Logistic Mixed Effect Model
Hands-on

By: Dr Hazlienor Mohd Hatta
drhazlienor@hotmail.com

1.0DATA ANALYSIS

1.1 Background

A researcher conducted a study across 7 different hospitals in a state involving 1250 patients to investigate factors that predict mortality among patients who have experienced a myocardial infarction (MI). Given that hospitals vary in terms of resources, patient demographics, and treatment protocols, the baseline risk of post-MI mortality is different for each hospital.

1.2 Method of Statistical Analysis

To analyse the factors associated with post-myocardial infarction (post-MI) mortality among patients, we will employ a multilevel logistic regression model. This approach is particularly suited for our dataset since patients are nested within hospitals, thus potentially introducing intra-hospital correlation that standard binary logistic regression might not account for.

Data analysis is being conducted using R Studio IDE for R software.

1.2.1 Data Preparation

- **Dataset:** mi_mortality.xlsx
- **Outcome Variable:** Post-MI mortality (binary: 0 = alive, 1 = dead).
- **Predictor Variables:**
    - **Age:** Continuous variable.
    - **Gender:** Categorical variable (binary: female, male).
    - **Underlying Diabetes Mellitus (DM):** Categorical variable (binary: no, yes).
    - **Chronic Kidney Disease (CKD):** Categorical variable (binary: no, yes).
    - **Treatment:** Categorical variable (binary: noninvasive, invasive)
    - **Duration of admission:** Continuous variable (in days).
    - **Hospital:** Categorical variable indicating the hospital where the patient was treated (1-7). Hospital 1 has no specialist, hospital 2-3 have general medicine specialist, and hospital 4-7 have cardiac specialist.

