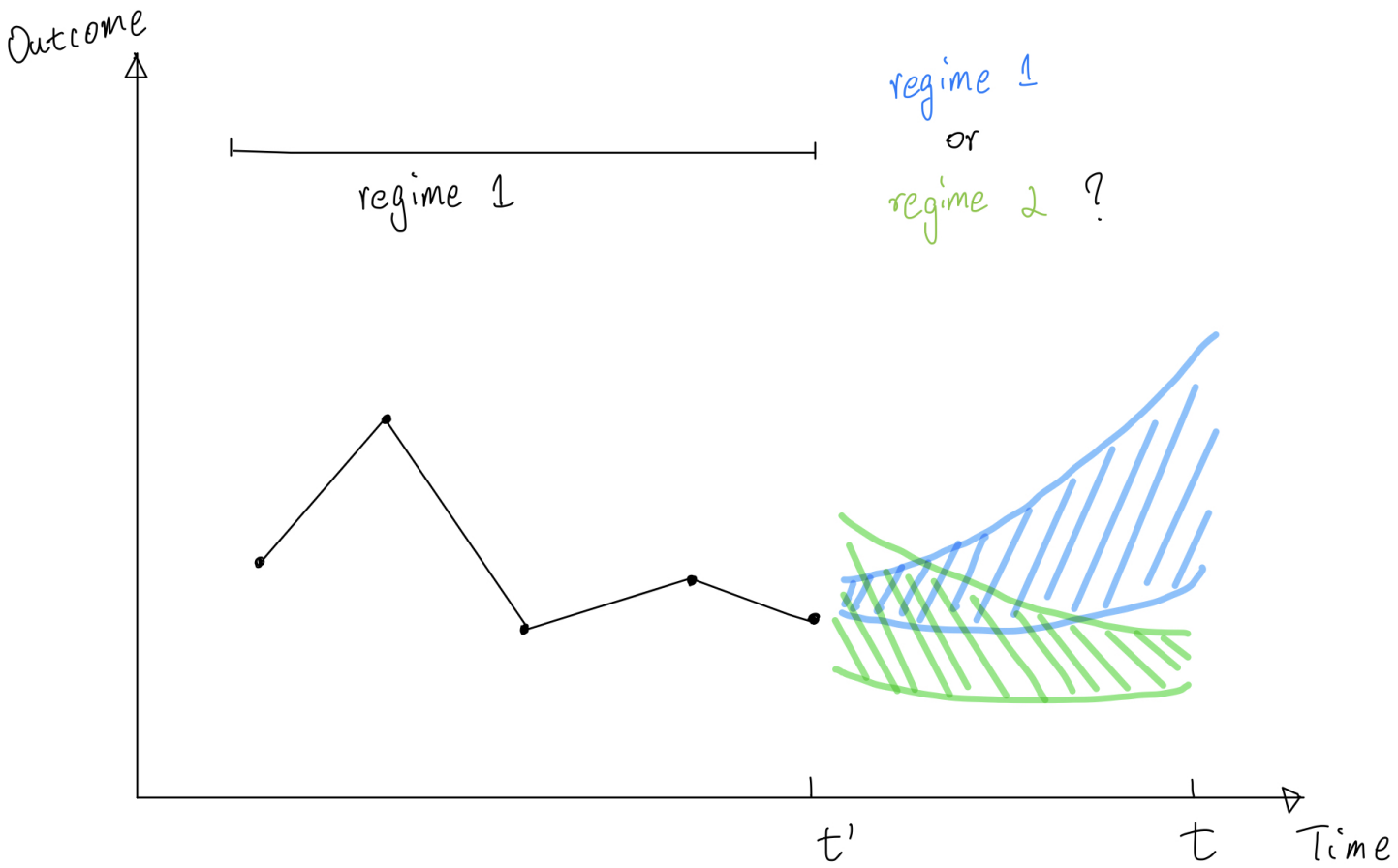
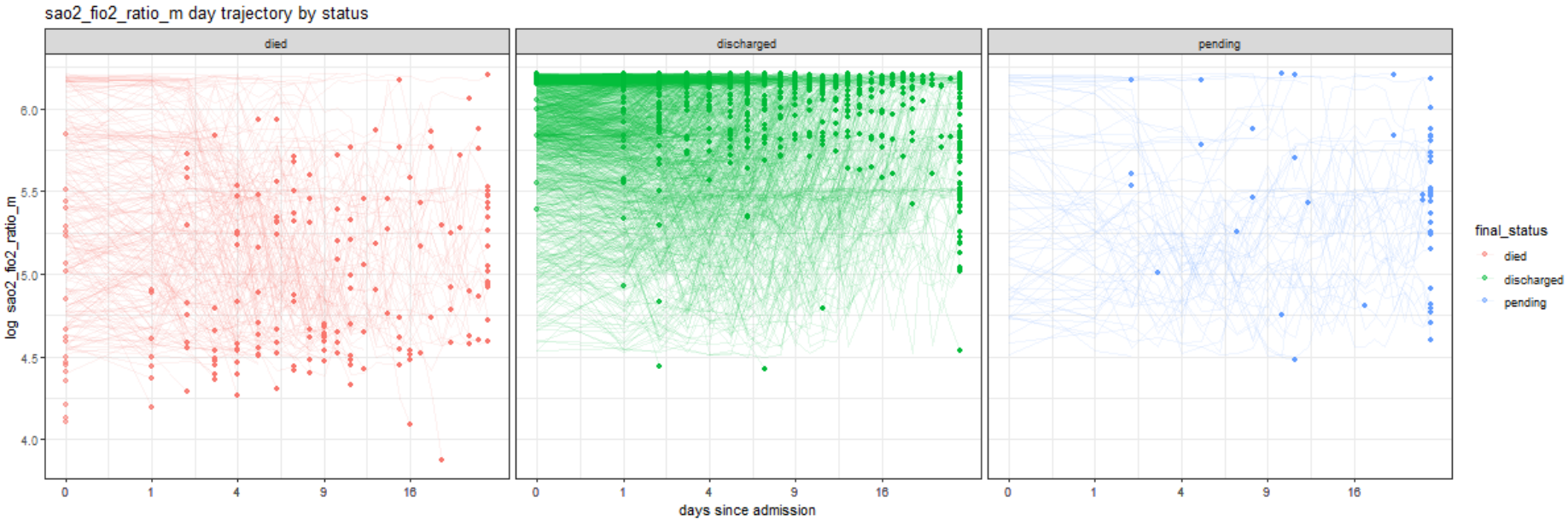
Causal Framework for Individualized Treatment Evaluation using Multivariate Generalized Mixed Effect Models with Longitudinal Data

