

# DATA 621 Project: Multi-centre modelling for Newborn Length of Stay in Intensive Care Units based on Diagnoses

Authors: Luke Larter and Trevor Seeger

## Background/Rationale

Newborn infants are particularly vulnerable to health upsets, thus understanding the determinants of intensive care unit (ICU) outcomes for newborns is integral for predicting how patients will recover. Additionally, complexity and duration of care received in neonatal intensive care unit (NICU) can be predictive of long-term effects on developmental outcomes (Subedi, Deboer, & Scharf, 2017). As such, much research has addressed the factors shaping the health outcomes and survival prospects of newborns admitted to the ICU (Foglia et al., 2017; Kurek Eken, Tütün, Özkaya, Karatekin, & Karateke, 2017; Meadow, Lagatta, Andrews, & Lantos, 2012). As prematurity is a common condition requiring hospitalization in the NICU, many studies have specifically lookeded at the NICU outcomes for premature infants (e.g. Manktelow et al., 2010; Kono et al., 2011). However, many health conditions may also require NICU admission for infants who have been brought to term. As infants at differing levels of prematurity will differ in their levels of physiological development, health conditions may interact with levels of prematurity to determine health outcomes (sengupta, 2013). As such, the current study aims to assess NICU patient health outcomes based on level of prematurity in addition to the combination of diagnoses patients have received using the MIMIC-III database. The Multiparameter Intelligent Monitoring in Intensive Care (MIMIC)-III dataset (Johnson et al., 2016) is an freely-accessible, de-identified, ICU-specific dataset with records spanning 10 years. As such, it has been extensively used to test big data solutions in ICU situations. However, to our knowledge there have been no studies that focus completely on the newborns that go to the NICU using this data. While medically, it makes sense that different diagnoses leading to NICU admission will lead to differences in outcome and length of stay, being able to predict the length of stay has not been thoroughly researched. A 2015 systematic literature review found only 9

studies predicting length of stay (Seaton et al., 2016), but of these, only one study included the reason for admission as a predictor, and they studied 4702 very premature infants (23-32 weeks gestation) (Manktelow, Draper, Field, & Field, 2010). Other studies included other conditions of the baby, mother, and other factors. Therefore, it is important to gain the predictive insight into when any baby may be discharged either to other units or home. For parents, having this insight may be beneficial to their engagement and preventing caregiver burnout, as a parent that knows their baby is likely to not be in hospital long will likely suffer less stress, and a parent that knows their baby may be in the hospital longer will be able to find support systems and manage their stress (Peebles-Kleiger, 2000).

# Research Question/Objective

## Research Question

We posed the research question: Can the duration and discharge location of a newborn's ICU stay be modelled based on the diagnoses that led to their admission?

## Aim 1

Using a Cox proportional hazards model, We will investigate how combinations of different health issue diagnoses influence the likelihood of infants being discharge from NICU over time.

## Aim 2

As many health conditions may be commonly associated with premature birth, we will also see whether the prevalence of certain health conditions differs among infants who differ in their level of prematurity.

## Aim 3

We will investigate how patients' health outcomes (i.e. discharged to home, to further hospitalization, etc.) are influenced by the combination of health issues they are diagnosed with when entering NICU.

# Hypotheses:

We hypothesize that the causes for admission to the NICU will influence the length of stay in NICU and the discharge location. Specifically, we hypothesize that the presence of sepsis and prematurity will be most useful in predicting the outcomes due to the multi-system nature of these diagnoses. For example, a premature (or extremely premature) newborn will be a more complex case because they have not reached certain developmental milestones within the entire individual, whereas a malformation may only affect one part of one organ. Therefore, we believe it is prudent to test for interactions with prematurity across all of the other variables, because those organ-specific issues may be triggered or exacerbated by the newborn being born too early in their developmental process.

# Methods

The MIMIC data set includes patients admitted to critical care units consisting of records for 7830 neonates admitted to NICU between 2001 and 2008 in the United States (Johnson et al., 2016). In building our final data set for analysis, demographic information was extracted from the PATIENTS table. The outcomes of interest for our study were the length of patients' stay at the NICU, as well as the location to which a patient was discharged to after their stay at NICU. These data were included in the ADMISSIONS and ICUSTAYS tables. To examine how diagnoses with different health conditions influenced the outcome of interest, we used diagnoses present in the DIAGNOSIS\_ICD table.