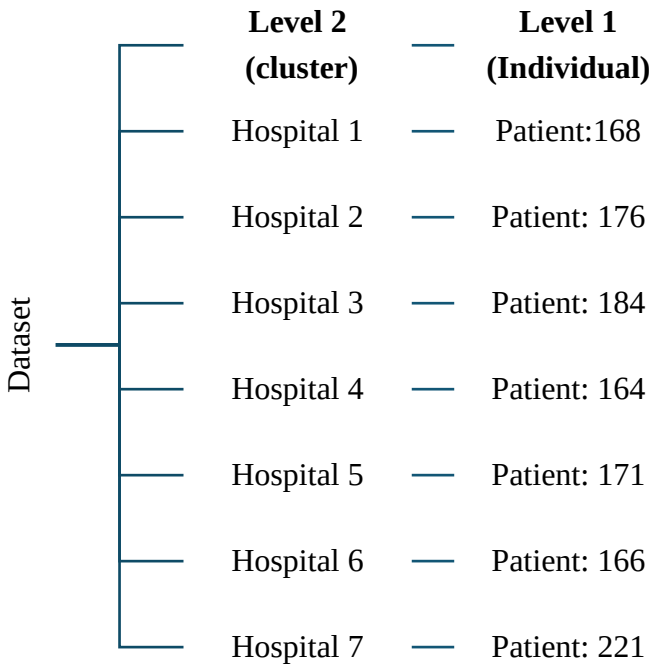


Figure 1.1 Hierarchical structure of the mi_mortality data

1.2.2 Data exploration and univariate analysis

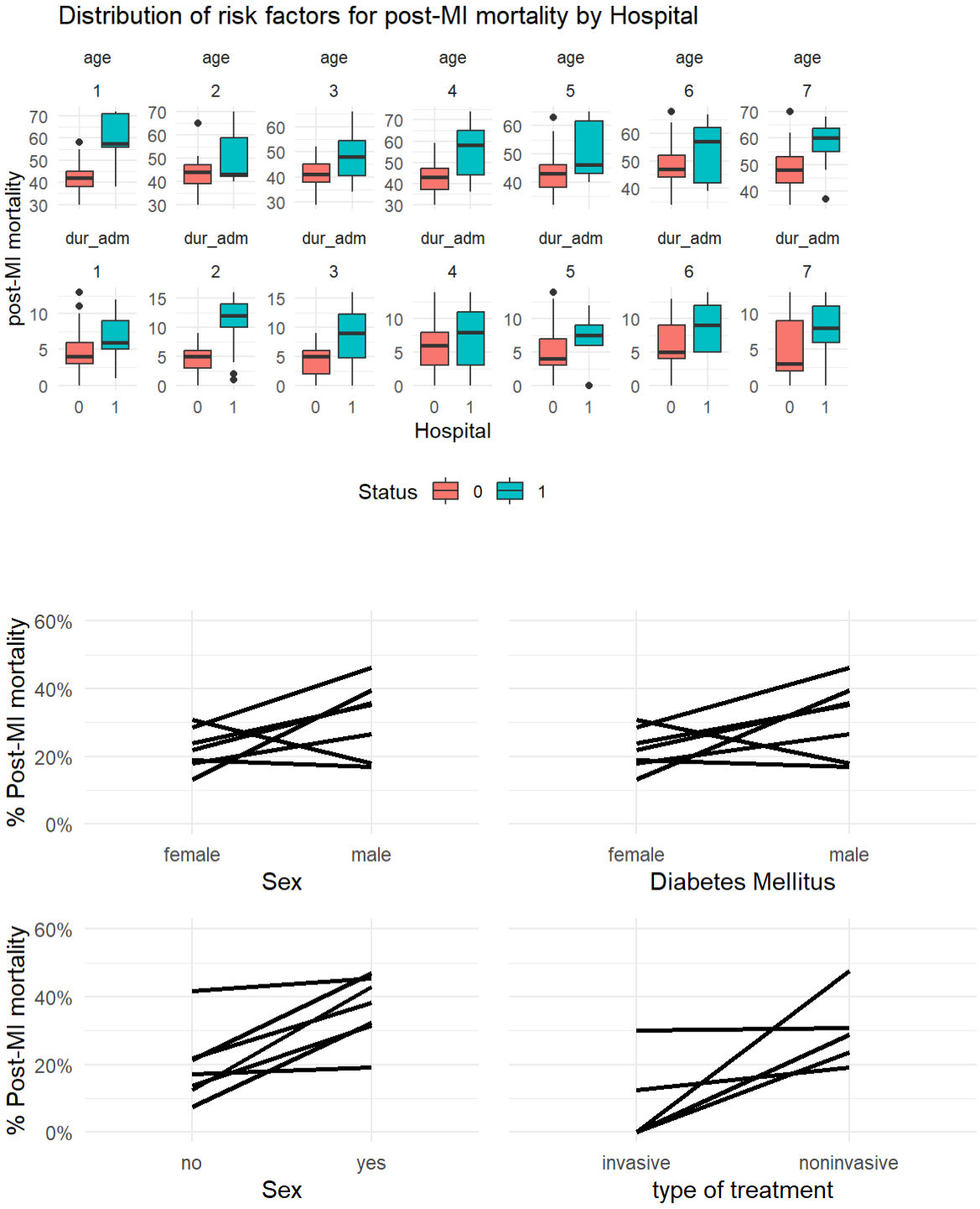


Figure 1.2. Distribution of post-MI mortality based on different risk factors by different hospital

In examining the percentage of post-MI mortality based on various factors across different hospitals, we observed notable variability in both intercept by hospital and slope by different factors. This variability suggests that the probability of post-MI mortality varies not only between hospitals but also in the way these factors influence mortality outcomes within each

hospital setting. We further examine the relationship between each independent variable and the outcome variable (post-MI mortality) individually through univariate analysis.

1.2.3 Model specification and estimation

Given the two-level structure of the data (patients (level 1) nested within hospitals (level 2)) with likely hospital-level variability, we will use a multilevel logistic regression model. The model will include both fixed effects for patient-level predictors and random effects to account for hospital-level variations.

Binary logistic regression model

We conducted binary logistic regression analysis using **glm** function.

Marginal (population average) model

We conducted the Generalized Estimating Equation (GEE) analysis using **gee** function from R package **gee** based on the following model:

$$\begin{aligned}\text{logit}(\text{P}(Y_{ij}=1)) &= \beta_0+\beta_1 Age_{ij}+\beta_2 Gender_{ij}+\beta_3 DM_{ij}+\beta_4 CKD_{ij}+\beta_5 treatment_{ij}\\ &+\beta_6 dur\_adm_{ij}\end{aligned}$$

Conditional model

The analyses were conducted using **glmer** function in R package **lme4** using maximum likelihood method.

