

STA 440 Case 5: Stroke Outcomes

Yura Heo, Anirudh Jain, Richard Jiang

2025-12-04

Background and Motivation

Researchers developed a new stroke awareness program with the goal of reducing the time needed to provide patients with life-saving acute stroke care delivery. The investigation was divided into two time periods: the pre-treatment period, during which researchers developed the program, and the program implementation period, which overlapped with the onset of the COVID-19 pandemic. The goals of the investigation are twofold: to determine the factors associated with “good” stroke care outcomes, defined as a discharge to home or a rehabilitation facility, and to determine if the implementation of the program improved patient outcomes.

Data and Exploratory Data Analysis

The researchers tracked stroke patient data from 2019-2020 on a quarterly basis, recording a variety of covariate variables, patient demographic information, and the hospital site where treatment took place. The resulting data set had missing data: therefore, it was necessary for us to develop an approach to handling the missingness before creating any models. In the exploratory analysis phase, we sought to first determine if data missingness was missing at random, missing completely at random, or missing not at random. Our findings on data missingness were used to inform our approach to creating an appropriate model for patient outcomes. Because one goal of the investigation is to determine the effectiveness of the program implementation relative to pre-implementation, we divided the time into three periods: a baseline period (Y1Q1-Y1Q2), a middle transition period as the program was first implemented (Y1Q3-Y2Q2), and the final implementation period (Y2Q3-Y2Q4), where the effects of the program might be the most detectable.

To begin our analysis, we first sought to determine which variables were most frequently missing. Table 3 reveals that the 4 variables with missing data were, in descending order, patient age, whether the patient was transported to the hospital via car or EMS, if the hospital was notified prior to patient arrival, if they were transported by EMS, and patient race. To determine whether the data is missing completely at random, we investigated whether there were patterns to the missingness. Figures 2 and 3 are heatmaps visualizing the proportion of data missing from the variables. Figure 2 displays the proportion of missingness in each variable by our three time periods, while Figure 3 displays variable missingness by hospital site. Inspection of Figure 2 reveals that data missingness appears worse for certain time quarters, most severely with missing age in the first period. Figure 3 reveals that certain hospitals, namely site IDs 140, 150, and 170, had high variable missingness, with age being the most frequently missing. Because missingness, specifically in age, seems to

vary with observed factors like age and time period, it is reasonable to assume that the data is not missing completely at random. The data being missing not at random is unlikely but not impossible based on these graphs. We cannot, for example, conclusively claim that the probability of age being missing does not depend on the unobserved age. However, because there are sites that have age missing completely for all patients, not just selectively older/younger ones, we believe the most likely scenario is the data being missing at random.

Model Rationale and Implementation

Model Rationale

We factor the joint distribution as

$$\begin{aligned} p(\text{age}, \text{notify}, \text{transport}, \text{race}, \text{home} \mid \mathbf{z}) &= p(\text{home} \mid \text{age}, \text{notify}, \text{transport}, \text{race}, \mathbf{z}) \\ &\quad \times p(\text{age} \mid \text{notify}, \text{transport}, \mathbf{z}) \\ &\quad \times p(\text{notify} \mid \text{transport}, \mathbf{z}) \\ &\quad \times p(\text{transport} \mid \mathbf{z}) \\ &\quad \times p(\text{race} \mid \text{site}). \end{aligned}$$

with \mathbf{z} the fully observed covariates (site, time stage, gender). This ordering mirrors the care workflow (arrival/notification precedes discharge) and lets us impute missing Age/notify/transport inside the MCMC rather than as a preprocessing step. This is critical because missingness varies by site and time period, making the assumption of missing completely at random unrealistic. Time is collapsed into three stages: baseline (Y1Q1–Y1Q2), mid (Y1Q3–Y2Q2), final (Y2Q3–Y2Q4) in order to stabilize time effects and reflect the program rollout. Site-level random intercepts allow modest heterogeneity across centers. Fixed effects use weakly informative $\mathcal{N}(0, 10^4)$ priors; variance components use half-uniform priors; and site-specific race probabilities follow a Dirichlet($\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$) prior.