# Patient recruitment and inclusion/exclusions criteria

This study utilized newborn infants recorded in the MIMIC-III database who had been admitted to NICU immediately subsequent to their birth ( $n=7830$ ). As we were interested in duration of stay in NICU, we narrowed our data set down to only contain infants whose stay in NICU was longer than 48 hours ( $n=450$ ). This was to ensure than we only had cases requiring prolonged hospitalization.

We wanted to determine how infants' combinations of diagnoses affected their health outcomes, however there were a total of 389 unique diagnoses applied to the 450 newborns in our data set. In looking at counts of diagnoses, many of the most common

diagnoses (with > 10 occurrences) could be grouped into 7 broad diagnosis categories (exact categorizations can be seen in the python data wrangling file). These are:

- i. Cardiac hemorrhages (e.g. intraventricular hemorrhages, pulmonary hemorrhages)
- ii. Malformations of the circulatory system (e.g. patent ductus arteriosus, ventricular septal defect)
- iii. Anomalous heart rates or blood pressure (e.g. neonatal bradycardia, hypotension)
- iv. Prematurity (3 levels: full-term, preterm, extreme preterm (<1kg; this cutoff is coded in the MIMIC data set))
- v. Respiratory issues (e.g. atelectasis, apnea, respiratory distress syndrome)
- vi. Sepsis
- vii. Jaundice

As these were the largest sensible groupings of diagnoses that emerged among the most common diagnoses, we then included any male and female infants who were diagnosed with 1 or more of these conditions in our data set. The above diagnoses are not mutually exclusive, and indeed many are commonly comorbid. For example, malformations of the circulatory system and anomalous heart rate are often found together, but they are not inextricably linked. As such, many patients had multiple of the above conditions. These categories are not all encompassing, for example conditions such as hernias, retrosternal fibroplasia, and metabolic acidosis appeared as somewhat common diagnoses. However, these and other diagnoses not fitting the scheme above could not be meaningfully grouped together, and so were dropped to avoid an excessive number of categories. Each of the 450 patients whose stay exceeded 48 hours had at least one diagnosis fitting into the above categorization scheme, meaning our final number of subjects was 450 infants.

## General outline of analysis

Aim 1 was addressed via a Cox proportional hazards regression modelling approach in R. The full model included presence/absence of all categories of health issue mentioned above, gender, and interactions between level of prematurity and all other health conditions. We decided to include this because preterm infants, especially extreme preterm, are less robust than infants brought to term, and so are often more vulnerable to health upsets (Platt, 2014). Thus, there are clinical grounds for including an interaction between preterm and other health variables. This full model was then pared down by removing non-significant interactions, then non-significant main effects

until only significant variables remained. At each stage of model reduction, proportional hazards assumptions were checked via Schoenfeld plots, likelihood ratio tests were preformed to confirm variable removal would not worsen the model, and the lack of a confounding effect of removed variables was checked (removal did not change betas of remaining variables by > 10%).

Aim 2 was addressed via proportion tests in R. Significant differences in disease prevalence among prematurity groups were investigated with post-hoc pairwise proportion tests, with Bonferroni corrections for multiple comparisons.

Aim 3 was addressed by multinomial modelling in R. After data cleaning, there were only one individual in each of the death, long term care, and cancer groups, therefore these patients were removed as the groups could not be meaningfully combined with others. Testing was carried on with patients who were discharged to short term care, home health care, and home. We first tested interactions between prematurity and other factors, as they are clinically likely to share some overlap, which will extend to the complexity of each case, and through to whether a patient can be discharged home or require more complex care after their NICU stay. Likelihood ratio tests were performed on each interaction, and if these interactions were significant a larger model would have been built. After model building for the interactions was complete, all main effects that did not have main effects estimates that we could conclude were non-zero were removed from the model, one at a time, tested with a likelihood ratio test, and the main effect with the least effect on the model was removed. The updated model followed the same process until all variables had at least one non-zero estimate or the likelihood ratio test, which indicates that removing the variable significantly reduced the variability captured by the model.

## Analysis

Analyses were performed under in R version 3.6.2 (R Core Team, 2007) using RStudio (RStudio Team, 2015), using data wrangling processes from tidyverse (Wickham & Henry, 2020) and dplyr (Wickham, François, Henry, & Müller, 2020). Figures will be made in ggplot (Wickam, 2009) and ggfortify (Tang, Horikoshi, & Li, 2016).

