Analysis of Ordinal Longitudinal Data under Case-Control Sampling: Studying Mortality in Critically Ill Patients

Chiara Di Gravio
CMStatistics 2023
December 16, 2023

The CLOVERS Study

The CLOVERS study was a randomized clinical trial comparing the effect of two resuscitation strategies on mortality and ARDS

The trial recruited 1,563 hospitalized patients with sepsis before being stopped due to inefficacy

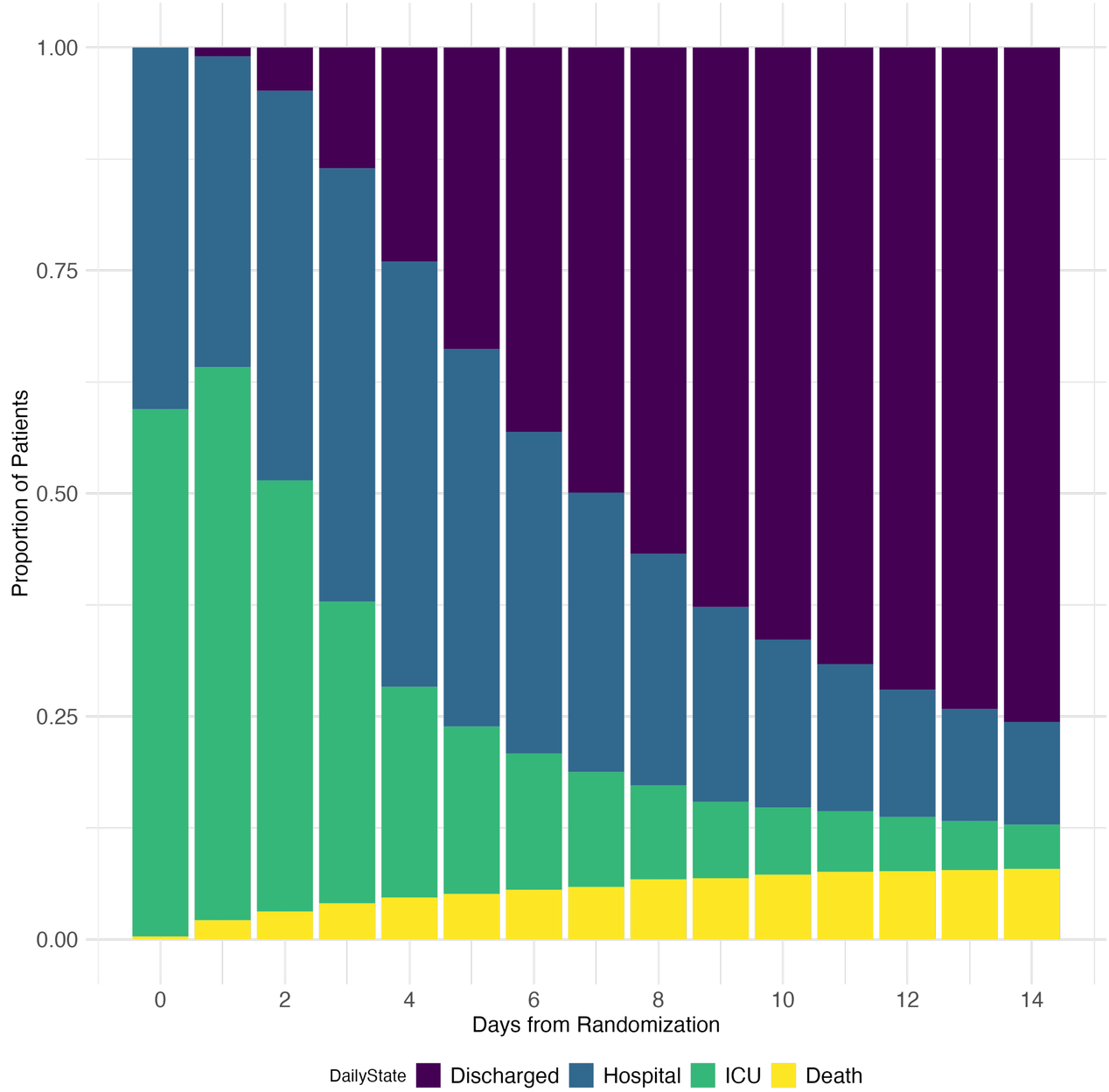
At recruitment, blood samples were collected and stored for later use

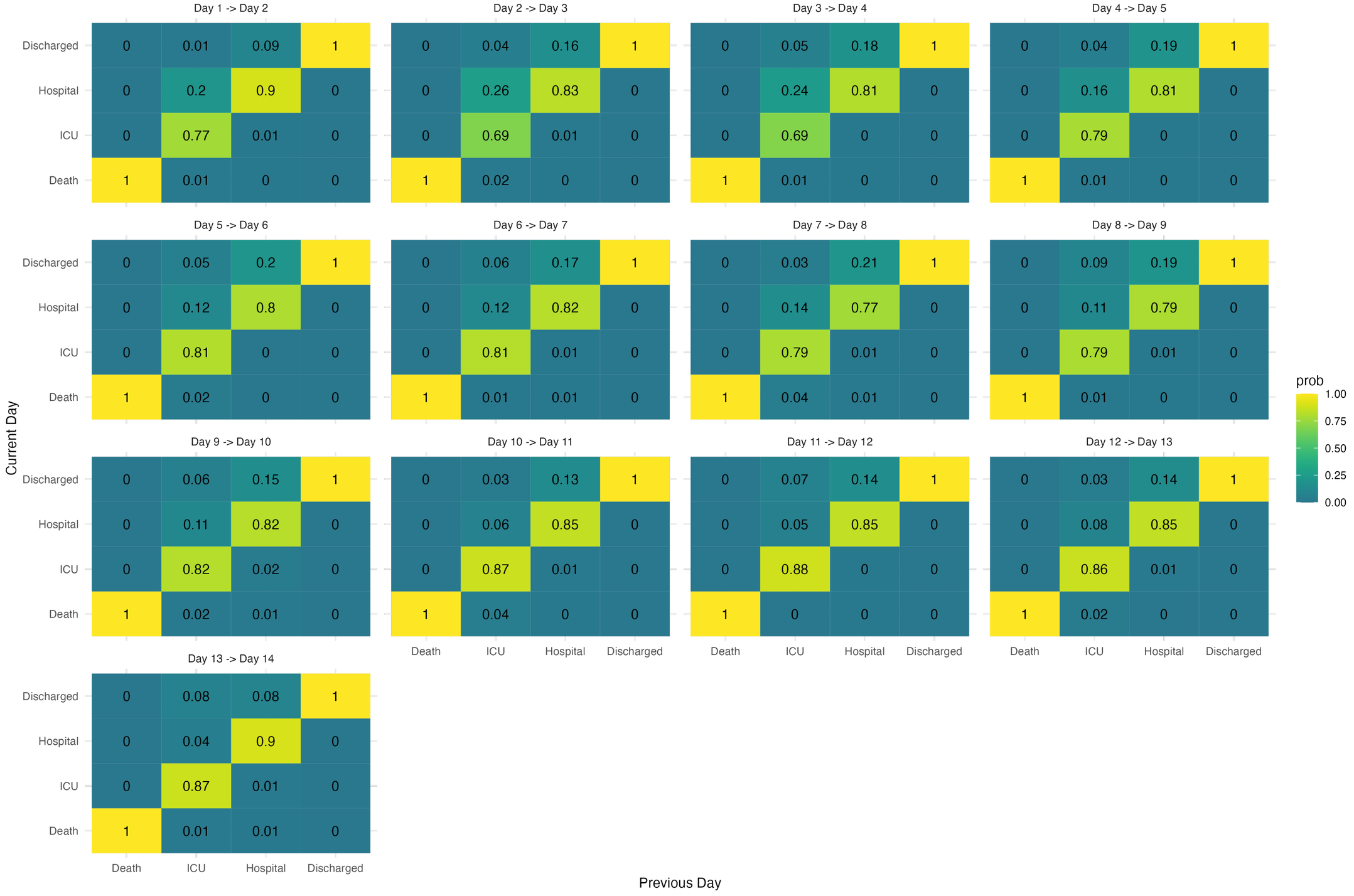
We want to use the blood samples collected in the CLOVERS study to:

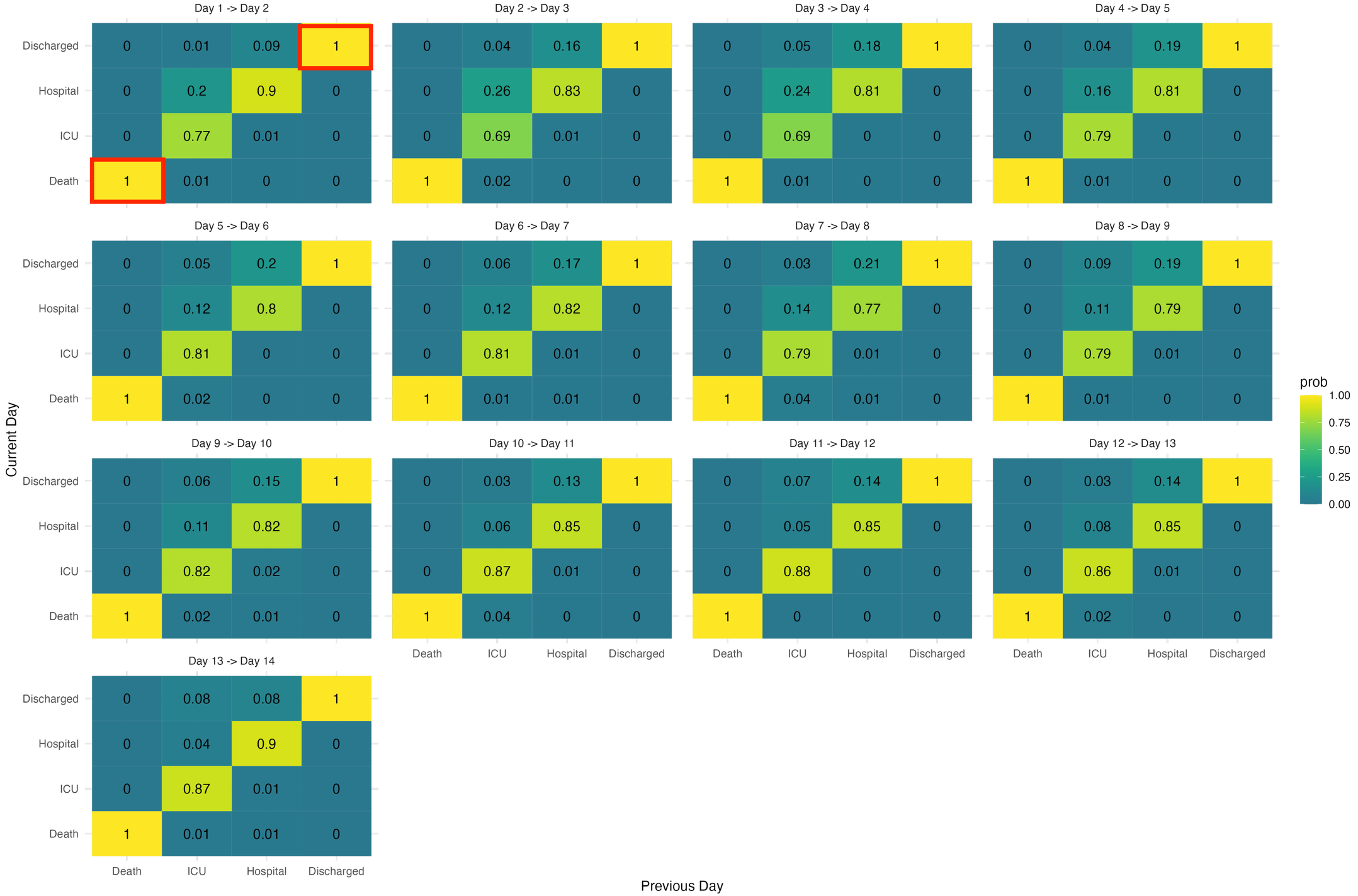
- Measure levels of glycocalyx degradation
- Study the relationship between glycocalyx degradation and mortality