Model Implementation

1. **Outcome (home vs. other):** $\text{home}_i \sim \text{Bernoulli}(p_i)$

$$\begin{aligned} \text{logit}(p_i) &= \beta_0 + \beta_{\text{age}} \text{Age}_i + \beta_{\text{gender}} \text{Gender}_i + \beta_{\text{ems}} \text{Transport}_i + \beta_{\text{not}} \text{Notify}_i + \beta_{\text{tpa}} \text{TPA}_i \\ &\quad + \beta_{\text{thr}} \text{Thromb}_i + \beta_{\text{tpaC}} \text{TPAComp}_i + \beta_{\text{thrC}} \text{ThrComp}_i + \beta_{\text{time}}[t_i] + \beta_{\text{race}}[r_i] + u_{\text{site}[i]}^{(out)}. \end{aligned}$$

This model estimates associations between demographic and pre-hospital factors and the probability of a good discharge outcome. Time-stage effects are coded relative to the baseline period (Y1Q1–Y1Q2), and Caucasian is used as the reference race category.

2. **Age | transport, notify, site, time stage, gender**

$$\text{Age}_i \sim \mathcal{N}(\alpha_0 + \alpha_{\text{time}}[t_i] + \alpha_{\text{gender}} \text{Gender}_i + \alpha_{\text{tr}} \text{Transport}_i + \alpha_{\text{not}} \text{Notify}_i + u_{\text{site}[i]}^{(age)}, \sigma_{\text{age}}^2).$$

This lets age vary by site, time, and pre-hospital factors, reflecting real-world care patterns.

3. Notify | transport, site, time stage, gender

$$\text{Notify}_i \sim \text{Bernoulli}(\pi_i^{(not)}), \quad \text{logit}(\pi_i^{(not)}) = \delta_0 + \delta_{\text{tr}} \text{Transport}_i + \delta_{\text{time}}[t_i] + \delta_{\text{gender}} \text{Gender}_i + u_{\text{site}[i]}^{(not)}.$$

A close to 0 probability is assigned when $\text{Transport}[i]=0$, representing that notification is effectively unavailable for car arrivals.

4. Transport | site, time stage, gender

$$\text{Transport}_i \sim \text{Bernoulli}(\pi_i^{(tr)}), \quad \text{logit}(\pi_i^{(tr)}) = \gamma_0 + \gamma_{\text{time}}[t_i] + \gamma_{\text{gender}} \text{Gender}_i + u_{\text{site}[i]}^{(tr)}.$$

Transport patterns may vary by site and program stage.

5. Race | site

$$\text{Race}_i \sim \text{Categorical}(\pi_{\text{site}[i], \cdot}^{(race)}), \quad \pi_{s, \cdot}^{(race)} \sim \text{Dirichlet}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right).$$

This allows each site to have its own racial distribution, with a weakly informative Jeffreys prior.

Time-stage effects are deviations from the baseline stage (fixed to zero); site random effects are mean-zero Normals with their own scale parameters. We run 3 chains with 2,000 burn-in iterations and 8,000 post-burn-in draws.

Model Evaluation

We assessed convergence and fit using multiple diagnostics:

- **Trace plots and effective sample sizes:** Core parameters (age, EMS, TPA, complications, time-stage effects, site SDs) show well-mixed “fuzzy caterpillars” with effective sample sizes in the hundreds to thousands, indicating good mixing despite the hierarchical structure (see Figure 1 and Table 1).
- **Gelman–Rubin (R-hat):** All substantive parameters have PSRFs ≈ 1.00 – 1.02 ; race and time-stage effects mix well despite small subgroup sizes. Intercept and age show mild autocorrelation but remain below typical concern thresholds (upper CI ≈ 1.04); see Table 2.
- **Posterior predictive checks:** Residuals split by outcome behave as expected for binary data; Pearson residuals only spike when the model was highly confident and the outcome disagreed. The Bayesian calibration curve tracks the diagonal closely, with only slight under-prediction in mid–high probability bins, indicating well-calibrated probabilities overall (see Figure 4, Figure 5, Figure 7, Figure 8, and Table 4).