Random intercept model

$$\begin{aligned}\text{logit}(\text{P}(Y_{ij}=1)) &= \beta_0+\beta_1 Age_{ij}+\beta_2 Gender_{ij}+\beta_3 DM_{ij}+\beta_4 CKD_{ij}+\beta_5 treatment_{ij}\\ &+\beta_6 dur\_adm_{ij}+u_{0j}\end{aligned}$$

$\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\beta_5,\beta_6$ are fixed effects representing the average effects of predictors on post-MI mortality across all hospitals.
u_{0j} is the random intercept for hospital *j*, capturing hospital-level baseline mortality differences.

Random intercept + single random slope model

$$\begin{aligned}\text{logit}\left(\text{P}(Y_{ij}=1)\right) &= \beta_0+\beta_1 Age_{ij}+\beta_2 Gender_{ij}+\beta_3 DM_{ij}+\beta_4 CKD_{ij}+\beta_5 treatment_{ij}\\ &+\beta_6 dur\_adm_{ij}+u_{0j}+u_6 dur\_adm_{ij}\end{aligned}$$

u_{6j} is random slope for duration of admission respectively for hospital *j*

Random intercept + multiple random slope model

$$\text{logit}\left(\text{P}(Y_{ij}=1)\right) = \beta_0 + \beta_1 Age_{ij} + \beta_2 Gender_{ij} + \beta_3 DM_{ij} + \beta_4 CKD_{ij} + \beta_5 treatment_{ij} + \beta_6 dur\_adm_{ij} + u_{0j} + u_5 treatment_{ij} + u_6 dur\_adm_{ij}$$

u_{5j}, u_{6j} are random slopes for treatment and duration of admission respectively for hospital j

Model comparison

We use **anova** function and compare the AIC of binary logistic regression to random intercept model, random intercept with single random slope model, and random intercept with multiple random slopes model. The AIC of random effect model was lower than the binary logistic regression model, hence inclusion of random effect is justified. The AIC of random intercept with multiple random slope model was significantly smaller than other random effect model, hence it is selected as the preliminary final model.

1.2.4 Inference

Confidence Intervals: 95% confidence intervals for fixed effects were calculated to provide a range of plausible values for the population parameters using **tidy** function from **broom.mixed** package

1.2.5 Prediction

Predicted log odds and probabilities of post-MI mortality were generated for individual patients based on their characteristics and hospital-level effects using **augment** function from **broom.mixed** package and **predict** function (type 'link' = log odds, 'response' = probability).

1.2.5 Model fitness assessment

Outlier assessment was conducted using **pairscore.fnc** of **languageR** package. Fitted values against the residuals plot was generated. AUROC was calculated using **pROC** package. Model performance was calculated using **confusionMatrix** function from **caret** package.

2.0 RESULTS

Table 2.1. The association between different risk factors and pot-MI mortality outcomes (*n*=1250)

Characteristic	Overall *n* = 1250[1]	Alive[1] *n*=902	Dead[1] n=348	p-value[2]
Hospital				<0.001
1	168 (13%)	138 (15%)	30 (9%)	
2	176 (14%)	122 (14%)	54 (16%)	
3	184 (15%)	104 (12%)	80 (23%)	
4	164 (13%)	128 (14%)	36 (10%)	
5	171 (14%)	129 (14%)	42 (12%)	
6	166 (13%)	124 (14%)	42 (12%)	
7	221 (18%)	157 (17%)	64 (18%)	
Age, years	46.6 (9.4)	44.2 (7.6)	53.0 (10.7)	<0.001
Sex				<0.001
Female	480 (38%)	378 (42%)	102 (29%)	
Male	770 (62%)	524 (58%)	246 (71%)	
Treatment				<0.001
Invasive	125 (10%)	109 (12%)	16 (4.6%)	
Non-invasive	1,125 (90%)	793 (88%)	332 (95%)	
Diabetes mellitus	702 (56%)	440 (49%)	262 (75%)	<0.001
Hypertension	572 (46%)	406 (45%)	166 (48%)	0.400
Chronic kidney disease	576 (46%)	365 (40%)	211 (61%)	<0.001
Duration of admission, days	6.0 (4.0)	5.2 (3.4)	8.2 (4.4)	<0.001