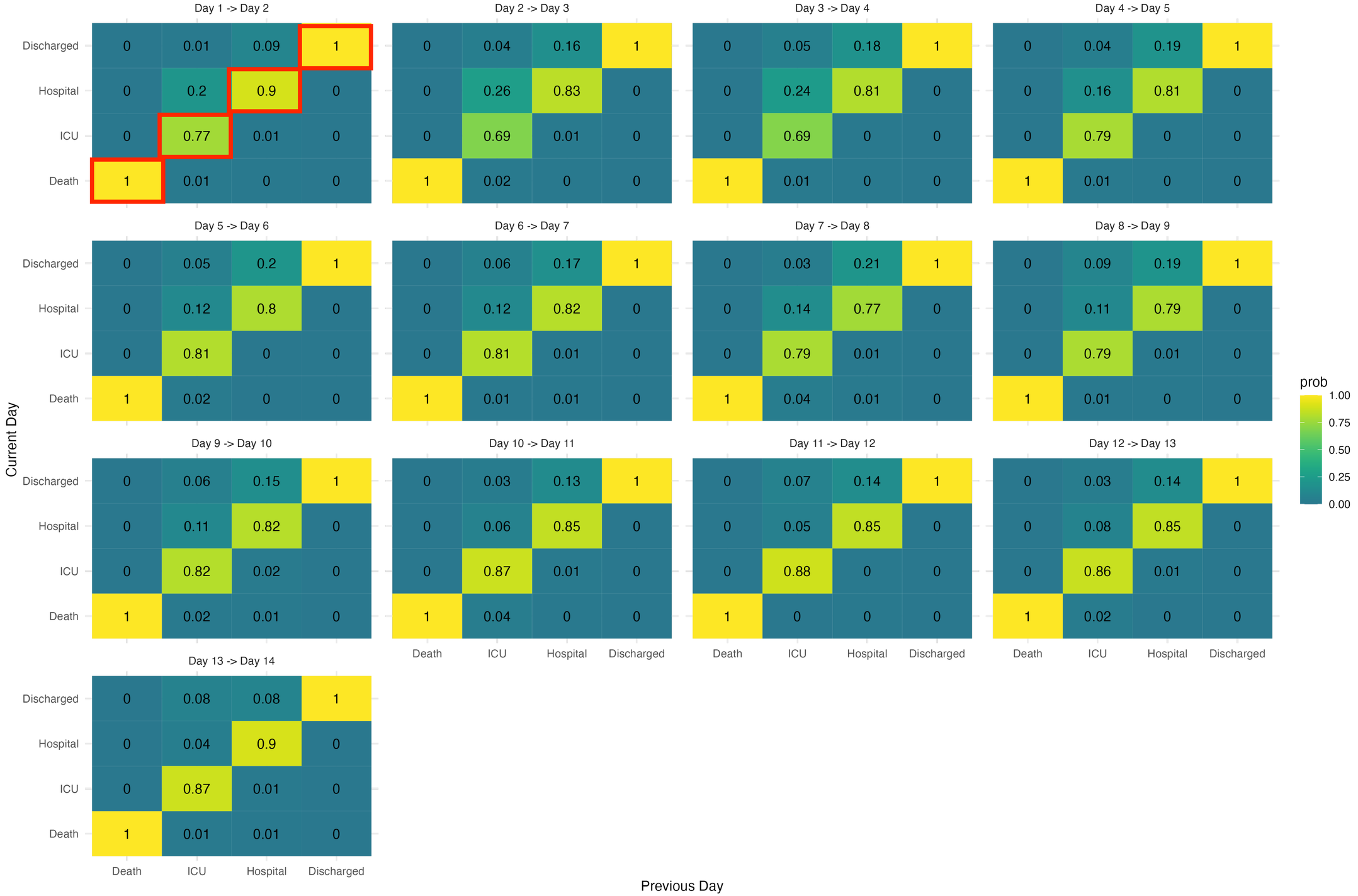
For the first 14 days we have information on where each patient was:

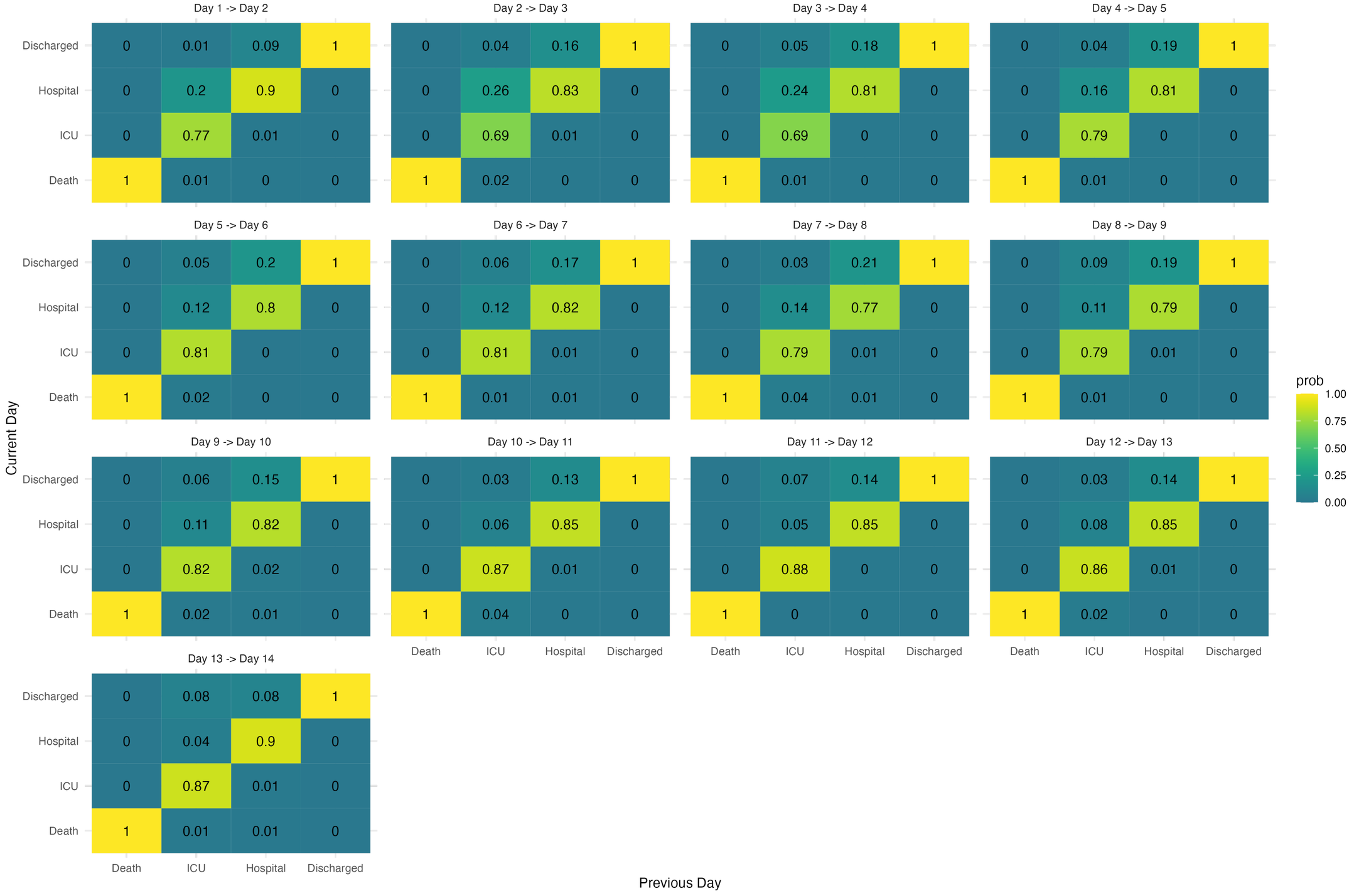
- Discharge
- Hospital
- Hospital + ICU
- Death




Day 1 -> Day 2
Day 2 -> Day 3
Day 3 -> Day 4
Day 4 -> Day 5
Day 5 -> Day 6
Day 6 -> Day 7
Day 7 -> Day 8
Day 8 -> Day 9
Day 9 -> Day 10
Day 10 -> Day 11
Day 11 -> Day 12
Day 12 -> Day 13
Day 13 -> Day 14
Discharged
Hospital
ICU
Death
Current Day
Previous Day
prob
1.00
0.75
0.50
0.25
0.00


Day 1 -> Day 2
Day 2 -> Day 3
Day 3 -> Day 4
Day 4 -> Day 5
Day 5 -> Day 6
Day 6 -> Day 7
Day 7 -> Day 8
Day 8 -> Day 9
Day 9 -> Day 10
Day 10 -> Day 11
Day 11 -> Day 12
Day 12 -> Day 13
Day 13 -> Day 14
Discharged
Hospital
ICU
Death
Current Day
Previous Day
prob
1.00
0.75
0.50
0.25
0.00


Day 1 -> Day 2
Day 2 -> Day 3
Day 3 -> Day 4
Day 4 -> Day 5
Day 5 -> Day 6
Day 6 -> Day 7
Day 7 -> Day 8
Day 8 -> Day 9
Day 9 -> Day 10
Day 10 -> Day 11
Day 11 -> Day 12
Day 12 -> Day 13
Day 13 -> Day 14
Discharged
Hospital
ICU
Death
Current Day
Death
ICU
Hospital
Discharged
Previous Day
prob
1.00
0.75
0.50
0.25
0.00


Day 1 -> Day 2
Day 2 -> Day 3
Day 3 -> Day 4
Day 4 -> Day 5
Day 5 -> Day 6
Day 6 -> Day 7
Day 7 -> Day 8
Day 8 -> Day 9
Day 9 -> Day 10
Day 10 -> Day 11
Day 11 -> Day 12
Day 12 -> Day 13
Day 13 -> Day 14
Discharged
Hospital
ICU
Death
Current Day
Previous Day
prob
1.00
0.75
0.50
0.25
0.00

Who are we going to sample?

Budget and time constraints allowed us to collect information on glycocalyx degradation on 600 of the 1,563 patients enrolled in the CLOVERS trial

Who are we going to sample?

Budget and time constraints allowed us to collect information on glycocalyx degradation on 600 of the 1,563 patients enrolled in the CLOVERS trial

Everyone who died and/or developed ARDS are sampled with probability one. The remaining patients are sampled using simple random sampling until we reached a total of 600 patients

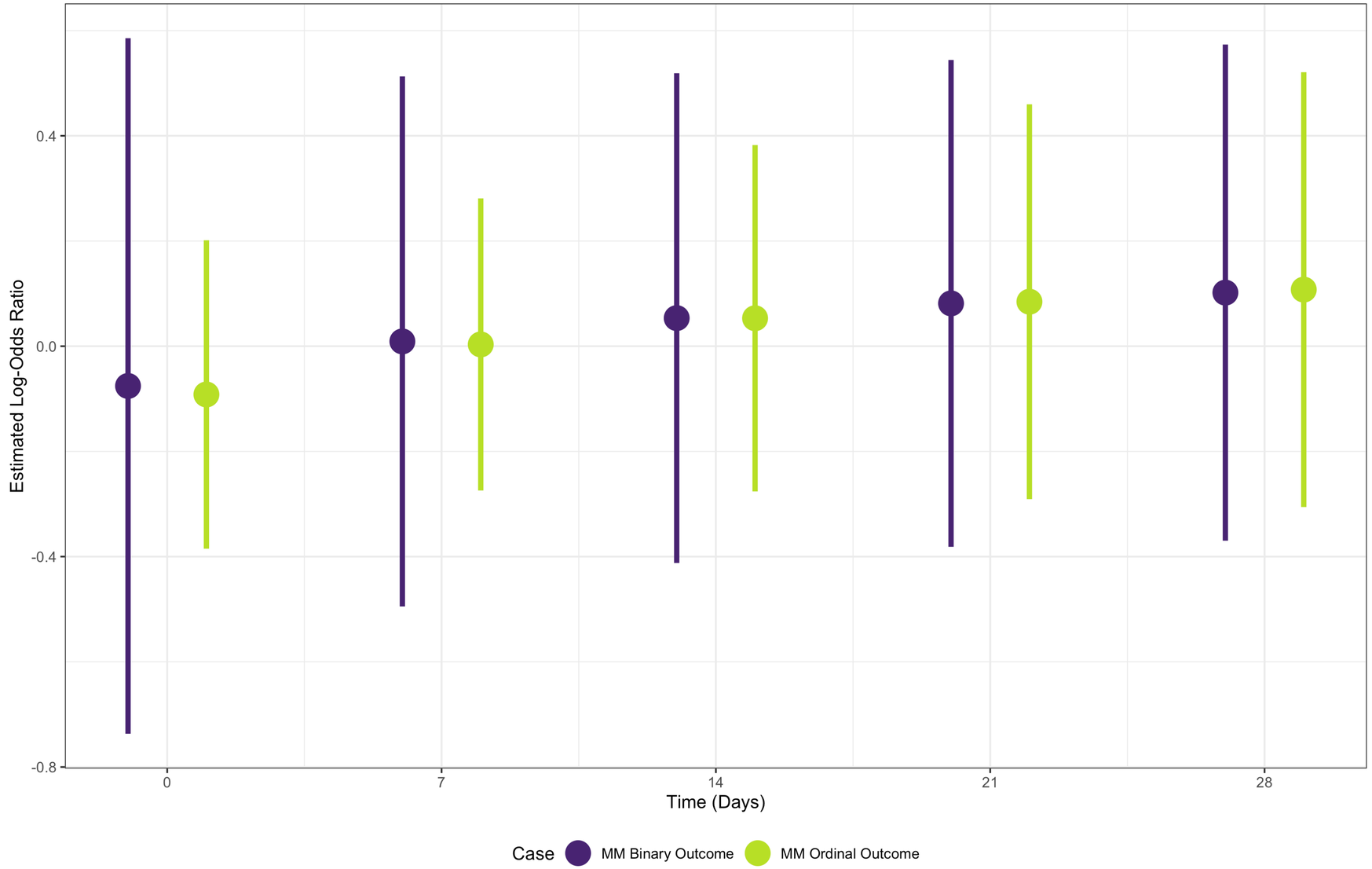
How are we going to analyze the data?