Results

The posterior summary Table (Table 5) and the (Table 6) indicate several strong associations. Each additional year of age is associated with roughly a 6–7% decrease in the odds or a good discharge outcome (OR = 0.94). Patients arriving via EMS had about 65% lower odds of being discharged home or to rehabilitation compared to those arriving by car (OR = 0.34) consistent with EMS patients having more severe strokes. Complications were among the strongest predictors of poor

outcomes: thrombectomy complications were associated with an approximately 89% reduction in the odds of a good discharge, and tPA complications with a 75% reduction. Race differences were also noticeable: African American patients had about 35% lower odds of good outcomes compared to Caucasian patients.

Time effects provided insight into whether outcomes improved over the study period. Compared with baseline, the mid-study period and final periods were associated with roughly 20-30% higher odds of a good discharge outcome. When translated into predicted probabilities for a reference patient (female, Caucasian, median age(70), car arrival, no complications, average site), the model estimates an absolute improvement of about 2-4 percentage points from baseline to the final study period. Taken together with the residual and calibration diagnostics, these findings suggest modest but consistent evidence that outcomes improved over time.

Conclusion

This study combined exploratory data analysis with a joint Bayesian modeling framework to identify predictors of good stroke discharge outcomes and to evaluate whether outcomes improved following implementation of a stroke awareness program. Exploratory analyses highlighted systematic missingness in age, transport, and notification variables across time and sites, motivating a joint modeling approach capable of imputing missing covariates coherently within the likelihood.

Posterior inference revealed several strong predictors of patient outcomes. Older age, EMS transport, and treatment complications substantially reduced a patient's likelihood of discharge to home or rehabilitation, while race disparities were also evident. Importantly, the estimated time-stage effects suggest that the program may have contributed to improved patient outcomes: for a reference patient, the predicted probability of a good outcome increased modestly, approximately 2-4 percentage points, from baseline to the final study period.

Overall, the joint Bayesian model provided a principled and flexible framework for handling missing data, characterizing associations, and assessing temporal improvements in outcomes.

Limitations and Future Directions

Although the joint model imputes missing values probabilistically, it still relies on the assumption that missingness is ignorable given the modeled covariates; if missingness depends on unobserved severity or workflow factors, residual bias may persist. In terms of site-level heterogeneity, substantial relies on the assumption that missingness is ignorable given the modeled covariates.

Future work could incorporate site-level predictors, patient severity measures, or dynamic time trends to more precisely quantify program effects. Extensions that explore interactions between demographic and clinical variables, or assess whether program benefits were equitably distributed across patient subgroups, would provide additional insight.

Appendix

Convergence diagnostics

Table 1: Effective sample sizes for core parameters.

parameter	ess
beta0	81.9
beta_age	115.5
beta_ems	334.9
beta_gender	4180.1
beta_not	839.3
beta_race[1]	0.0
beta_race[2]	3712.7
beta_race[3]	7226.8
beta_thr	1328.2
beta_thrC	10823.4
beta_time[1]	0.0
beta_time[2]	2073.3
beta_time[3]	2583.7
beta_tpa	1271.5
beta_tpaC	7564.6
sigma_age	4691.7
sigma_site_age	1795.9
sigma_site_not	4187.7
sigma_site_out	1008.2
sigma_site_tr	2022.1

Table 2: Gelman–Rubin PSRF for monitored parameters.

parameter	point_est	upper_ci
beta0	1.035	1.114
beta_age	1.025	1.084
beta_ems	1.002	1.006
beta_gender	1.002	1.005
beta_not	1.005	1.018
beta_race[1]	NaN	NaN
beta_race[2]	1.004	1.014
beta_race[3]	1.001	1.005
beta_thr	1.011	1.039
beta_thrC	1.001	1.003
beta_time[1]	NaN	NaN
beta_time[2]	1.001	1.001
beta_time[3]	1.001	1.001
beta_tpa	1.010	1.034

Table 2: Gelman–Rubin PSRF for monitored parameters.