[1] n (%); Mean (SD)
[2] Pearson's Chi-squared test (categorical), Independent *t*-test

The analysis of post-myocardial infarction (post-MI) mortality revealed significant differences across hospitals (p <0.001) with varying mortality rates, ranging from 9% to 23%. Additionally, age, sex, underlying diabetes mellitus, chronic kidney disease, type of treatment, duration of admission, and hospital admitted were significantly associated with post-MI mortality.

We fitted a logistic mixed effect model (estimated using ML and Nelder-Mead optimizer) to predict post-MI mortality with age, sex, underlying diabetes mellitus, chronic kidney disease, type of treatment, and duration of admission. The model included duration of admission, type of treatment, and admitted hospital as random effect. The model's total explanatory power is substantial (conditional R2 =0.67) and the part related to the fixed effects alone (marginal R2) is of 0.56.

The effect of the studied risk factors on post-MI mortality while considering the random effect of hospital, duration of admission, and type of treatment is depicted in **Table 2.2-2.3**, and **Figure 2.1-2.3**, and is summarized by the following model:

$$\begin{aligned}\text{logit}\Big(\text{P}\big(Post-MI\ mortality_{ij}=1\big)\Big) &= -14.273+u_{0hospital}+0.17*Age+1.18\\ &*Gender(male=1,female=0)_{ij}+1.09*\text{DM}(yes=1,no=0)+1.22\\ &*CKD(yes=1,no=0)+1.22*treatment(noninvasive=1,invasive=0)\\ &+u_5 dur_{adm(hospital)}+u_6 treatment(noninvasive=1,invasive=0)_{(hospital)}\end{aligned}$$

For the estimate value of $u_{0hospital}$, $u_5 dur_{adm(hospital)}$, and $u_6 treatment(noninvasive = 1, invasive = 0)_{(hospital)}$, please refer **Table 2.3**.

Table 2.2. Multilevel Logistic Regression Results for Post-MI Mortality Risk Factors with Random Intercept (Hospital) and Random Slope (Age), *n*=1250

Characteristic	Unadjusted estimate		Adjusted estimate			
	log(OR)[1]	OR[1]	log(OR)[1]	OR[1]	95% CI[1]	p-value
Age, years	0.106	1.112	0.168	1.183	1.155, 1.211	<0.001
Sex						
Female	—	—	—	—	—	
Male	0.554	1.74	1.179	3.252	2.156, 4.905	<0.001
Diabetes Mellitus						
No	—	—	—	—	—	
Yes	1.163	3.199	1.094	2.987	2.032, 4.389	<0.001
Chronic Kidney Disease						
No	—	—	—	—	—	
Yes	0.818	2.266	1.219	3.384	2.333, 4.910	<0.001
Treatment			0.255	1.291	1.170, 1.425	<0.001
Invasive	—	—	—	—	—	
Non-invasive	1.048	2.852	1.453	4.278	0.395, 46.367	0.232
Duration of admission, days	0.204	1.226	0.255	1.291	1.170, 1.425	<0.001

Intercept for multilevel logistic regression -14.273, [1] OR = Odds Ratio, CI = Confidence Interval

Random Effects	
σ2	3.29
τ00 hospital	4.60
τ11 hospital.dur_adm	0.01
τ11 hospital.treatmentnoninvasive	5.89
ρ01	-0.44
	-1.00
ICC	0.25
N hospital	7
Observations	1250
Marginal R^2 / Conditional R^2	0.555 / 0.668

Table 2.3. Summary of random effect generated from mixed effect logistic regression analysis - random intercept for different hospitals and random slope (duration of admission, treatment)

hospital	Intercept (u_{0j})	duration of admission (u_{5ij})	treatment noninvasive(u_{6ij})
1	2.0238366	-0.05447862	-2.2774811
2	0.5997651	0.13069308	-0.8704736
3	-3.7831399	0.20408397	4.1211749
4	0.5531807	-0.06081919	-0.5613605
5	-1.4110052	-0.07807459	1.7423918
6	1.6583693	-0.08719161	-1.8095738
7	0.563648	-0.04120093	-0.599637

Fixed effect interpretation

For each year increase in age, the odds of post-MI mortality increase by 1.18 (95% CI: 1.15, 1.21, p-value <0.001), when adjusted for sex, underlying diabetes mellitus, chronic kidney disease, duration of admission, treatment type, and random effect of hospital, treatment type and duration of admission.