Read data (data cleaned and wrangled in python):

```

data=read.csv('datadfpretermNew.csv', header=T)

#Rename so that the appear as column headings in our descriptive table.
data$preterm[data$preterm == 0] <- 'Full term'
data$preterm[data$preterm == 1] <- 'Preterm'
data$preterm[data$preterm == 2] <- 'Extreme preterm'

head(data)

```

HAD...	<b>cardiac.hemorrhage</b>	<b>circulatory.system.malformation</b>
<int>	<dbl>	<dbl>
1 100135	0	0
2 100419	0	1
3 100545	0	0
4 100591	0	1
5 100665	0	0
6 101015	0	1

6 rows | 1-5 of 13 columns

Change all categorical variables to factors:

```

data$cardiac.hemorrhage=factor(data$cardiac.hemorrhage)
data$circulatory.system.malformation=factor(data$circulatory.system.malformation)
data$heart.rate.circulation.issues=factor(data$heart.rate.circulation.issues)
data$preterm=factor(data$preterm)
data$respiratory=factor(data$respiratory)
data$jaundice=factor(data$jaundice)
data$sepsis=factor(data$sepsis)

data <- within(data, preterm <- relevel(preterm, ref = 'Full term'))

```

# Description of Study Subjects

```
data1 = data

data$GENDER <-
  factor(data$GENDER, levels=c("F", "M"),
         labels=c("Female",
                  "Male"))

data$cardiac.hemorrhage <-
  factor(data$cardiac.hemorrhage, levels=c(0,1),
         labels=c("Absent",
                  "Present"))

data$circulatory.system.malformation <-
  factor(data$circulatory.system.malformation, levels=c(0,1),
         labels=c("Absent",
                  "Present"))

data$heart.rate.circulation.issues <-
  factor(data$heart.rate.circulation.issues, levels=c(0,1),
         labels=c("Absent",
                  "Present"))

data$respiratory <-
  factor(data$respiratory, levels=c(0,1),
         labels=c("Absent",
                  "Present"))

data$sepsis <-
  factor(data$sepsis, levels=c(0,1),
         labels=c("Absent",
                  "Present"))

data$jaundice <-
  factor(data$jaundice, levels=c(0,1),
         labels=c("Absent",
                  "Present"))

data$DISCHARGE_LOCATION <-
  factor(data$DISCHARGE_LOCATION, levels=c( "DEAD/EXPIRED", "DISC-T
RAN CANCER/CHLDRN H", "HOME", "HOME HEALTH CARE", "LONG TERM CARE H
OSPITAL", "SHORT TERM HOSPITAL"),
         labels=c("Dead", "Discharged to Cancer Care/ Children's Ho
spital", "Home", "Home with Health Care", "Long Term Care Facilit
y", "Short Term Hospital"))
```

```

label(data$GENDER)           <- "Gender"
label(data$cardiac.hemorrhage)    <- "Cardiac Hemorrhages"
label(data$circulatory.system.malformation)      <- "Circulatory System Malformation"
label(data$heart.rate.circulation.issues)        <- "Heart Rate or Circulation Dysfunction"
label(data$respiratory)          <- "Respiratory Issues"
label(data$sepsis)              <- "Sepsis"
label(data$jaundice)            <- "Jaundice"
label(data$DISCHARGE_LOCATION)   <- "Discharge Location"
label(data$LOS)                 <- "Length of Stay"



```

	<b>Full term (n=38)</b>	<b>Extreme preterm (n=171)</b>	
<b>Gender</b>			
Female	17 (44.7%)	86 (50.3%)	1
Male	21 (55.3%)	85 (49.7%)	1
<b>Cardiac Hemorrhages</b>			
Absent	32 (84.2%)	126 (73.7%)	1
Present	6 (15.8%)	45 (26.3%)	2
<b>Circulatory System Malformation</b>			
Absent	21 (55.3%)	65 (38.0%)	1
Present	17 (44.7%)	106 (62.0%)	5
<b>Heart Rate or Circulation Dysfunction</b>			
Absent	21 (55.3%)	52 (30.4%)	7
Present	17 (44.7%)	119 (69.6%)	1
<b>Respiratory Issues</b>			