parameter	point_est	upper_ci
beta_tpaC	1.002	1.007
sigma_age	1.006	1.020
sigma_site_age	1.005	1.011
sigma_site_not	1.000	1.001
sigma_site_out	1.007	1.021
sigma_site_tr	1.003	1.004

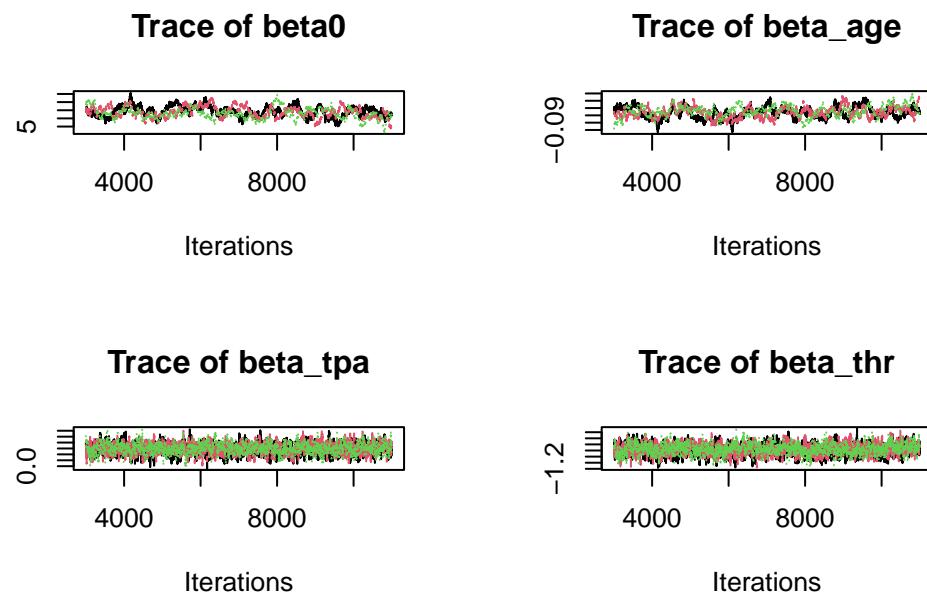


Figure 1: Trace plots for selected parameters.

Residual diagnostics

Table 3: Missingness Per Variable

```
# A tibble: 17 x 2
  variable      n_missing
  <chr>          <int>
1 age              802
2 pre_hosp_notify    203
3 notify_num        203
4 emsvscar           94
5 transport_num       94
6 race2              83
7 home_or_rehab       58
8 home_num             58
9 site_id                 0
10 time2                  0
11 gender                  0
12 had_thrombectomy        0
13 had_tpa                  0
14 tpa_complic                0
15 thr_complic                0
16 gender_num                 0
17 time_stage                 0
```