We want to understand the relationship between glycocalyx degradation (yes/no) and mortality at day 1 and at day 14

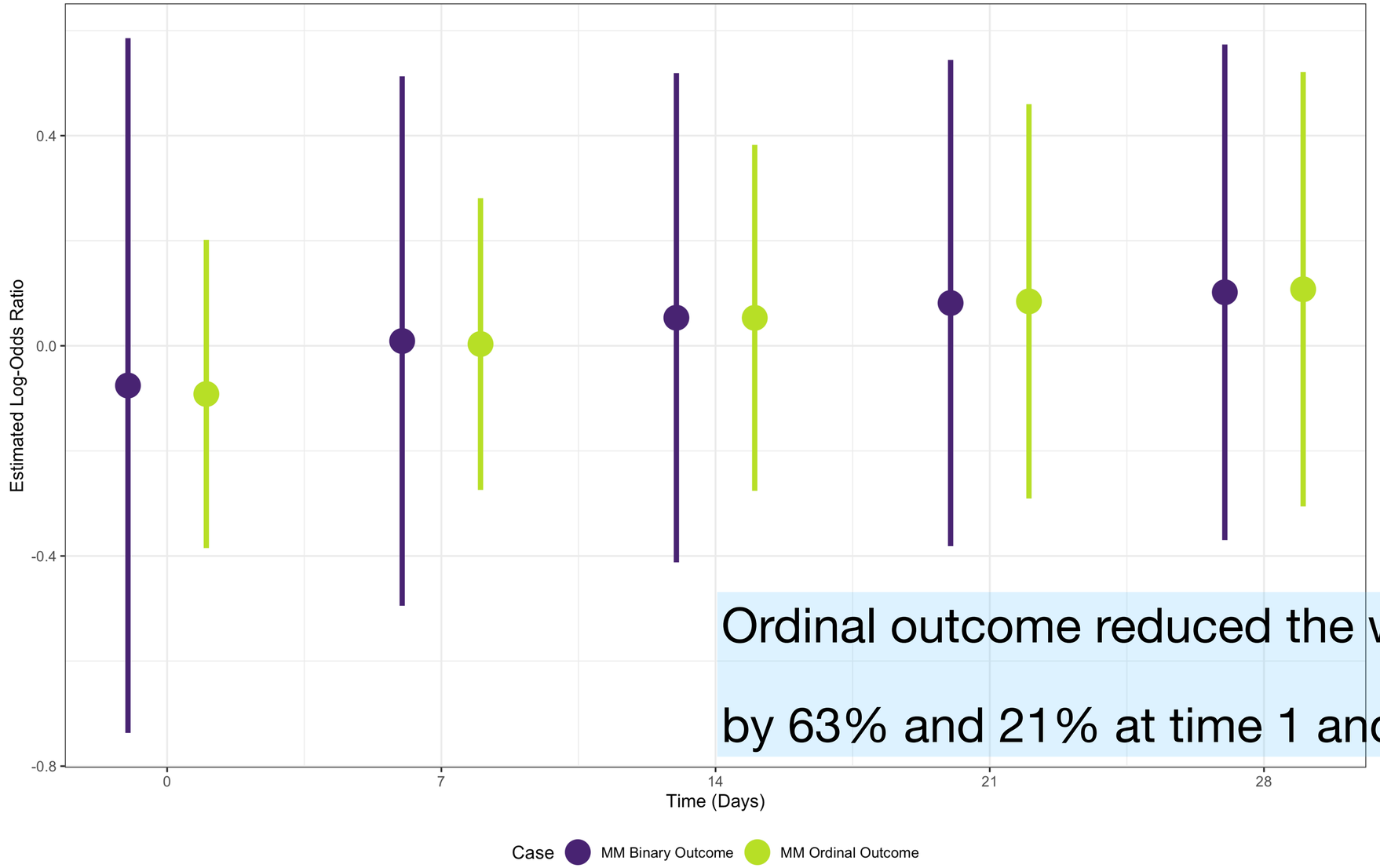
We define the outcome in two ways:

- binary longitudinal outcome: death vs alive (hospital, hospital+ICU, discharge)
- ordinal longitudinal outcome: death, hospital + ICU, hospital, discharge

Why Aren’t We Only Looking at a Binary Outcome?



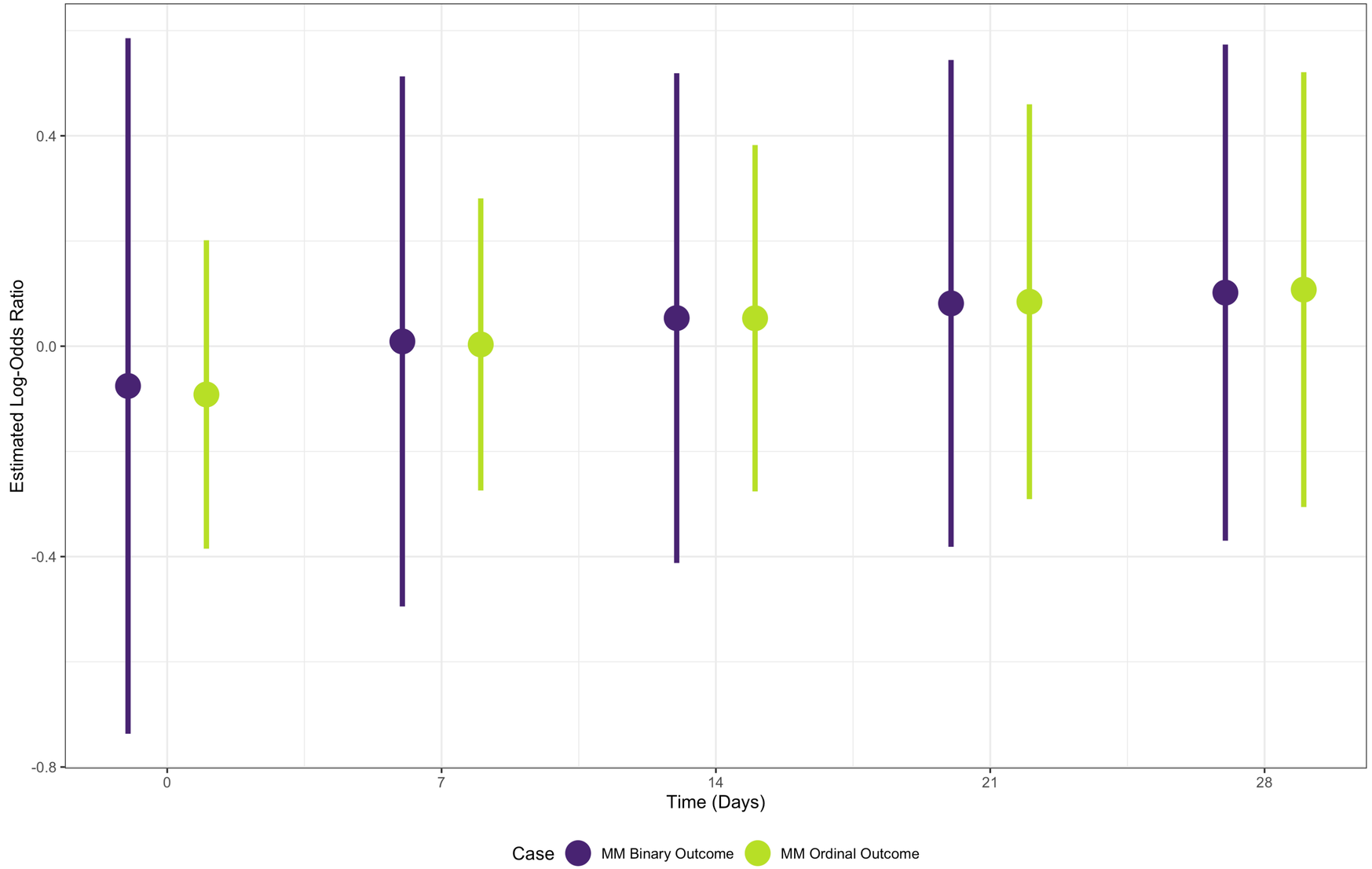
Why Aren’t We Only Looking at a Binary Outcome?



Ordinal outcome reduced the width of the C.I.

by 63% and 21% at time 1 and 28

Why Aren’t We Only Looking at a Binary Outcome?



The Marginalized Transition Model

The Model

The marginalized transition model is identified by two generalized linear models:

$$h\{E(Y_{ij} \mid X_i, \boldsymbol{Z}_i, T_{ij})\} = \alpha_0 + \beta_x X_i + \beta_t T_{ij} + \beta_{xt} X_i T_{ij} + \boldsymbol{\beta}_z^{\mathrm{T}} \boldsymbol{Z}_i$$

$$g\{E(Y_{ij} \mid X_i, \boldsymbol{Z}_i, Y_{i(j-1)})\} = \Delta_{ijk} + \sum_{s=1}^{K-1} \gamma^{ks} I\{Y_{i(j-1)} = s\}$$

where:

- Y_{ij} is the outcome state for subject i at time j
- K is the total number of states
- X_i is an indicator of the presence of glycocalyx degradation
- $\boldsymbol{Z}_i$ is a matrix of baseline covariates
- Δ_{ijk} links the marginal and the conditional mean model

The Model

The marginalized transition model is identified by two generalized linear models:

Marginal Mean Model

$$h\{E(Y_{ij} \mid X_i, \boldsymbol{Z}_i, T_{ij})\} = \alpha_0 + \beta_x X_i + \beta_t T_{ij} + \beta_{xt} X_i T_{ij} + \boldsymbol{\beta}_z^{\mathrm{T}} \boldsymbol{Z}_i$$

$$g\{E(Y_{ij} \mid X_i, \boldsymbol{Z}_i, Y_{i(j-1)})\} = \Delta_{ijk} + \sum_{s=1}^{K-1} \gamma^{ks} I\{Y_{i(j-1)} = s\}$$

where:

- Y_{ij} is the outcome state for subject i at time j
- K is the total number of states
- X_i is an indicator of the presence of glycocalyx degradation
- $\boldsymbol{Z}_i$ is a matrix of baseline covariates
- Δ_{ijk} links the marginal and the conditional mean model

The Model

The marginalized transition model is identified by two generalized linear models:

Marginal Mean Model

$$h\{E(Y_{ij} \mid X_i, \boldsymbol{Z}_i, T_{ij})\} = \alpha_0 + \beta_x X_i + \beta_t T_{ij} + \beta_{xt} X_i T_{ij} + \boldsymbol{\beta}_z^{\mathrm{T}} \boldsymbol{Z}_i$$