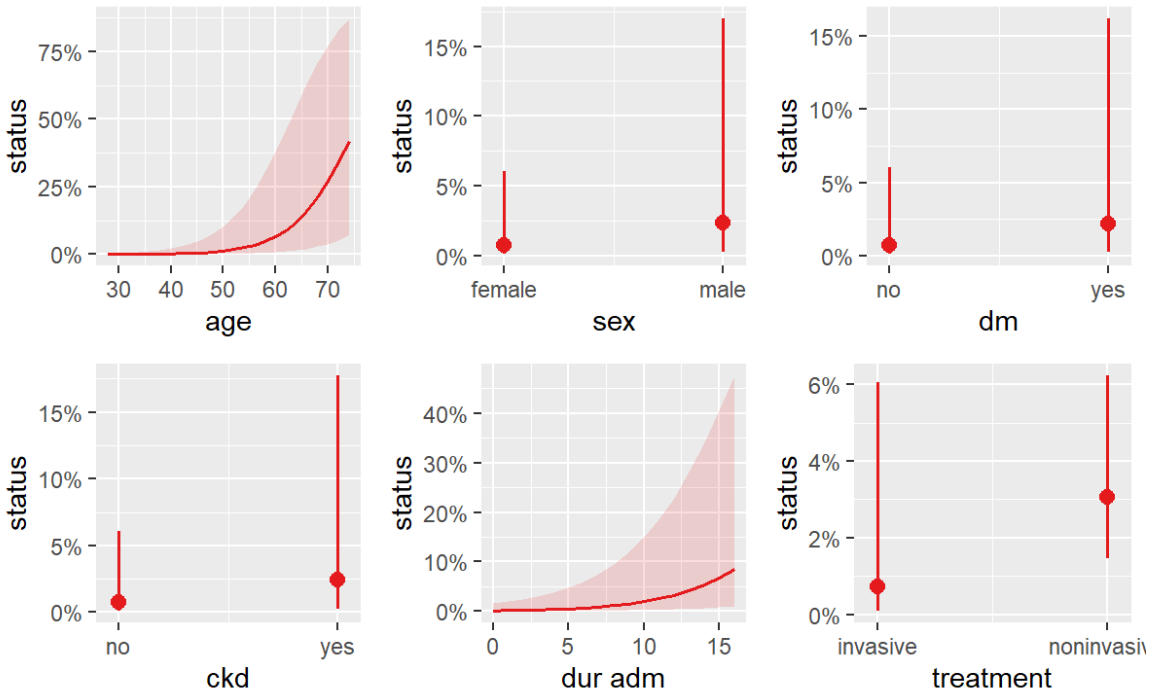
Being male is associated with a 3.25 (95% CI: 2.16, 4.91, p-value <0.001) increase in the odds of post-MI mortality compared to being female), when adjusted for age, underlying diabetes mellitus, chronic kidney disease, duration of admission, treatment type, and random effect of hospital, treatment type and duration of admission.

Having diabetes mellitus is associated with a 2.99 (95% CI: 2.03, 4.39, p-value <0.001) increase in the odds of post-MI mortality compared to not having diabetes mellitus, when adjusted for age, sex, chronic kidney disease, duration of admission, treatment type, and random effect of hospital, treatment type and duration of admission.

Having chronic kidney disease is associated with a 3.38 (95% CI: 2.33, 4.91, p-value <0.001) increase in the odds of post-MI mortality compared to not having chronic kidney disease, when adjusted for age, sex, underlying diabetes mellitus, duration of admission, treatment type, and random effect of hospital, treatment type and duration of admission.

Receiving non-invasive treatment for myocardial infarction is associated with a 4.28 (95% CI: 0.40, 46.37, p-value = 0.023) increase in the odds of post-MI mortality compared to receiving invasive treatment, when adjusted for age, sex, underlying diabetes mellitus, chronic kidney disease, duration of admission, and random effect of hospital, treatment type and duration of admission.