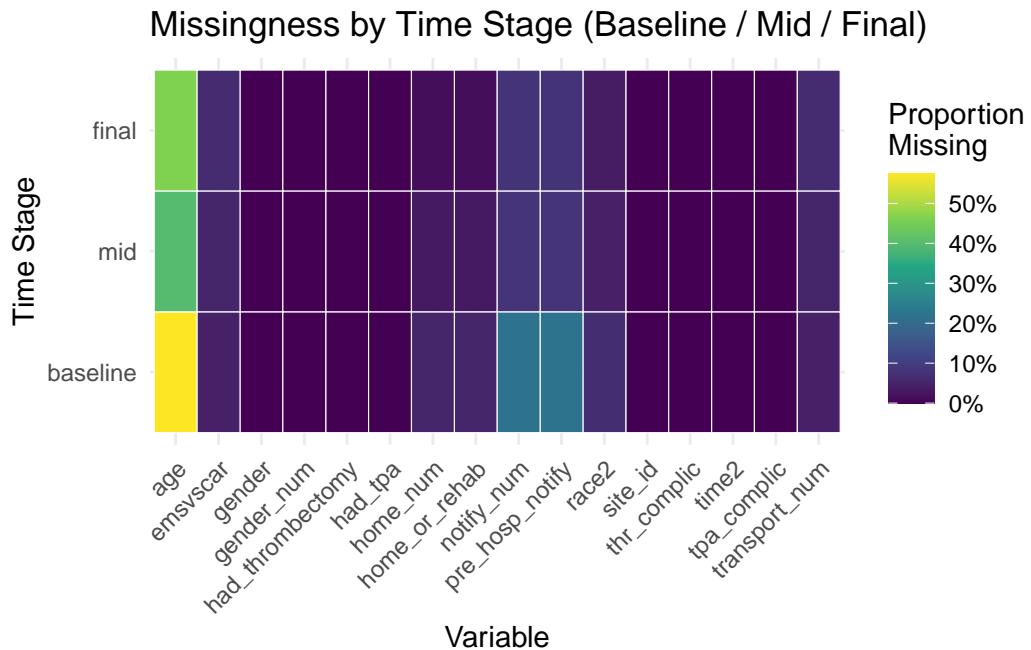


Figure 2: Missingness by Time Stage

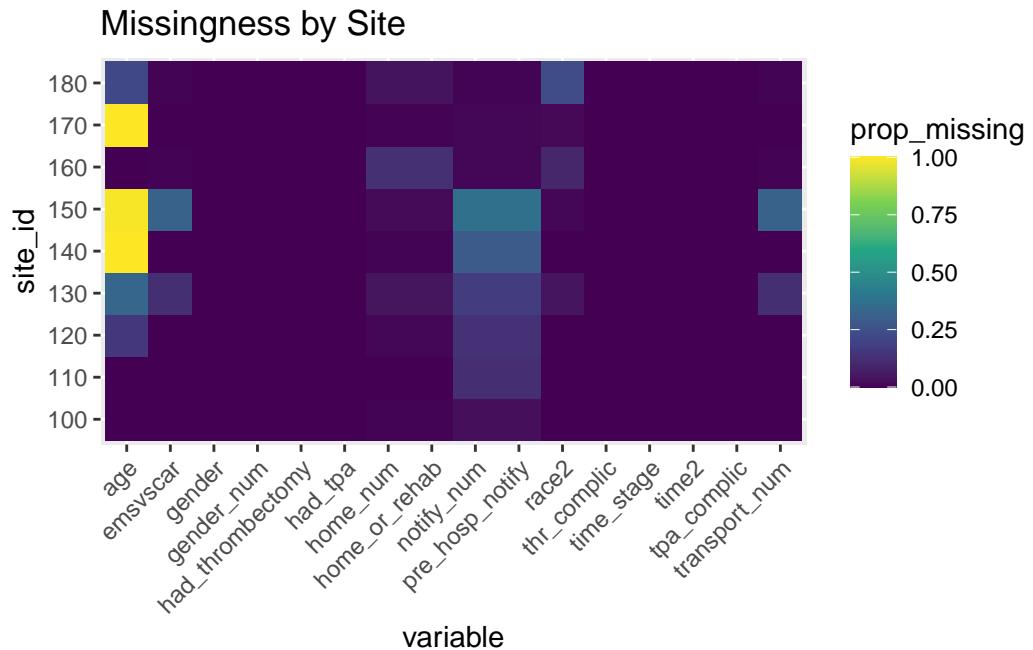


Figure 3: Missingness Aggregated by Site

Table 4: Summary of posterior residuals.