Conditional Mean Model

$$g\{E(Y_{ij} \mid X_i, \boldsymbol{Z}_i, Y_{i(j-1)})\} = \Delta_{ijk} + \sum_{s=1}^{K-1} \gamma^{ks} I\{Y_{i(j-1)} = s\}$$

where:

- Y_{ij} is the outcome state for subject i at time j
- K is the total number of states
- X_i is an indicator of the presence of glycocalyx degradation
- $\boldsymbol{Z}_i$ is a matrix of baseline covariates
- Δ_{ijk} links the marginal and the conditional mean model

A natural choice for the link functions is the logit link:

$$\text{logit}\{P(\boldsymbol{Y}_i \leq k \,|\, X_i, T_{ij}, \boldsymbol{Z}_i)\} = \alpha_{0,k} + \beta_x X_i + \beta_t T_{ij} + \beta_{xt} X_i T_{ij} + \boldsymbol{\beta}_z^{\mathrm{T}} \boldsymbol{Z}_i$$

$$\log \left\{ \frac{P(Y_{ij} = k \,|\, Y_{i(j-1)}, X_i, T_{ij}, \boldsymbol{Z}_i)}{P(Y_{ij} = K \,|\, Y_{i(j-1)}, X_i, T_{ij}, \boldsymbol{Z}_i)} \right\} = \Delta_{ijk} + \sum_{s=1}^{K-1} \gamma^{ks} I\{Y_{i(j-1)} = s\}$$

- When K = 2 (binary case), the marginal mean and the conditional mean models become logistic regression models
- When K > 2 (ordinal case), the marginal mean model is a proportional odds model while the conditional mean model is a multinomial regression

Dealing with Absorbing States

The marginal mean model assumes that the association between the ordinal outcome and time is captured by a single coefficient. This is **not true** when there are absorbing states

We relax the proportional odds assumption for time in the marginal mean model

$$\text{logit}\{P(\boldsymbol{Y}_i \leq k \,|\, X_i, T_{ij}, \boldsymbol{Z}_i)\} = \alpha_0 + \beta_x X_i + T_{ij}[\beta_{t,1} + \beta_{t,2} I(Y \leq 2) + \beta_{t,3} I(Y \leq 3)]$$
$$+ \beta_{xt} X_i T_{ij} + \boldsymbol{\beta}_z^{\mathrm{T}} \boldsymbol{Z}_i$$

The difference in the log-odds of death for those with and without glycocalyx degradation at time T_{ij} is $\beta_x + T_{ij}\beta_{xt}$

We relax the proportional odds assumptions for other variables

Dealing with Absorbing States

The marginal mean model assumes that the association between the ordinal outcome and time is captured by a single coefficient. This is **not true** when there are absorbing states

We relax the proportional odds assumption for time in the marginal mean model

$$\text{logit}\{P(\boldsymbol{Y}_i \leq k \,|\, X_i, T_{ij}, \boldsymbol{Z}_i)\} = \alpha_0 + \beta_x X_i + T_{ij}[\beta_{t,1} + \beta_{t,2} I(Y \leq 2) + \beta_{t,3} I(Y \leq 3)]$$
$$+ \beta_{xt} X_i T_{ij} + \boldsymbol{\beta}_z^{\mathrm{T}} \boldsymbol{Z}_i$$

The difference in the log-odds of death for those with and without glycocalyx degradation at time T_{ij} is $\beta_x + T_{ij}\beta_{xt}$

We relax the proportional odds assumptions for other variables

Dealing with Absorbing States

The marginal mean model assumes that the association between the ordinal outcome and time is captured by a single coefficient. This is **not true** when there are absorbing states

We relax the proportional odds assumption for time in the marginal mean model

$$\text{logit}\{P(\boldsymbol{Y}_i \leq k \,|\, X_i, T_{ij}, \boldsymbol{Z}_i)\} = \alpha_0 + \beta_x X_i + T_{ij}[\beta_{t,1} + \beta_{t,2} I(Y \leq 2) + \beta_{t,3} I(Y \leq 3)]$$
$$+ \beta_{xt} X_i T_{ij} + \boldsymbol{\beta}_z^{\mathrm{T}} \boldsymbol{Z}_i$$

The difference in the log-odds of death for those with and without glycocalyx degradation at time T_{ij} is $\beta_x + T_{ij}\beta_{xt}$

We relax the proportional odds assumptions for other variables

Dealing with Absorbing States

The marginal mean model assumes that the association between the ordinal outcome and time is captured by a single coefficient. This is **not true** when there are absorbing states

We relax the proportional odds assumption for time in the marginal mean model

$$\text{logit}\{P(\boldsymbol{Y}_i \leq k \,|\, X_i, T_{ij}, \boldsymbol{Z}_i)\} = \alpha_0 + \beta_x X_i + T_{ij}[\beta_{t,1} + \beta_{t,2} I(Y \leq 2) + \beta_{t,3} I(Y \leq 3)]$$

$$+\beta_{xt} X_i T_{ij} + \boldsymbol{\beta}_z^{\mathrm{T}} \boldsymbol{Z}_i$$

The difference in the log-odds of death for those with and without glycocalyx degradation at time T_{ij} is $\beta_x + T_{ij}\beta_{xt}$

We relax the proportional odds assumptions for other variables

Dealing with Absorbing States

The marginal mean model assumes that the association between the ordinal outcome and time is captured by a single coefficient. This is **not true** when there are absorbing states

We relax the proportional odds assumption for time in the marginal mean model

$$\text{logit}\{P(\boldsymbol{Y}_i \leq k \,|\, X_i, T_{ij}, \boldsymbol{Z}_i)\} = \alpha_0 + \beta_x X_i + T_{ij}[\beta_{t,1} + \beta_{t,2} I(Y \leq 2) + \beta_{t,3} I(Y \leq 3)]$$
$$+ \beta_{xt} X_i T_{ij} + \boldsymbol{\beta}_z^{\mathrm{T}} \boldsymbol{Z}_i$$

The difference in the log-odds of death for those with and without glycocalyx degradation at time T_{ij} is $\beta_x + T_{ij}\beta_{xt}$

We relax the proportional odds assumptions for other variables

Estimating Model's Parameters

Given a set of parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, the fitting algorithm can be summarised in three steps:

1. Compute Δ_{ijk} for each subject i, time j, and state k

2. Calculate the likelihood

$$\prod_{i=1}^{N} \left\{ \prod_{k=1}^{K} \left[P(Y_{i1}) = k \,|\, X_i, \boldsymbol{Z}_i \right]^{I(Y_{i1}=k)} \prod_{j=2}^{n_i} \prod_{k=1}^{k} \left[P(Y_{ij}) = k \,|\, Y_{ij-1}, X_i, \boldsymbol{Z}_i \right]^{I(Y_{ij}=k)} \right\}$$

3. Maximize the likelihood using a Newton-Raphson approach

The Estimation Procedures

The Estimation Procedures

1. **Weighted Likelihood Estimator**

   Considers only the 600 subjects with information on glycocalyx degradation, and weights their contribution to the likelihood by the inverse of their sampling probability

2. **SMLE**

   Includes the 600 subjects with information on glycocalyx degradation and information on outcome and covariates for the remaining subjects

3. **Multiple Imputation**

   Includes the 600 subjects with information on glycocalyx degradation and information on outcome and covariates for the remaining subjects

Multiple Imputation

Let V be an indicator of whether a subject has X measured. Our design samples based on the observed outcome:

$$P(X_i \mid \mathbf{Z}_i, \mathbf{Y}_i, V_i = 0) = P(X_i \mid \mathbf{Z}_i, \mathbf{Y}_i) = P(X_i \mid \mathbf{Z}_i, \mathbf{Y}_i, V_i = 1)$$

We build an imputation model for X in the unsampled subjects directly from the observed data without accounting for the design. Using Bayes' theorem, we derive

$$\log\left(\frac{P(\mathbf{Y}_i \mid \mathbf{X}_i \geq k, \mathbf{Z}_i)}{P(\mathbf{Y}_i \mid \mathbf{X}_i < K, \mathbf{Z}_i)}\right) + \log\left(\frac{P(\mathbf{X}_i \leq k \mid \mathbf{Z}_i)}{P(X_i < K \mid \mathbf{Z}_i)}\right)$$

We build *M* multiply imputed datasets, fit the target model to each one and combine the estimates using Rubin's rule

Multiple Imputation

Let V be an indicator of whether a subject has X measured. Our design samples based on the observed outcome:

$$P(X_i \,|\, \mathbf{Z}_i, \mathbf{Y}_i, V_i = 0) = P(X_i \,|\, \mathbf{Z}_i, \mathbf{Y}_i) = P(X_i \,|\, \mathbf{Z}_i, \mathbf{Y}_i, V_i = 1)$$

We build an imputation model for X in the unsampled subjects directly from the observed data without accounting for the design. Using Bayes' theorem, we derive

$$\log\left(\frac{P(\mathbf{Y}_i \,|\, \mathbf{X}_i \geq k, \mathbf{Z}_i)}{P(\mathbf{Y}_i \,|\, \mathbf{X}_i < K, \mathbf{Z}_i)}\right) + \log\left(\frac{P(\mathbf{X}_i \leq k \,|\, \mathbf{Z}_i)}{P(X_i < K \,|\, \mathbf{Z}_i)}\right)$$

We build *M* multiply imputed datasets, fit the target model to each one and combine the estimates using Rubin's rule