Motivation



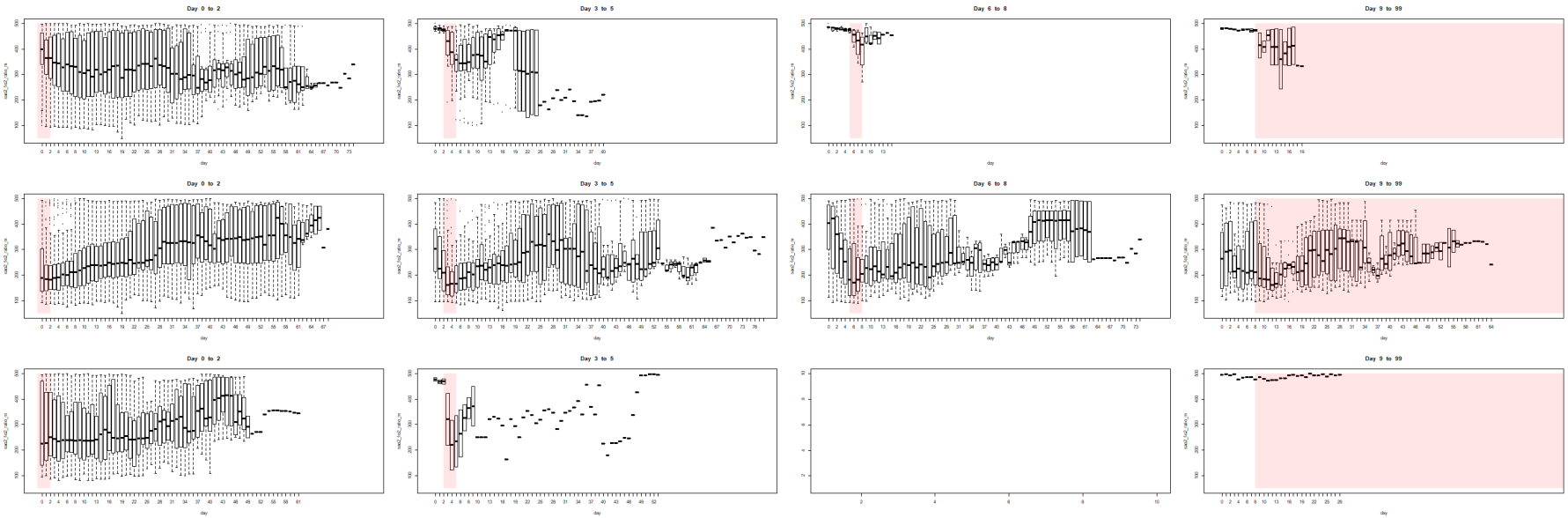
Individualized causal effect of treatment regimes on important biomarkers and endpoints