mean_resid	sd_resid	extreme
0.004	0.413	0.249

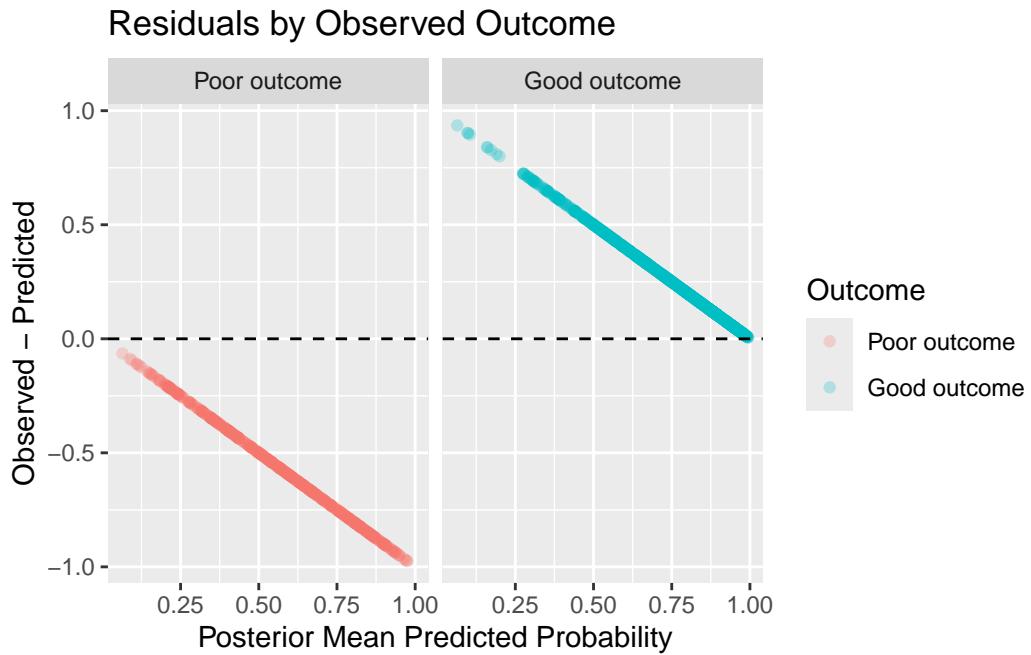


Figure 4: Residuals by observed outcome.

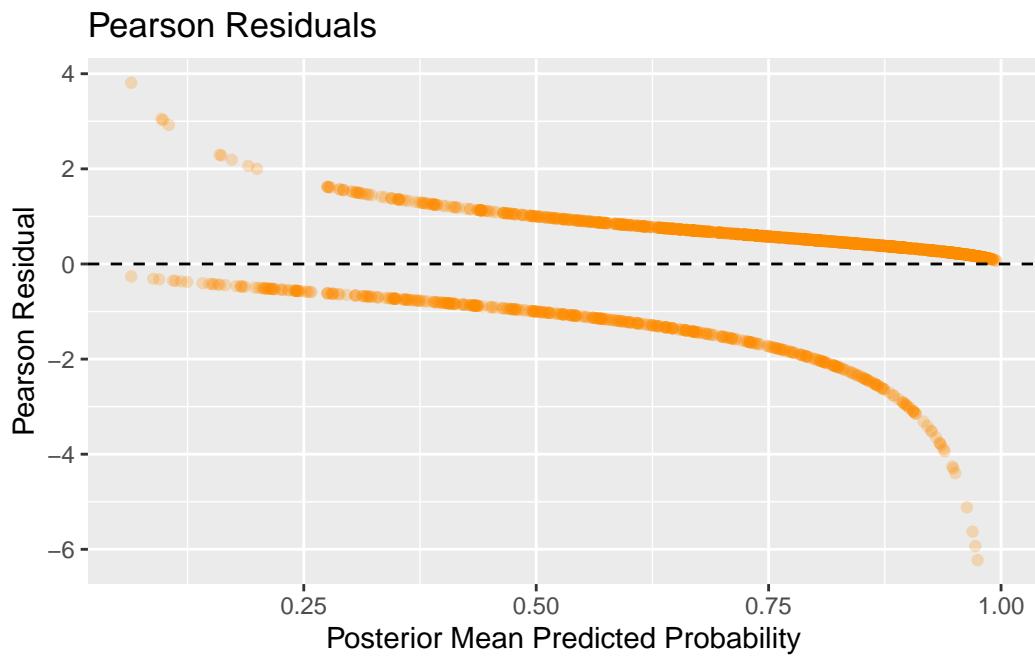


Figure 5: Pearson residuals vs. posterior mean predicted probability.