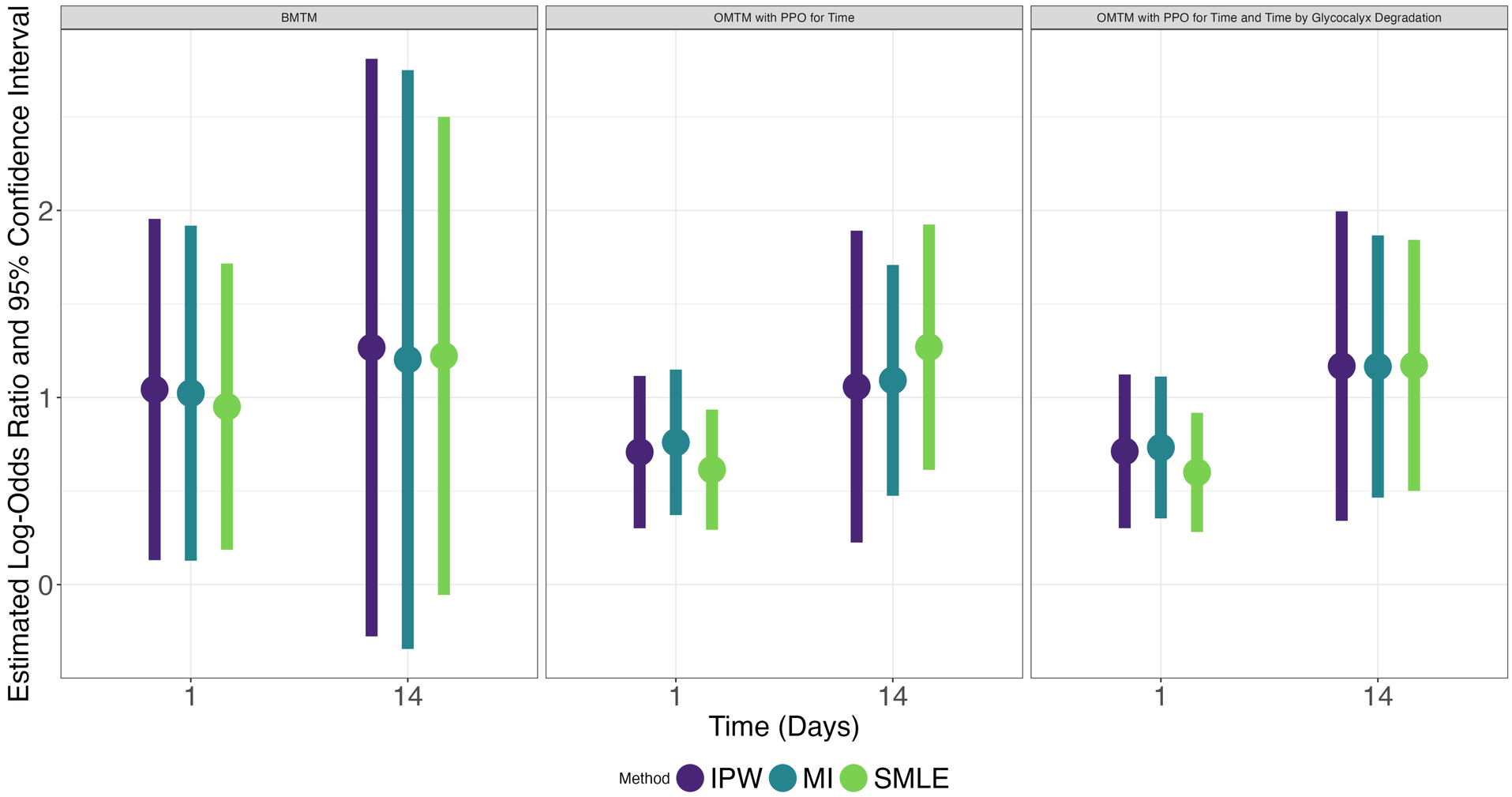
The Results

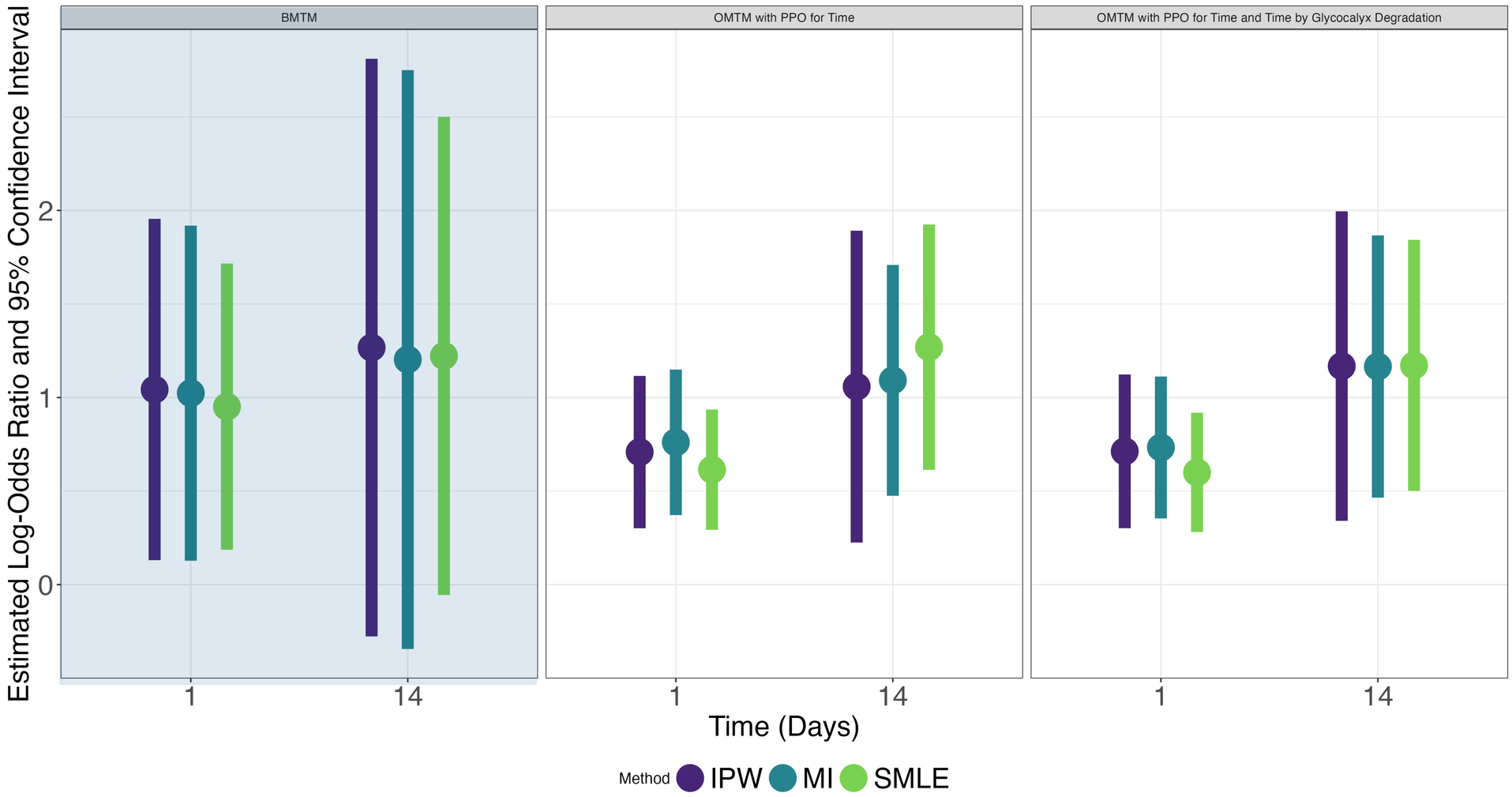
We considered three models:

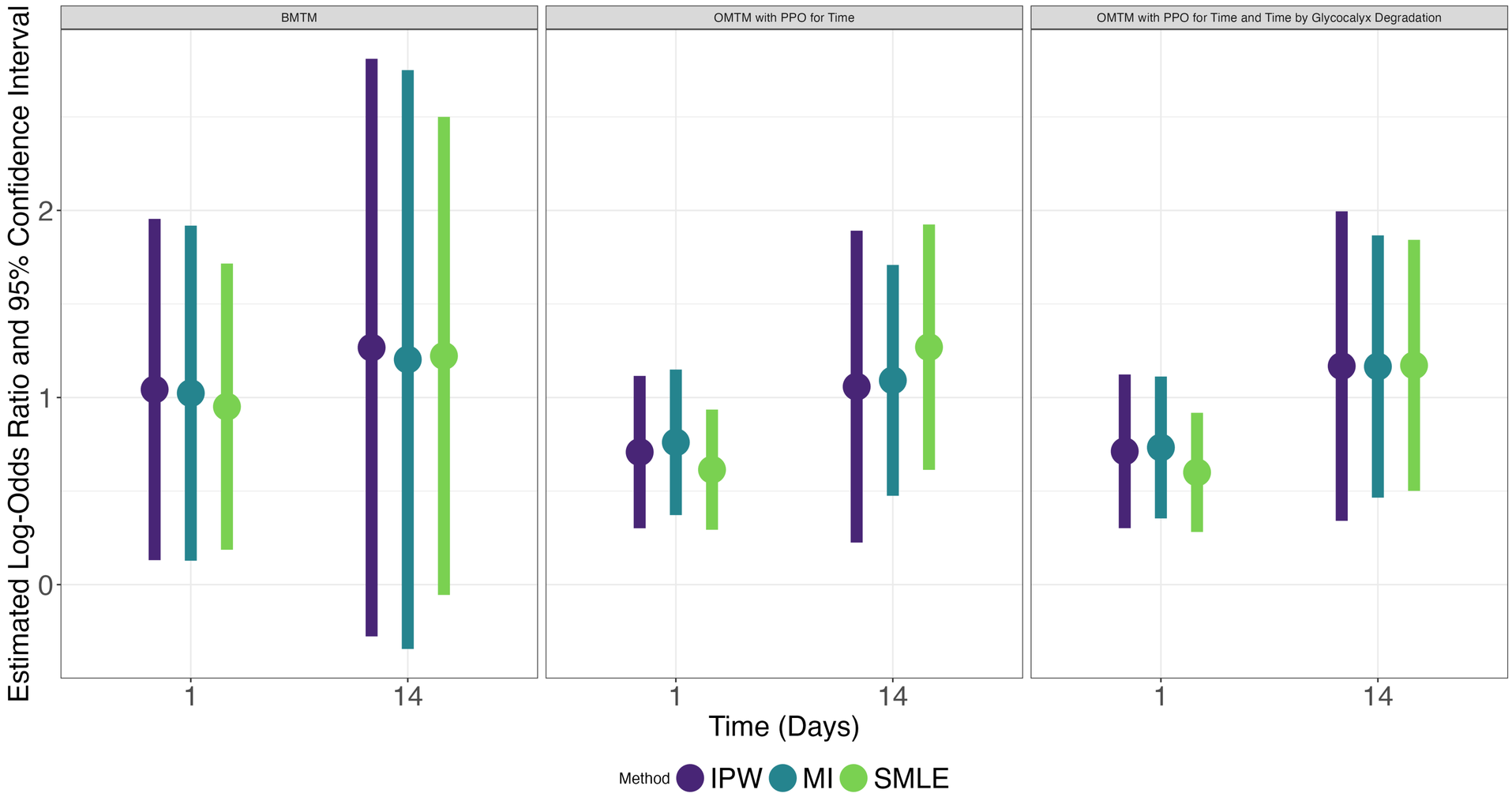
- Marginalized transition model with a binary outcome (BMTM)
- Marginalized transition model with an ordinal outcome (OMTM)
  - Relax the proportional odds assumption for time
  - Relax the proportional odds assumption for time and time by glycocalyx degradation

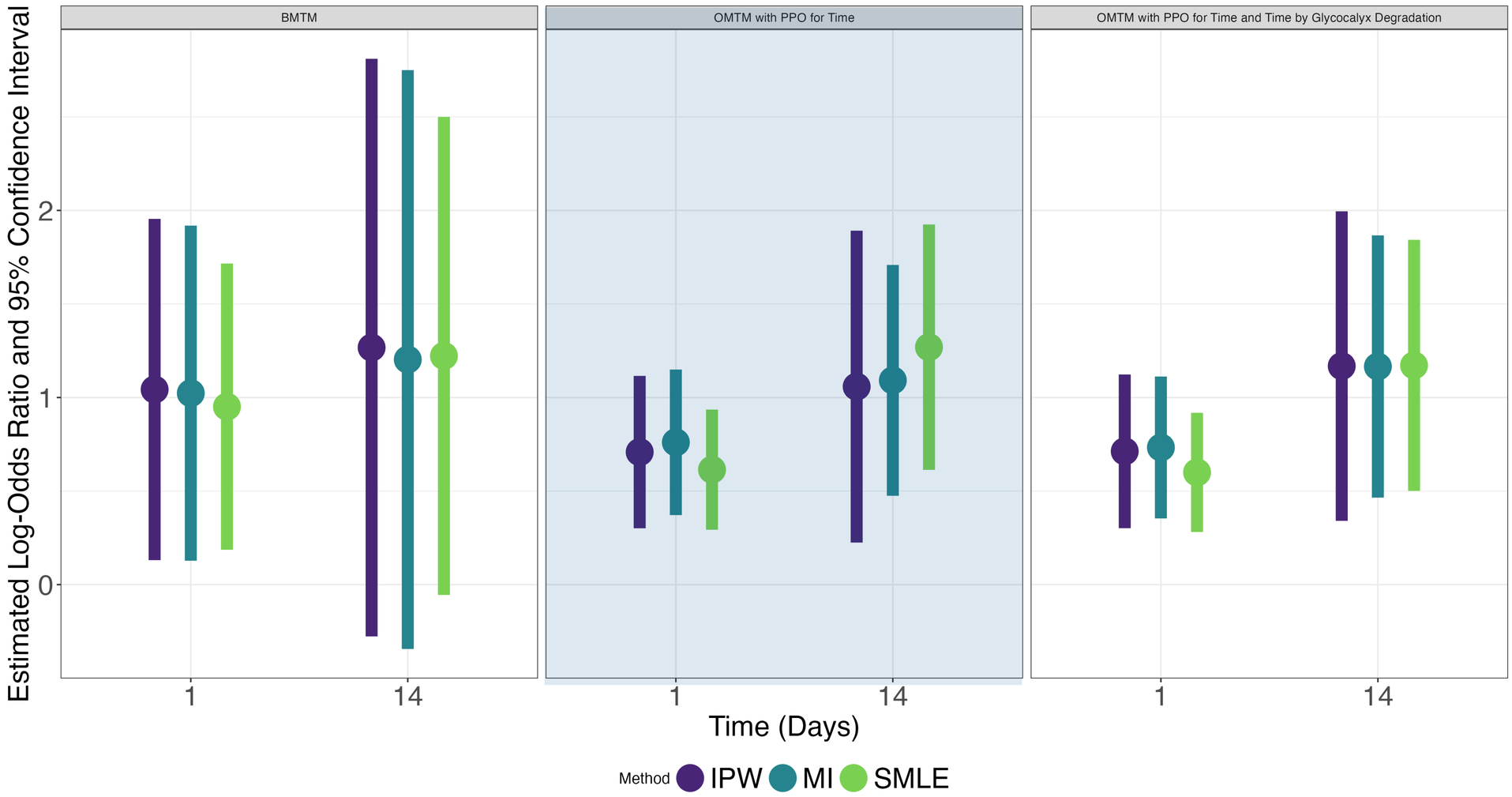
For each model we estimated the parameters with IPW, SMLE and MI