For each day increase in duration of admission, the odds of post-MI mortality increase by 1.29 (95% CI: 1.17. 1.43, p-value <0.001), when adjusted for age, sex, underlying diabetes mellitus, chronic kidney disease, treatment type, and random effect of hospital, treatment type and duration of admission.



Status = post-MI mortality; dm = diabetes mellitus; ckd = chronic kidney disease; dur adm = duration of admission

Figure 2.1. Predicted probability of post-MI mortality based on different risk factors

Random effect interpretation

The random effect of hospital, duration of admission and treatment type on post-MI mortality is visualized by **Figure 2.2-2.3.**

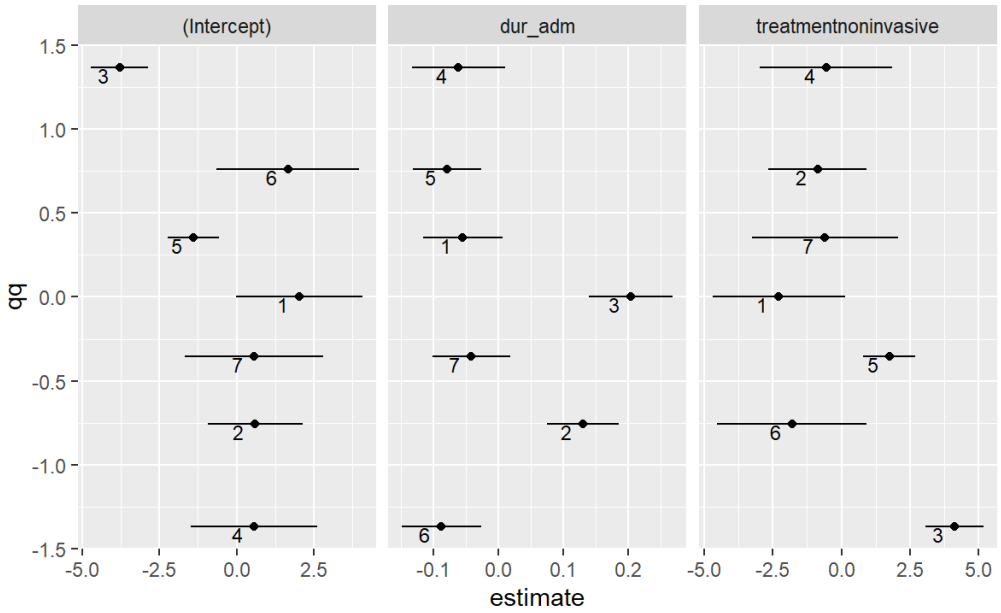


Figure 2.2. Random effects of hospital, duration of admission and type of treatment on post-MI mortality