Distribution of Posterior Residuals

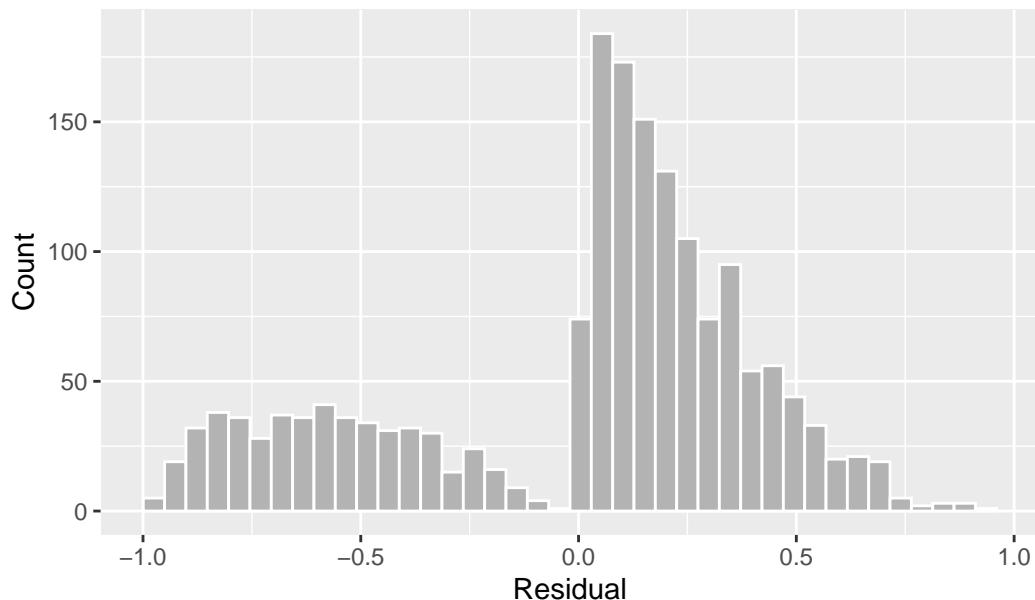


Figure 6: Distribution of posterior residuals.

Bayesian Calibration Curve

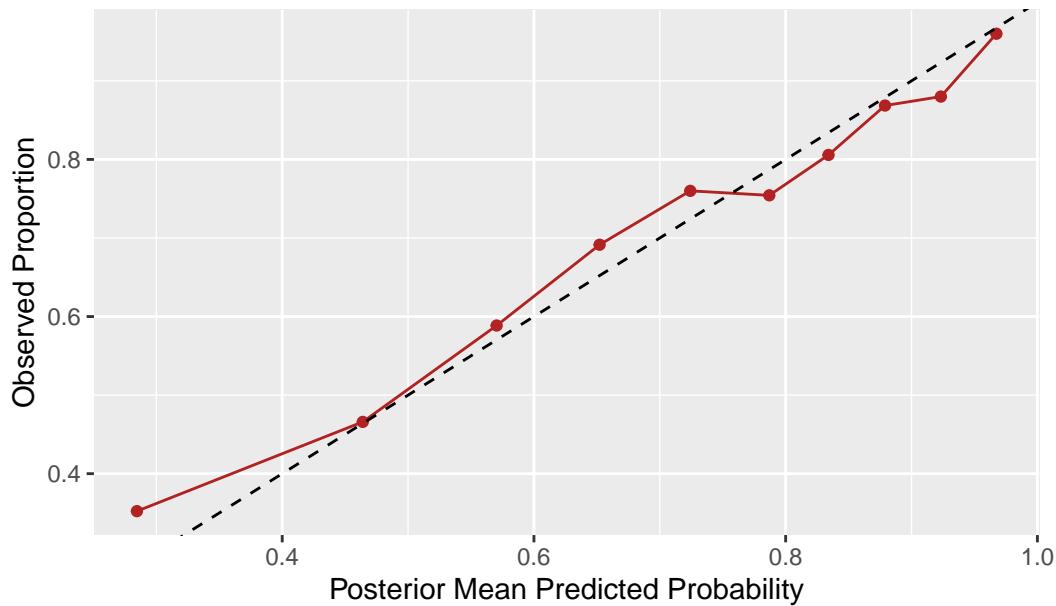


Figure 7: Bayesian calibration curve (deciles of predicted probability).