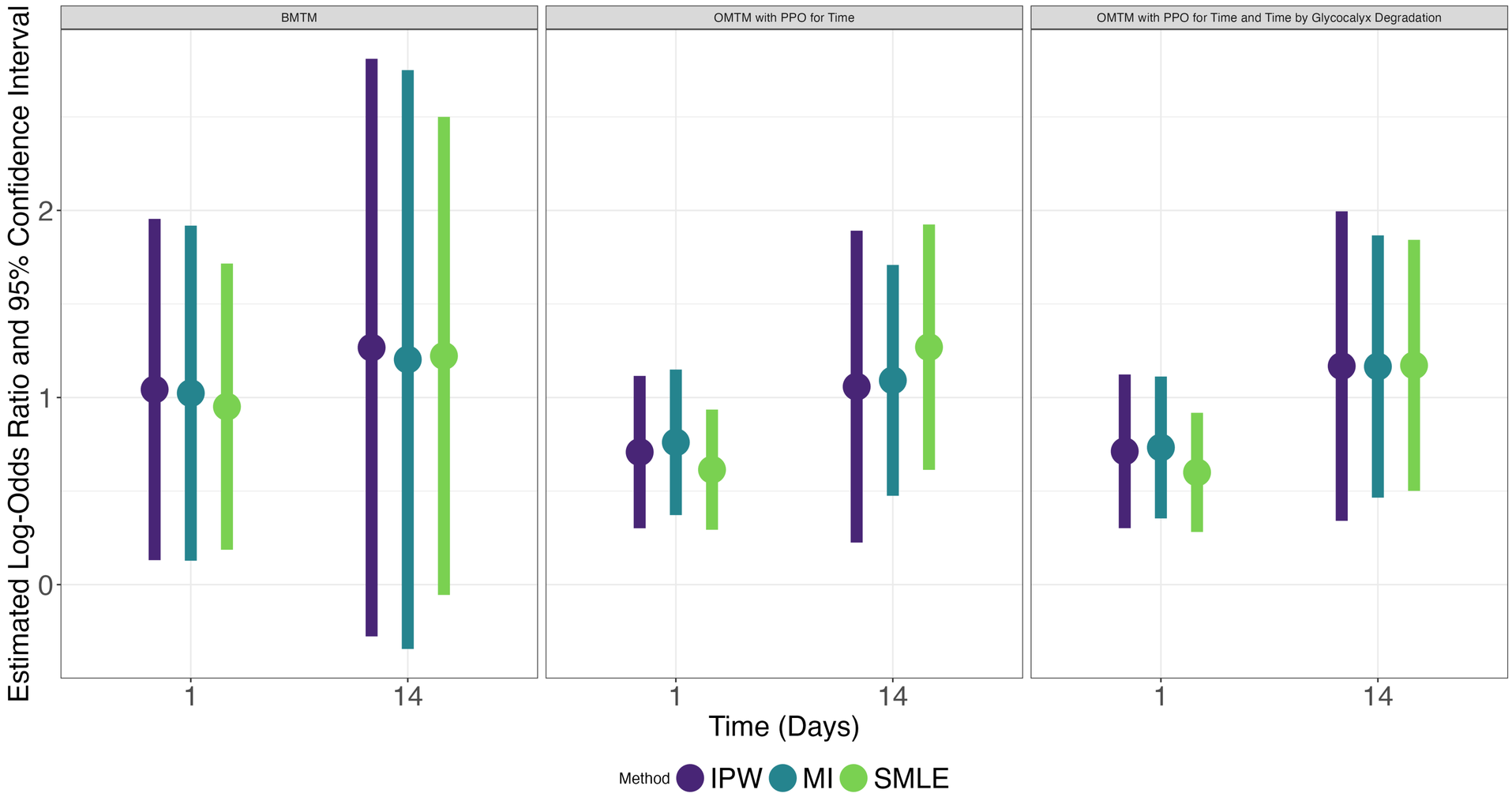
All models were adjusted for glycocalyx degradation, time, age, sex, ARDS, SOFA score and time by glycocalyx degradation