Covid Data



- Data source: The Johns Hopkins Covid-19 Precision Medicine Analytics Platform (PMAP) Registry
- All patients consecutively admitted with confirmed SARS-CoV-2 infection by microbiological testing from 3/4/2020-6/25/2020

Covid Data



- Understand how individual's mortality and important biomarkers vary as a function of ventilation initiation

Challenges

- Causal inference is usually about population causal effect.
  - Individual causal effect is unidentifiable in most cases.
- Informative patient heterogeneity in treatment assignment.
  - Clinician's tendency in making treatment decisions from unobserved factors; we assume the existence of unobserved time-invariant confounders.
  - With unmeasured confounding, standard approaches in longitudinal causal inference may lead to biased estimates. (*Yang and Lok, 2018*)
- Unit fixed effects regression models are widely used to adjust for unobserved time-invariant confounders
  - Drawback: trade-off between causal dynamics and time-invariant unobservables. (*Imai and Kim, 2019*)

Contributions

Propose a Bayesian framework for quantifying the individual counterfactual benefit of dynamic treatment regimes

- relaxes the no unmeasured confounders assumption
- summarizes the evidence relevant to clinical decisions of a single patient
- defines the causal effect for a specific patient at a day t as a functional of
  - observed history of that patient up to day t' ($t' < t$)
  - parameters estimated based on data from similar patients
  - counterfactual paths of that patient during days $(t', t]$, accounting for unobserved time-invariant confounders

Notations

For patient i at day t,

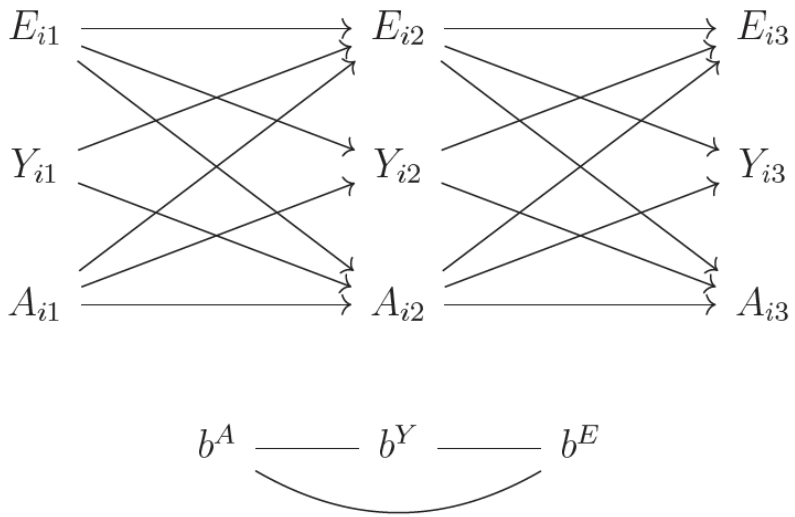
- Y_{it}: biomarker(s), i.e. continuous lung efficiency score
- E_{it}: competing risk endpoints, i.e. death and discharge
- A_{it}: indicator of first-time ventilation
- V: baseline covariates

Model Specification

Multivariate mixed model of time-varying components

$$Y_{it} = f_Y(V, A_{i,t-1}; \beta^Y, b_i^Y, \epsilon_{it})$$
$$A_{it} = f_A(V, A_{i,t-1}, Y_{i,t-1}; \beta^A, b_i^A)$$
$$E_{it} = f_E(V, A_{i,t-1}, Y_{i,t-1}; \beta^E, b_i^E)$$
$$(b_i^Y, b_i^A, b_i^E) \sim MVN(0, G)$$

The random effect (b_i^Y, b_i^A, b_i^E) captures a vector of unobserved time-invariant confounders in a flexible manner.