	<b>Full term (n=38)</b>	<b>Extreme preterm (n=171)</b>	
Absent	3 (7.9%)	2 (1.2%)	
Present	35 (92.1%)	169 (98.8%)	2
<b>Sepsis</b>			
Absent	30 (78.9%)	107 (62.6%)	2
Present	8 (21.1%)	64 (37.4%)	3
<b>Jaundice</b>			
Absent	11 (28.9%)	10 (5.8%)	
Present	27 (71.1%)	161 (94.2%)	2
<b>Discharge Location</b>			
Dead	0 (0%)	1 (0.6%)	
Discharged to Cancer Care/ Children's Hospital	1 (2.6%)	0 (0%)	
Home	8 (21.1%)	29 (17.0%)	6
Home with Health Care	26 (68.4%)	99 (57.9%)	1
Long Term Care Facility	0 (0%)	1 (0.6%)	
Short Term Hospital	3 (7.9%)	41 (24.0%)	
<b>Length of Stay</b>			
Mean (SD)	70.3 (19.6)	95.0 (27.3)	6
Median [Min, Max]	62.7 [49.0, 116]	90.7 [48.2, 172]	65.

```
data = data1
```

# Addressing Aim 1): Cox Proportional Hazard Modelling

Model and Variables:

We built a cox proportional hazards regression model to determine whether infant gender and diagnosis with certain categories of health insult affect the likelihood of being discharged from the neonatal intensive care unit, based on their length of stay in

the NICU. Thus, the event which we modeled with our time-to-event model is a positive health outcome; discharge from NICU. This analysis was conducted using the survival (Therneau & Grambsch, 2000), survminer (Kassambara, Kosinski, & Biecek, 2019), survivalMPL (Couturier, Ma, Heritier, & Manuguerra, 2017), and biostat3 (Karlsson & Clements, 2019) packages. The categorical variables that we used are:

- Level of prematurity (full-term, preterm, extreme preterm (<1kg))
- Whether the infant has been diagnosed with a cardiac hemorrhage (Y/N)
- Whether the infant has been diagnosed with a malformation of some part of their circulatory system (Y/N)
- Whether the infant has been diagnosed with anomalous heart rate or blood pressure (Y/N)
- Whether the infant has been diagnosed with respiratory issues (Y/N)
- Whether the infant has been diagnosed with sepsis (Y/N)
- Whether the infant has been diagnosed with jaundice (Y/N)
- Gender (M/F)

We also decided to include interactions between each of these health conditions and the infant's level of prematurity. We decided to include this because preterm infants, especially extreme preterm, are less robust than infants brought to term, and so are often more vulnerable to health upsets (Platt, 2014). Thus, there are clinical grounds for including an interaction between preterm and other health variables.

Only one infant in our final set of subjects died while in the NICU, and so was the only censored data point in our analysis.

```
full_model <- coxph(Surv(LOS, delta) ~ GENDER + preterm*cardiac.hemorrhage + preterm*circulatory.system.malformation + preterm*heart.rate.circulation.issues + preterm*respiratory + preterm*sepsis + preterm*jaundice, method=c("breslow"), data = data)
summary(full_model)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ GENDER + preterm * cardiac.hemorrhage +
##         preterm * circulatory.system.malformation + preterm * heart.rate.circulation.issues +
##         preterm * respiratory + preterm * sepsis + preterm * jaundice,
##       data = data, method = c("breslow"))
##
## n= 450, number of events= 449
##
##                                         coef
exp(coef)
## GENDERM                               0.06922
1.07167
## pretermExtreme preterm                -1.62635
0.19665
## pretermPreterm                         -0.67613
0.50858
## cardiac.hemorrhage1                  -0.54847
0.57784
## circulatory.system.malformation1     -1.09833
0.33343
## heart.rate.circulation.issues1       -0.23362
0.79166
## respiratory1                          -0.39926
0.67082
## sepsis1                                -0.86084
0.42281
## jaundice1                             -0.16890
0.84460
## pretermExtreme preterm:cardiac.hemorrhage1   0.78058
2.18275
## pretermPreterm:cardiac.hemorrhage1        0.39192
1.47982
## pretermExtreme preterm:circulatory.system.malformation1  1.04469
2.84251
## pretermPreterm:circulatory.system.malformation1        0.70432
2.02246
## pretermExtreme preterm:heart.rate.circulation.issues1    0.45009

