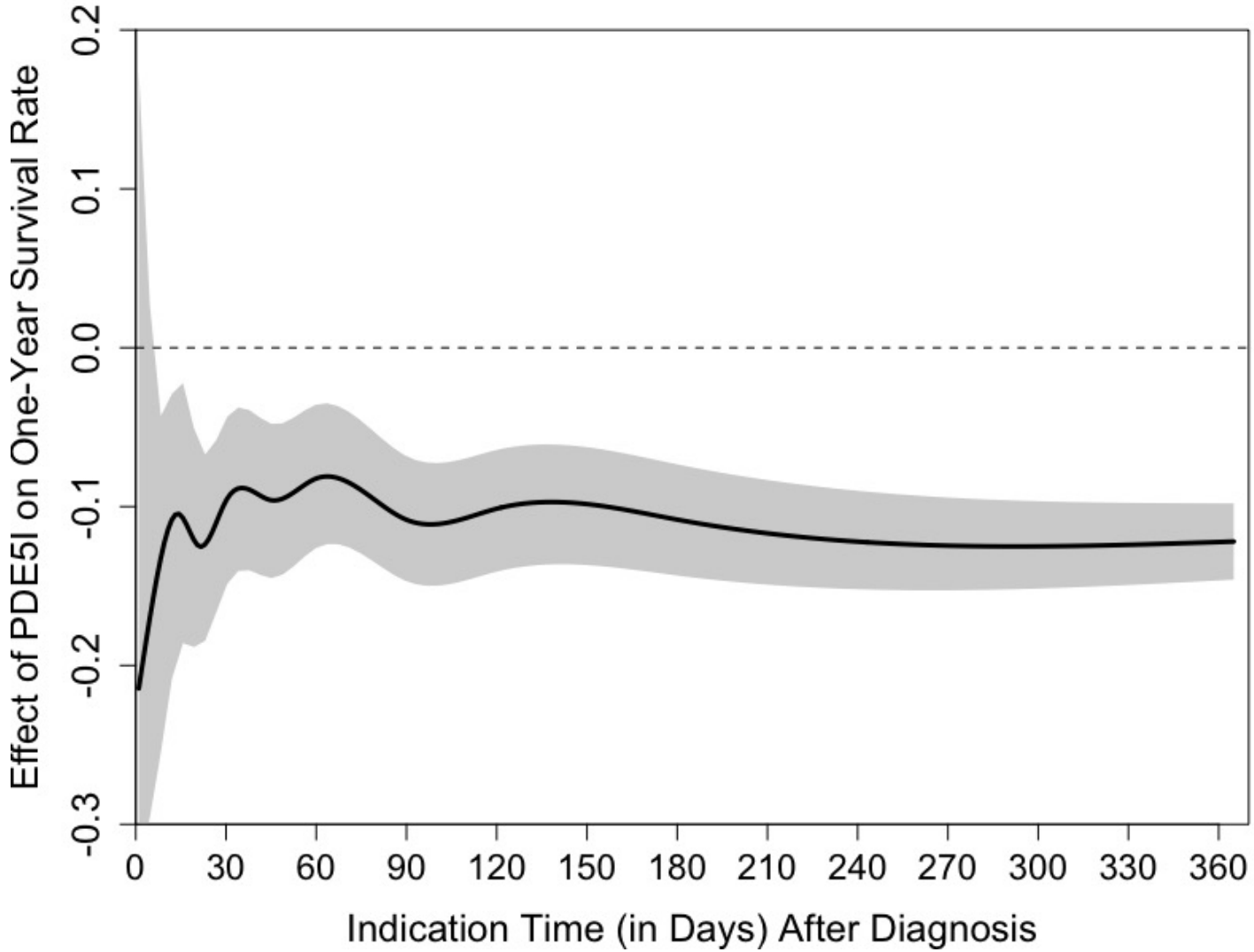
Some inferential procedure details

- Missing indication times (for untreated units) are inferred as part of MCMC posterior simulation.
- Autoregressive process on $(\theta_1, \ldots, \theta_K)$ is latent, but can be marginalized out using the Kalman Filter for computational efficiency.
- Summaries of aggregated average treatment effects can be obtained via appropriate averages over posterior samples.
- Can also report posterior summaries of
  - covariate effects on indication times, and
  - assignment probability to treatment over time.

Application to VA study

- Medical intervention of interest: Receipt of PDE5I prescription within one year of PH diagnosis.
- Outcome of interest: Survival one year after indication time.
- Final sample: 534 treated patients and 531 potential controls matched at PH diagnosis date selected from a pool of 167,000 untreated patients.
- Time-varying covariates included indicators for recent hospitalization, changes in medical history and presence of comoborbidities.

Application to VA study



Findings: the effect of initiating PDE5Is upon indication is a 10-11 percentage point decrease in survival rates.

Final thoughts

- Main innovation: Conceptualizing indication times for random assignment to control versus treatment.
- Still need to use standard methods for observational study analyses (e.g., bias reduction through matching, etc.)
- Framework permits flexible modeling choices for evolution in health status to trigger indication time, probability of control/treatment assignment, and so on.
- Please see full details in our paper.

Thanks for listening!

rmozer@bentley.edu
glickman@fas.harvard.edu

Paper: arxiv.org/abs/1909.06432
Slides: reaganmozer.com/ichps2020