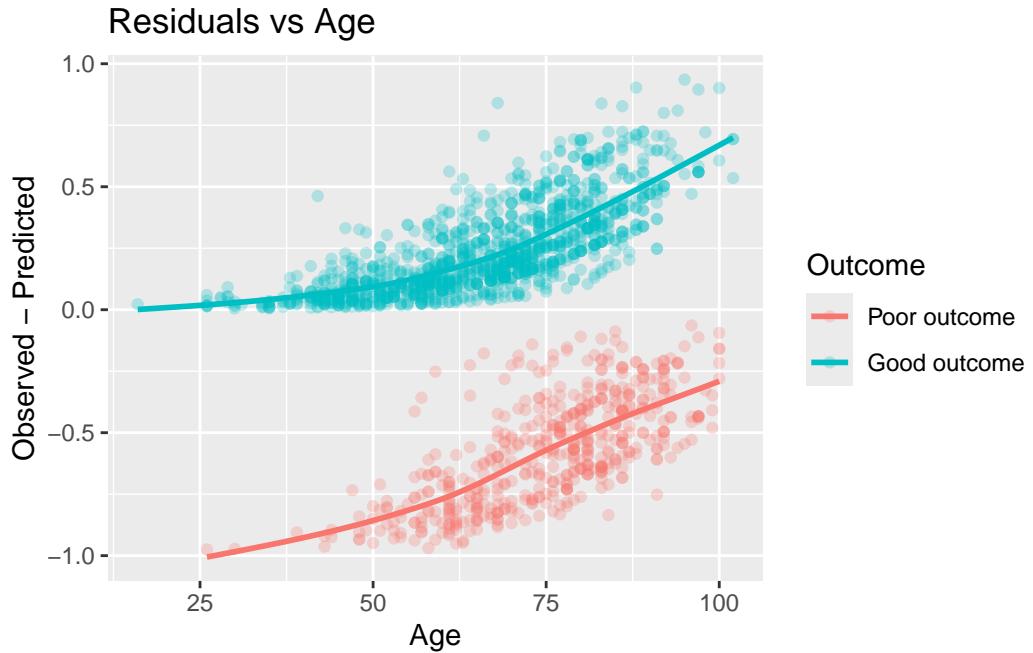


Figure 8: Residuals vs. age by observed outcome.

Posterior tables

Table 5: Posterior means and 95% credible intervals.

parameter	Mean	SD	2.5%	50%	97.5%
beta_age	-0.066	0.008	-0.080	-0.066	-0.051
beta_ems	-1.076	0.343	-1.770	-1.068	-0.429
beta_not	-0.007	0.227	-0.440	-0.012	0.437
beta_race[1]	0.000	0.000	0.000	0.000	0.000
beta_race[2]	-0.422	0.155	-0.728	-0.420	-0.121
beta_race[3]	-0.396	0.293	-0.958	-0.399	0.185
beta_thr	-0.564	0.173	-0.902	-0.564	-0.218
beta_thrC	-1.512	0.419	-2.360	-1.506	-0.714
beta_time[1]	0.000	0.000	0.000	0.000	0.000
beta_time[2]	0.184	0.168	-0.151	0.185	0.506
beta_time[3]	0.275	0.191	-0.097	0.276	0.654
beta_tpa	0.595	0.179	0.243	0.596	0.946
beta_tpaC	-1.435	0.401	-2.216	-1.435	-0.650
sigma_site_out	0.198	0.128	0.016	0.179	0.499

Table 6: Posterior odds ratios for key covariates.

parameter	OR_mean	OR_low	OR_high
beta_age	0.936	0.923	0.950
beta_ems	0.341	0.170	0.651
beta_not	0.993	0.644	1.548
beta_race[1]	1.000	1.000	1.000
beta_race[2]	0.656	0.483	0.886
beta_race[3]	0.673	0.384	1.203
beta_thr	0.569	0.406	0.804
beta_thrC	0.220	0.094	0.489
beta_time[1]	1.000	1.000	1.000
beta_time[2]	1.201	0.860	1.658
beta_time[3]	1.317	0.907	1.923
beta_tpa	1.813	1.275	2.577
beta_tpaC	0.238	0.109	0.522