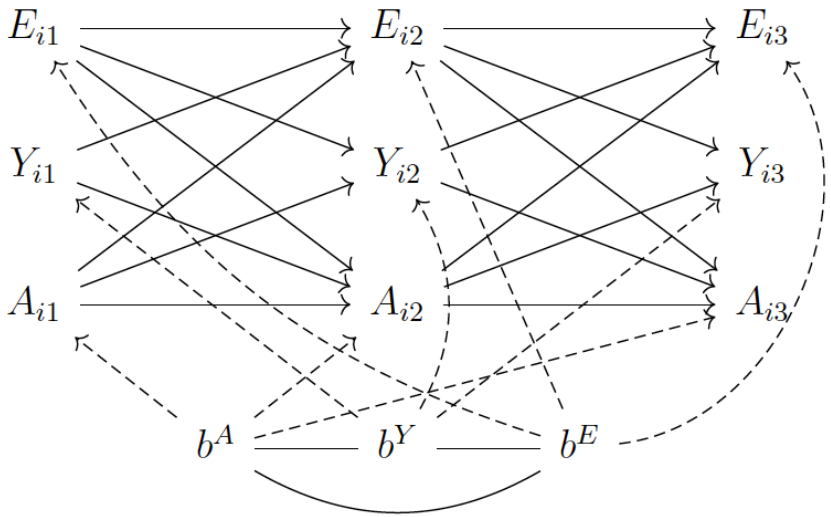


Model Specification

Multivariate mixed model of time-varying components

$$Y_{it} = f_Y(V, A_{i,t-1}; \beta^Y, b_i^Y, \epsilon_{it})$$
$$A_{it} = f_A(V, A_{i,t-1}, Y_{i,t-1}; \beta^A, b_i^A)$$
$$E_{it} = f_E(V, A_{i,t-1}, Y_{i,t-1}; \beta^E, b_i^E)$$
$$(b_i^Y, b_i^A, b_i^E) \sim MVN(0, G)$$

The random effect (b_i^Y, b_i^A, b_i^E) captures a vector of unobserved time-invariant confounders in a flexible manner.



Model Specification

Why multivariate mixed model?

- Repeated measurement makes it possible to partially inform individual heterogeneity that arises from unobserved covariates
- Connects the progression of biomarkers, events, and treatment assignments

Difficulties

- Estimation of patient heterogeneity usually uses complete history information
- Longitudinal causal estimation only uses up-to-date history information to infer on future counterfactuals

Model Estimation

Obtain posterior estimates of population level parameters $(\beta^Y, \beta^A, \beta^E, G)$

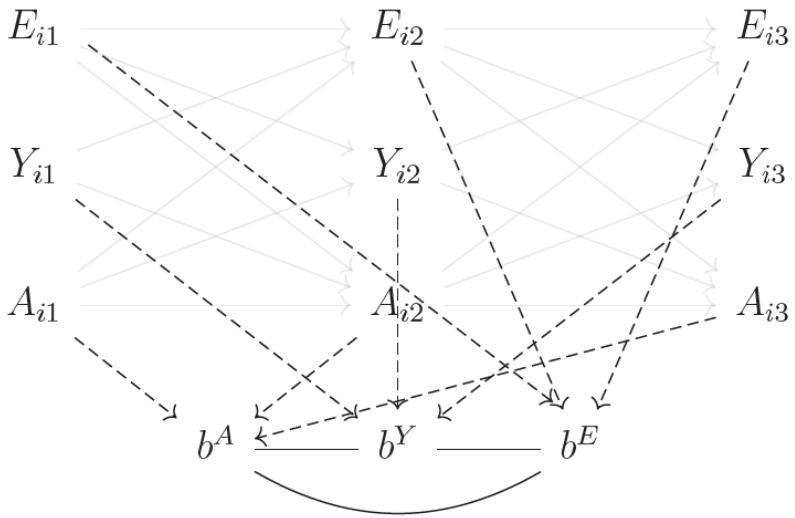


Figure: Information flow in the estimation of G

A "G-computation" Approach

Posterior sampling of individual counterfactual trajectories under a treatment regime of interest