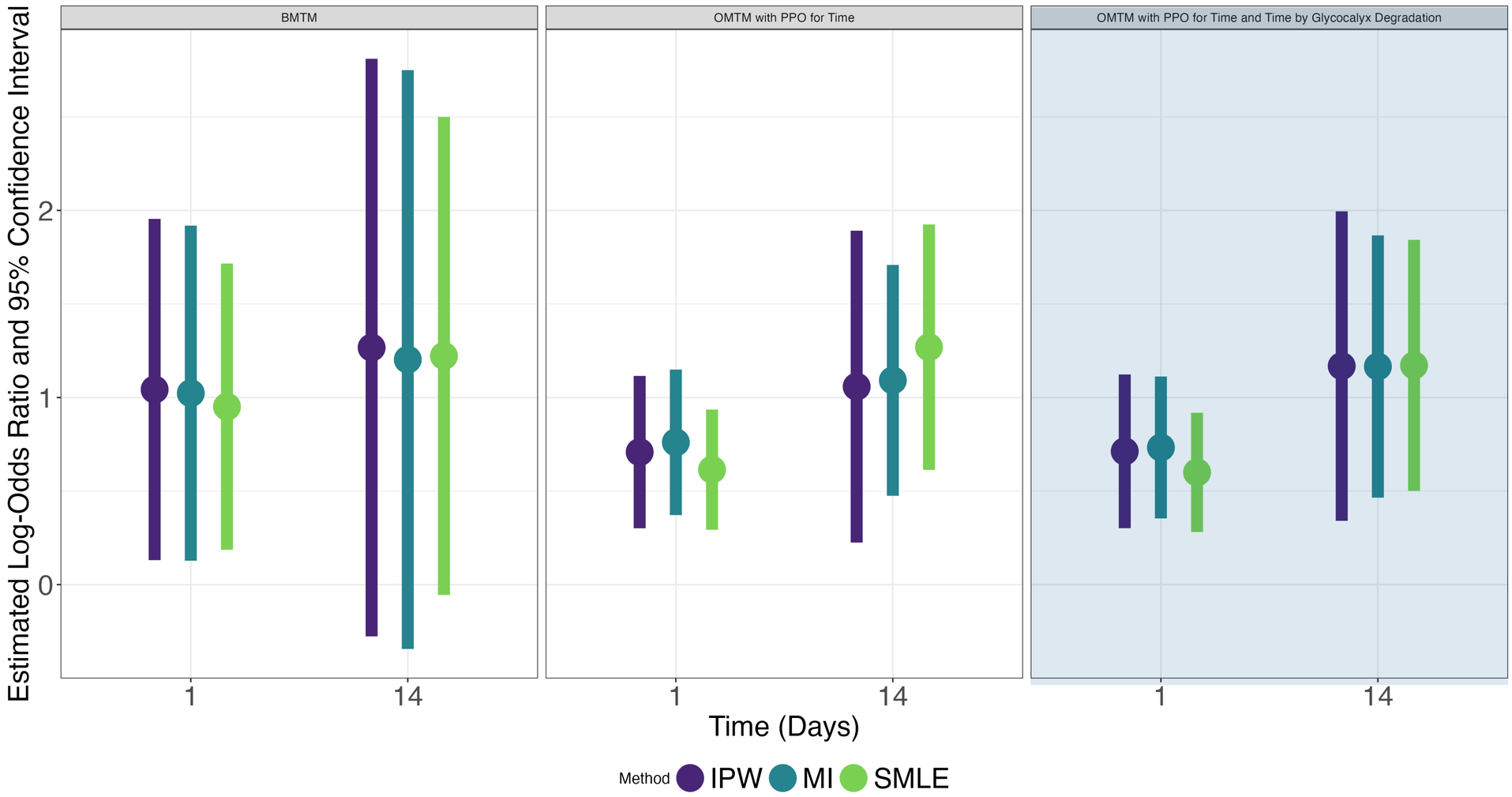

BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE

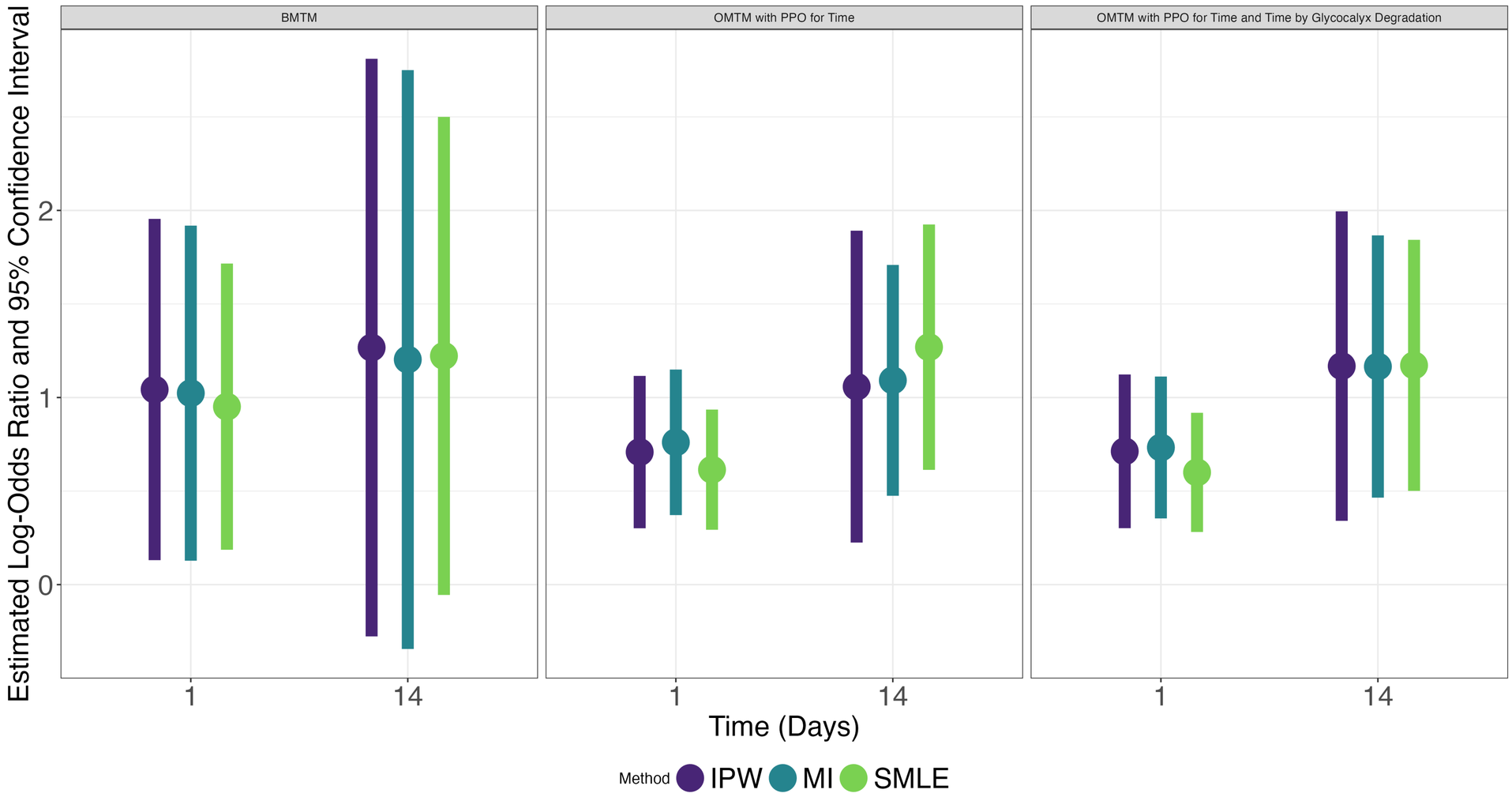

BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE


BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE

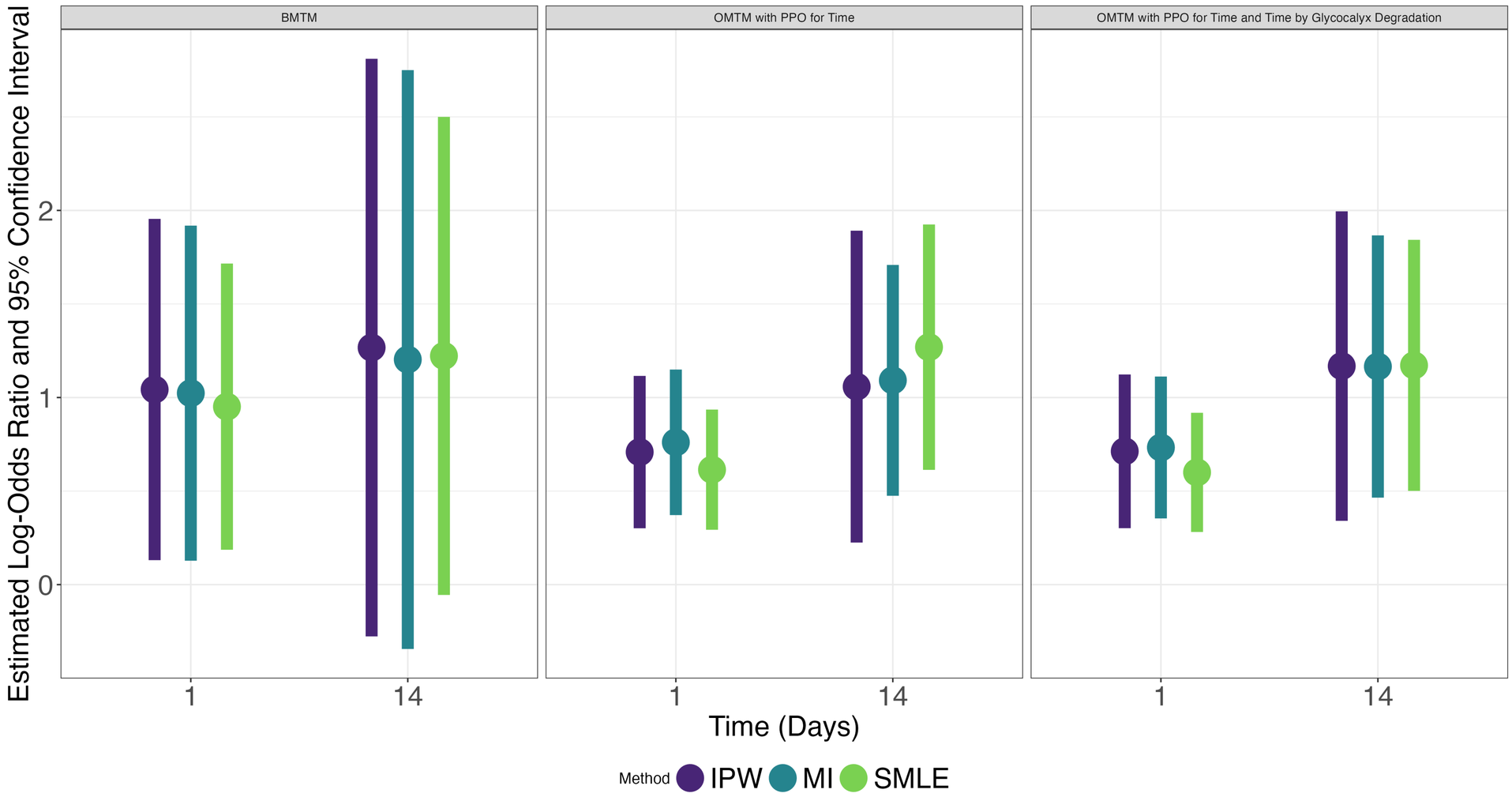

BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE

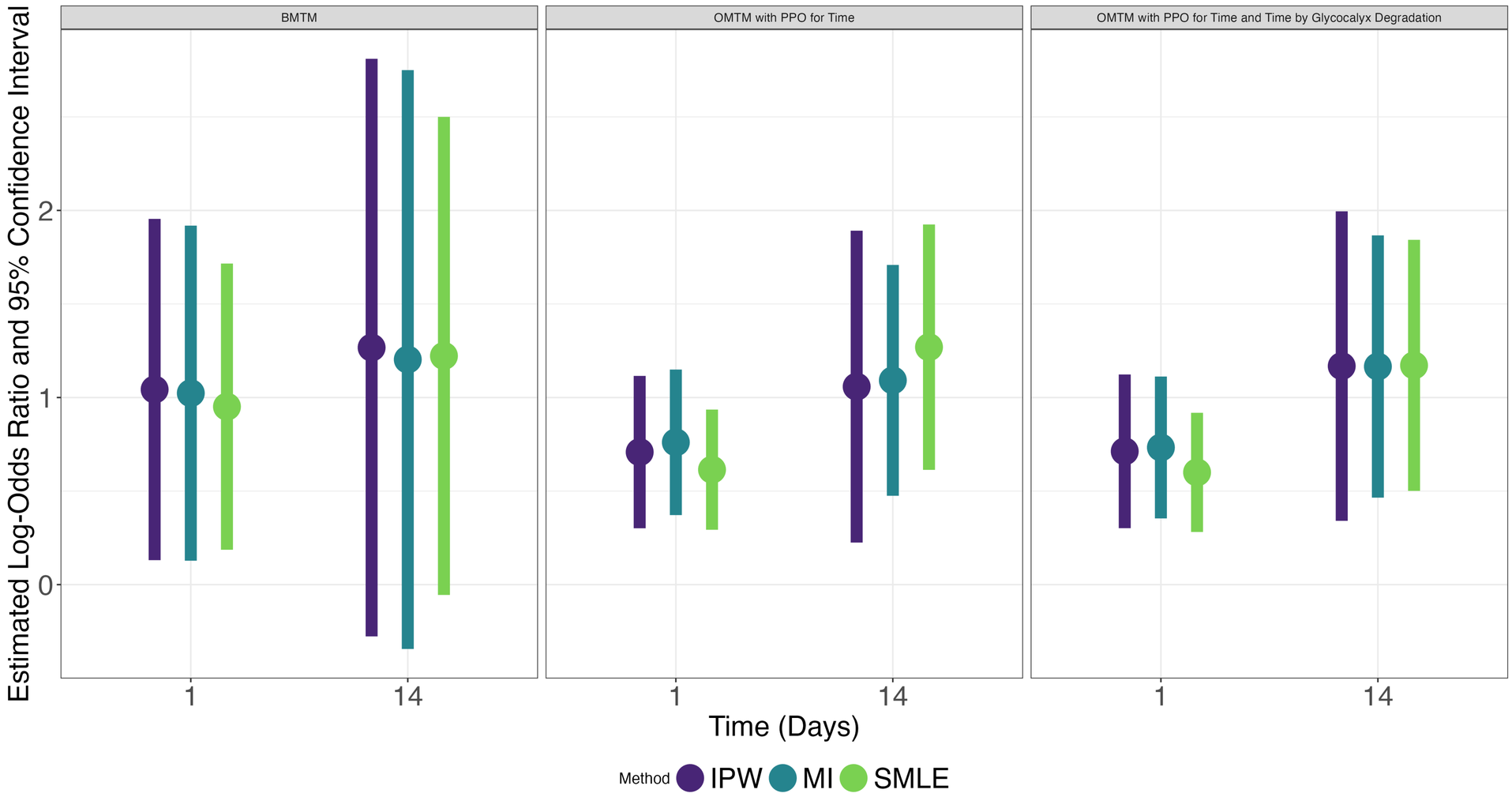

BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE

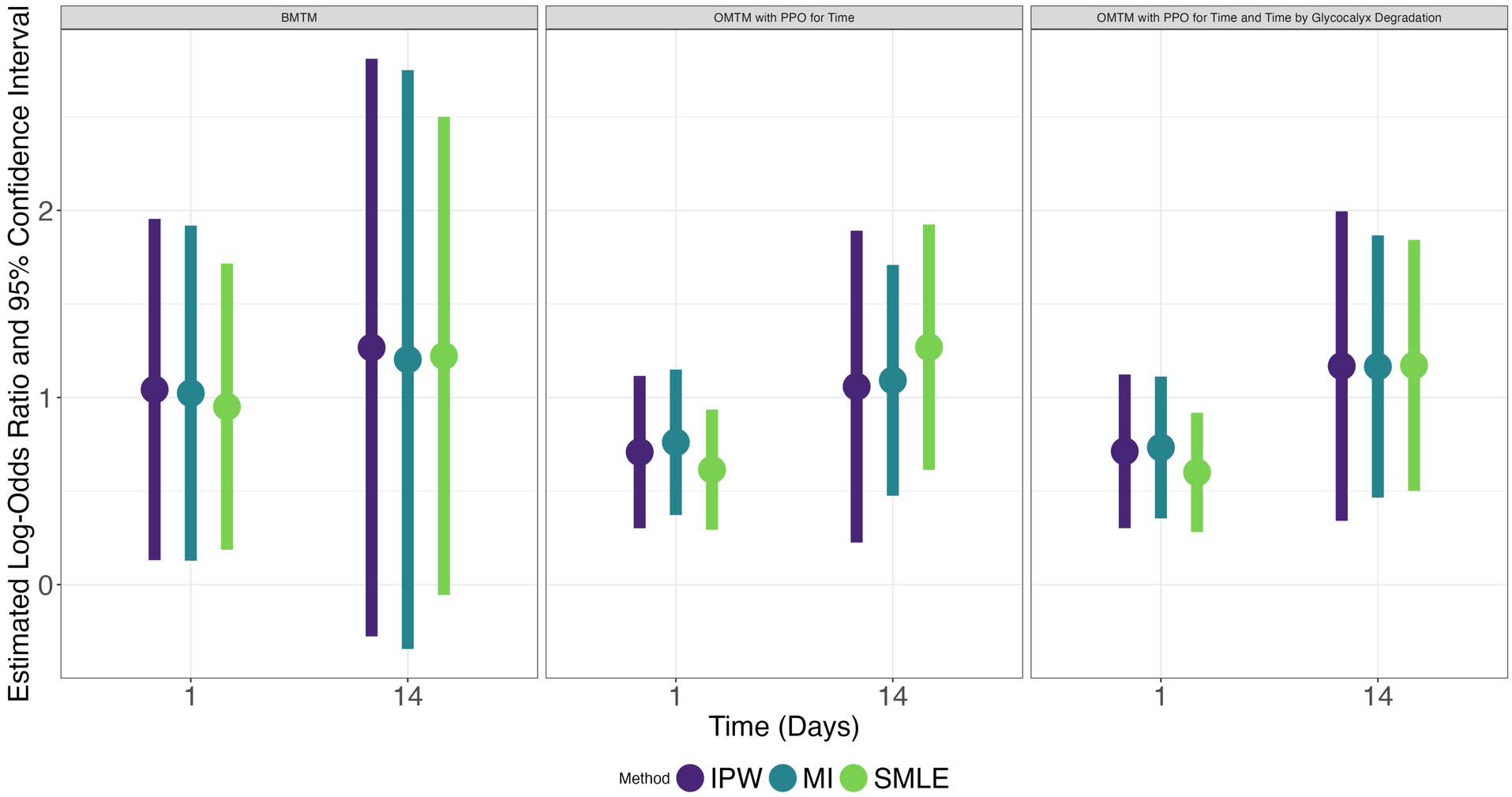

BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE


BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE

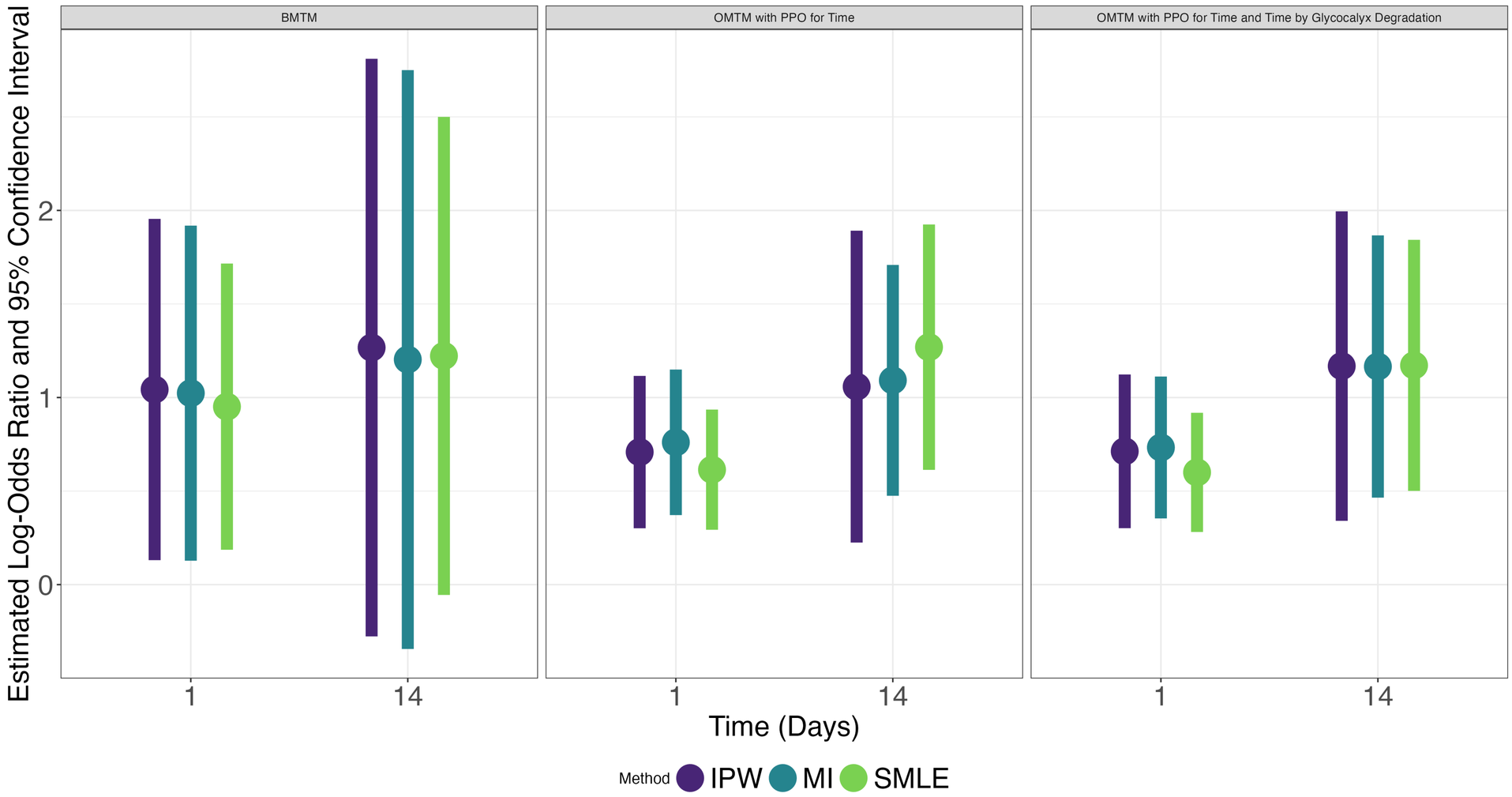
At day 1 people with glycocalyx degradation had higher odds of mortality

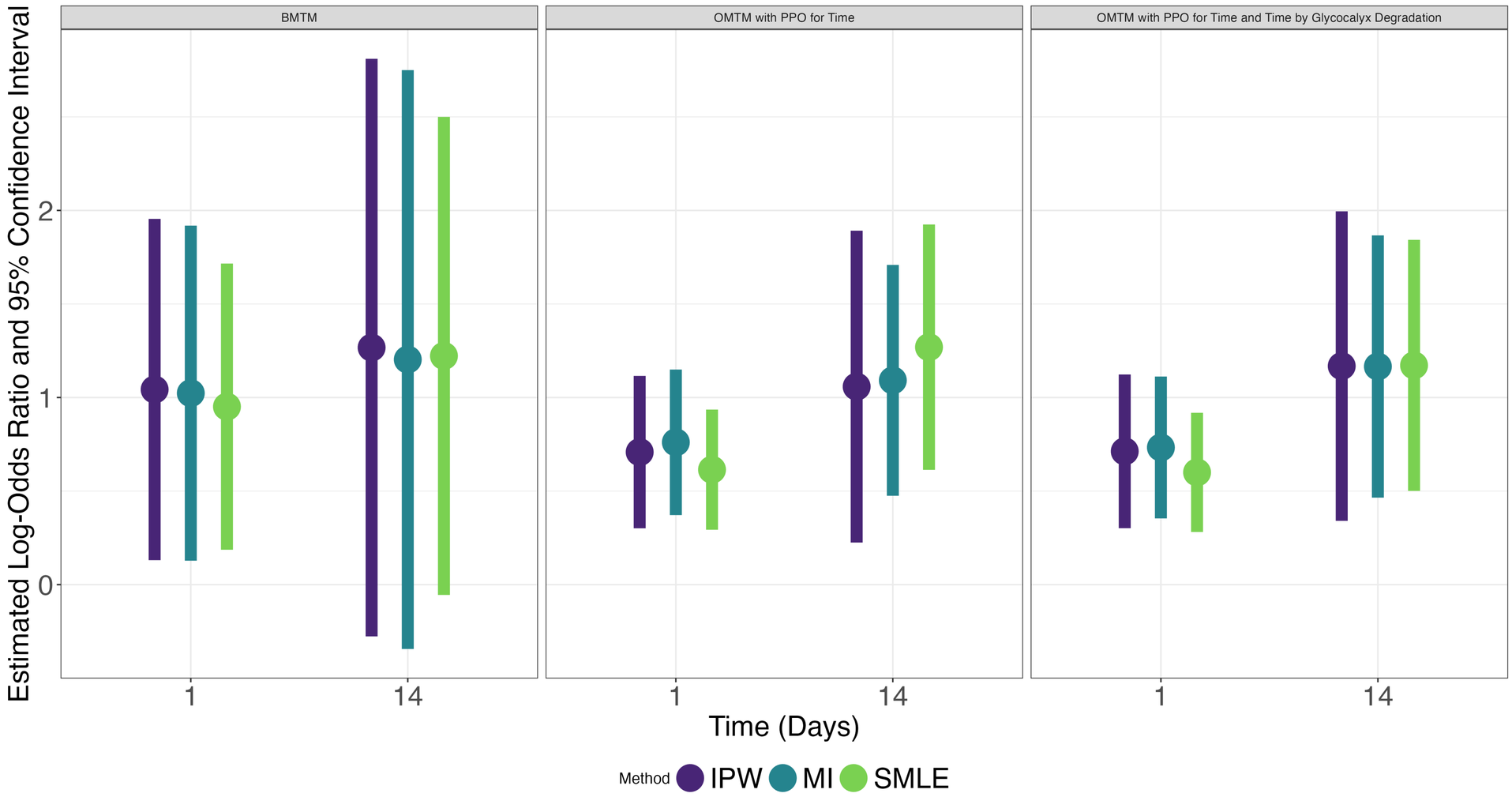



BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE

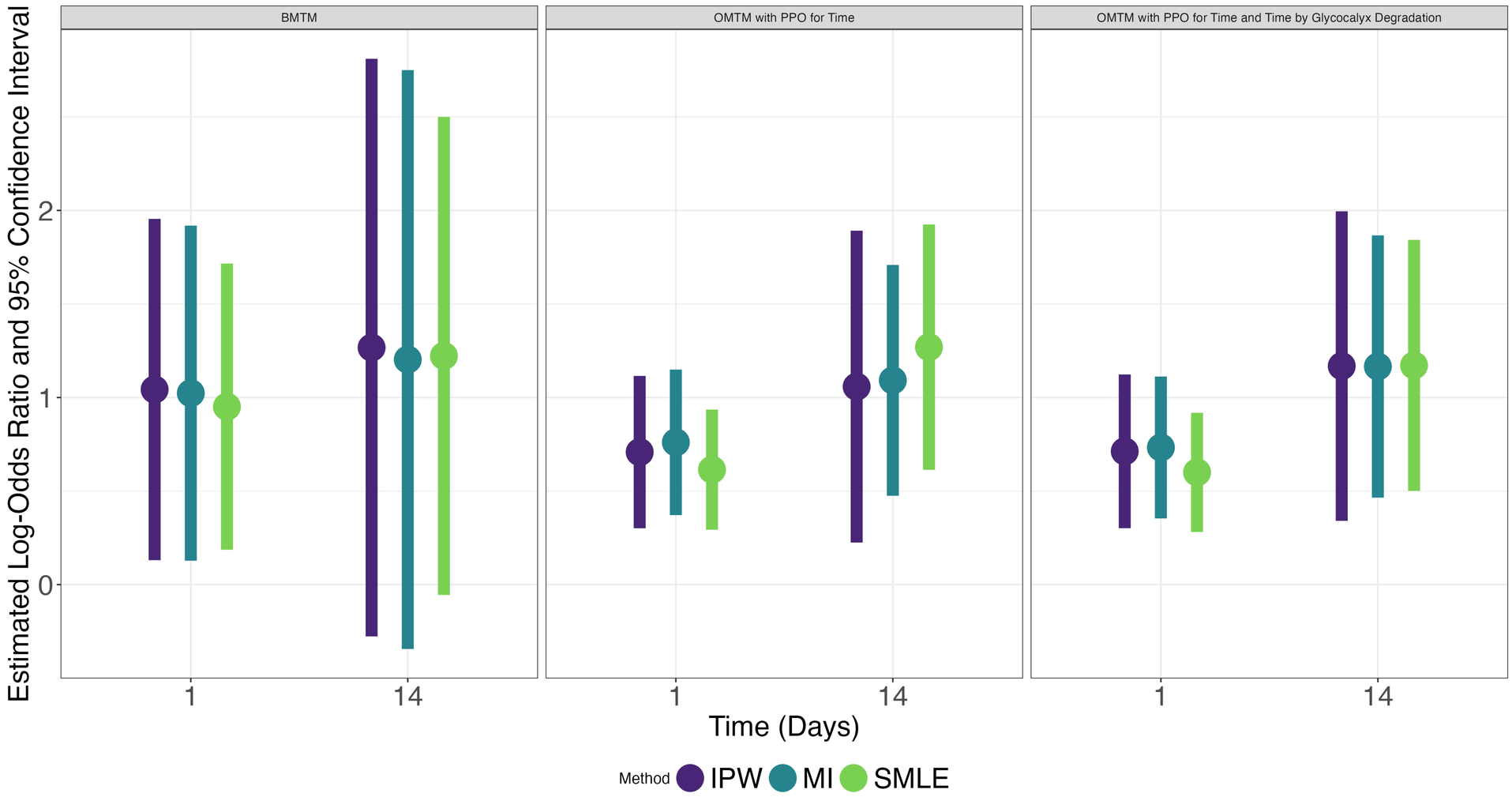


SMLE and MI reduced the width of the CI by 17% at day 14


BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE



The ordinal outcome reduced the width of the CI at day 1 and at day 14


BMTM
OMTM with PPO for Time
OMTM with PPO for Time and Time by Glycocalyx Degradation
Estimated Log-Odds Ratio and 95% Confidence Interval
2
1
0
1
14
1
14
1
14
Time (Days)
Method
IPW
MI
SMLE

Conclusion

Marginalized transition models with longitudinal ordinal or binary outcomes can be used to estimate the association between an exposure and mortality

Estimation efficiency can be increased when using ordinal outcome rather than a binary outcome

When all available information is included in the estimation procedure (SMLE or MI), we observed efficiency gains compared to methods that only include participants with complete data

Thank you!

Email: c.di-gravio@imperial.ac.uk

Reference

Schildcrout et al (2022). Model-assisted analyses of longitudinal, ordinal outcomes with absorbing states. Statistics in Medicine.

NHLBIP and Early Treatment of Acute Lung Injury Clinical Trial Network (2023). Early Restrictive or liberal fluid management for sepsis-induced hypotension. NEJM.

Di Gravio et al (2023+) Efficient Designs and Analysis of Two-Phase Studies with Longitudinal Binary Outcome. Biometrics

Appendix

Why Choosing a Marginalized Transition Model?

We want to model the marginal mean and the dependence **separately**

- Estimates are consistent as long as the marginal mean model is correctly specified

We want to be able to **directly** estimate a marginal effect

We want to have a likelihood-based method