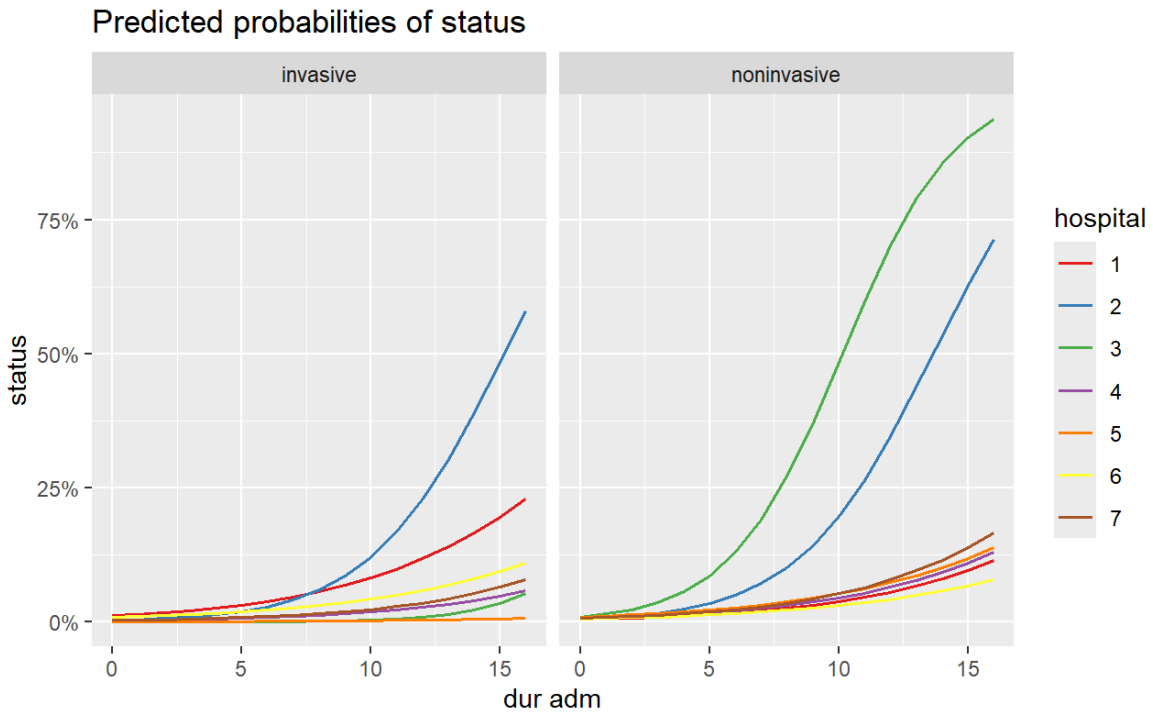


Figure 2.3. Predicted probability of post-MI mortality based on duration of admission and type of treatment across different hospitals

The analysis demonstrated variability in the association between duration of admission, treatment approaches on post-MI patient outcomes across different hospitals. Hospitals with cardiac specialists (4-7) generally show higher intercept estimates, indicating a higher baseline effect, given that these hospitals often manage more severe cases. The effect of non-invasive treatment varies significantly among hospitals that may hinted at different standard of protocol. Hospitals lacking cardiac specialists and opting for non-invasive treatments show higher mortality rates with longer admission durations among post-MI patients.

Model fitness and outlier assessment

The model demonstrates strong predictive performance with an accuracy of 86.2% (95% CI: 84.2%, 88.0%), sensitivity of 93.8% and specificity of 66.4%. The area under the receiver operating characteristic curve (AUROC) of 0.90 (95% CI: 0.88, 0.92) suggests excellent discriminative ability. There are no obvious outliers **(Figure 2.4)**.

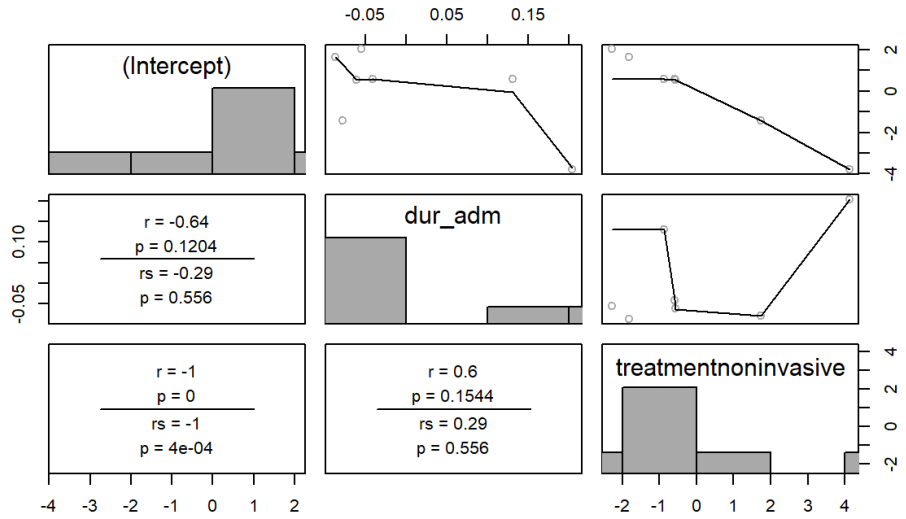


Figure 2.4. Pairwise plot of by-hospital random effects

APPENDIX

The dataset and the code can be retrieved from: https://github.com/drhazlienor/mixedmodel