Given history $(E_{i1}, Y_{i1}, A_{i1}, E_{i2}, Y_{i2})$, regime $(A_{i1}, A_{i2}) = (a_1, a_2) = \bar{a}$

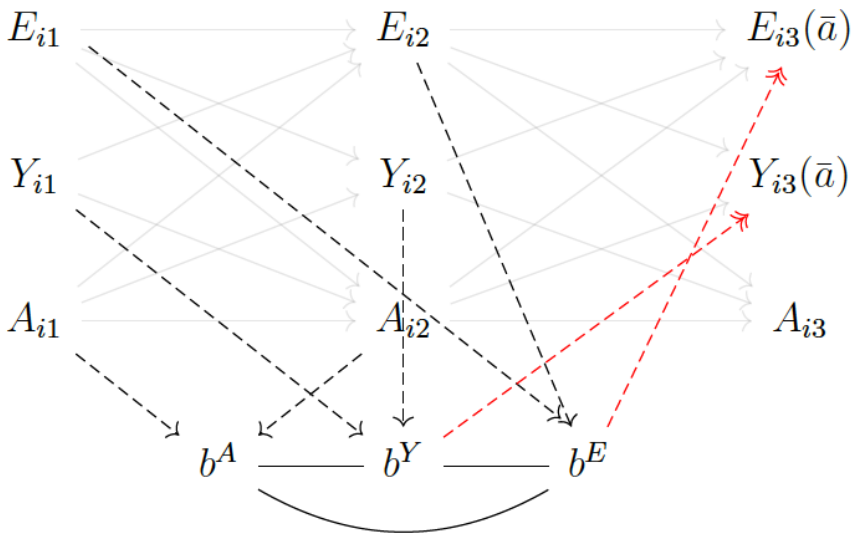


Figure: Information flow in computing potential outcomes

- Real-time update of random effect estimation
- Conditional on treatment heterogeneity b_i^A for the identifiability of individualized treatment effect based on the assumed DAG

A "G-computation" Approach

Posterior sampling of individual counterfactual trajectories under a treatment regime of interest

Given history $(E_{i1}, Y_{i1}, A_{i1}, E_{i2}, Y_{i2})$, regime $(A_{i1}, A_{i2}) = (a_1, a_2) = \bar{a}$

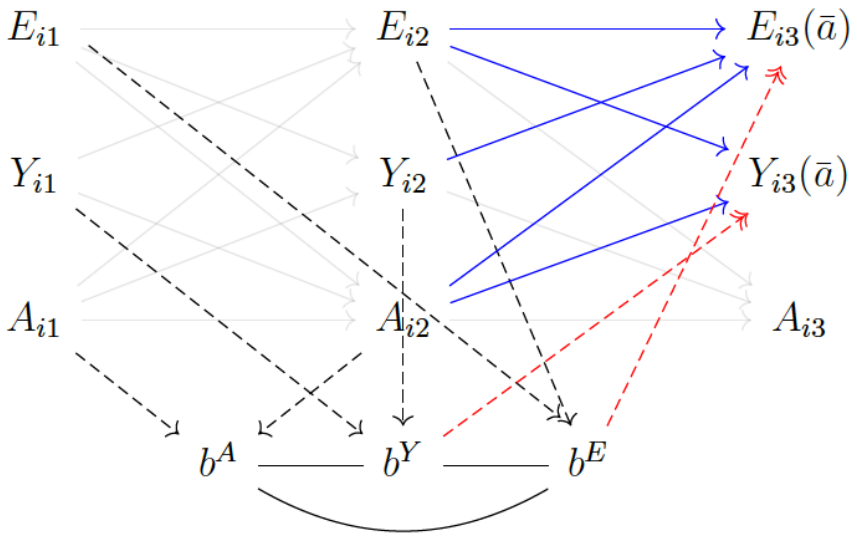


Figure: Information flow in computing potential outcomes

Causal Comparison

Compute target quantity for treatment regime comparison as a functional of the posterior counterfactual trajectories under different treatment options

- Example: Contrasts of cumulative hazards between two treatment regimes.

Preliminary Resuts

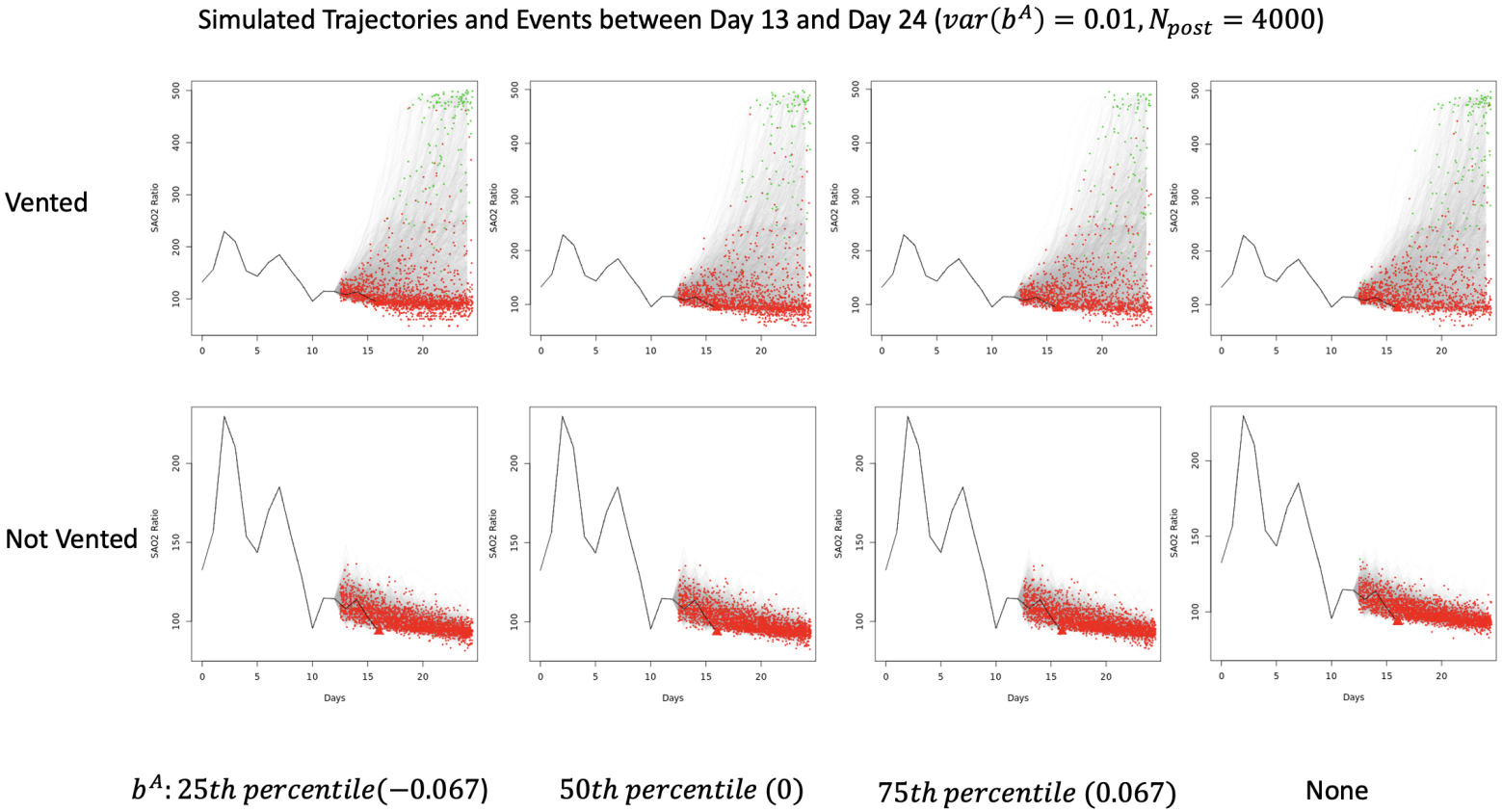


Figure: Venting the person on day 12 versus always no ventilation

Preliminary Resuts