```

1.56846	
## pretermPreterm:heart.rate.circulation.issues1	0.15662
1.16955	
## pretermExtreme preterm:respiratory1	-0.61113
0.54274	
## pretermPreterm:respiratory1	0.06276
1.06477	
## pretermExtreme preterm:sepsis1	0.48584
1.62553	
## pretermPreterm:sepsis1	0.27581
1.31759	
## pretermExtreme preterm:jaundice1	0.16426
1.17852	
## pretermPreterm:jaundice1	0.10689
1.11281	
##	se(coef)
z	
## GENDERM	0.09911
0.698	
## pretermExtreme preterm	1.07305
-1.516	
## pretermPreterm	0.90107
-0.750	
## cardiac.hemorrhage1	0.47178
-1.163	
## circulatory.system.malformation1	0.35956
-3.055	
## heart.rate.circulation.issues1	0.35961
-0.650	
## respiratory1	0.68468
-0.583	
## sepsis1	0.45641
-1.886	
## jaundice1	0.38649
-0.437	
## pretermExtreme preterm:cardiac.hemorrhage1	0.50599
1.543	
## pretermPreterm:cardiac.hemorrhage1	0.50471
0.777	
## pretermExtreme preterm:circulatory.system.malformation1	0.39583

2.639	
## pretermPreterm:circulatory.system.malformation1	0.38336
1.837	
## pretermExtreme preterm:heart.rate.circulation.issues1	0.40180
1.120	
## pretermPreterm:heart.rate.circulation.issues1	0.38792
0.404	
## pretermExtreme preterm:respiratory1	1.04999
-0.582	
## pretermPreterm:respiratory1	0.79738
0.079	
## pretermExtreme preterm:sepsis1	0.48688
0.998	
## pretermPreterm:sepsis1	0.49534
0.557	
## pretermExtreme preterm:jaundice1	0.53537
0.307	
## pretermPreterm:jaundice1	0.45931
0.233	
##	Pr (>   z   )
## GENDERM	0.48494
## pretermExtreme preterm	0.12961
## pretermPreterm	0.45304
## cardiac.hemorrhage1	0.24502
## circulatory.system.malformation1	0.00225
**	
## heart.rate.circulation.issues1	0.51592
## respiratory1	0.55980
## sepsis1	0.05928
.	
## jaundice1	0.66212
## pretermExtreme preterm:cardiac.hemorrhage1	0.12290
## pretermPreterm:cardiac.hemorrhage1	0.43744
## pretermExtreme preterm:circulatory.system.malformation1	0.00831
**	
## pretermPreterm:circulatory.system.malformation1	0.06618
.	
## pretermExtreme preterm:heart.rate.circulation.issues1	0.26263
## pretermPreterm:heart.rate.circulation.issues1	0.68641
## pretermExtreme preterm:respiratory1	0.56055

```
## pretermPreterm:respiratory1          0.93726
## pretermExtreme preterm:sepsis1      0.31835
## pretermPreterm:sepsis1              0.57767
## pretermExtreme preterm:jaundice1    0.75899
## pretermPreterm:jaundice1           0.81598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coef)
##
## GENDERM                           1.071
7
## pretermExtreme preterm             0.196
6
## pretermPreterm                     0.508
6
## cardiac.hemorrhage1               0.577
8
## circulatory.system.malformation1  0.333
4
## heart.rate.circulation.issues1   0.791
7
## respiratory1                      0.670
8
## sepsis1                            0.422
8
## jaundice1                          0.844
6
## pretermExtreme preterm:cardiac.hemorrhage1 2.182
7
## pretermPreterm:cardiac.hemorrhage1       1.479
8
## pretermExtreme preterm:circulatory.system.malformation1 2.842
5
## pretermPreterm:circulatory.system.malformation1        2.022
5
## pretermExtreme preterm:heart.rate.circulation.issues1 1.568
5
## pretermPreterm:heart.rate.circulation.issues1         1.169
5
```

## pretermExtreme	preterm:respiratory1	0.542
7		
## pretermPreterm:respiratory1		1.064
8		
## pretermExtreme	preterm:sepsis1	1.625
5		
## pretermPreterm:sepsis1		1.317
6		
## pretermExtreme	preterm:jaundice1	1.178
5		
## pretermPreterm:jaundice1		1.112
8		
##		exp (-coe
f)		
## GENDERM		0.93
31		
## pretermExtreme	preterm	5.08
53		
## pretermPreterm		1.96
62		
## cardiac.hemorrhage1		1.73
06		
## circulatory.system.malformation1		2.99
92		
## heart.rate.circulation.issues1		1.26
32		
## respiratory1		1.49
07		
## sepsis1		2.36
51		
## jaundice1		1.18
40		
## pretermExtreme	preterm:cardiac.hemorrhage1	0.45
81		
## pretermPreterm:cardiac.hemorrhage1		0.67
58		
## pretermExtreme	preterm:circulatory.system.malformation1	0.35
18		
## pretermPreterm:circulatory.system.malformation1		0.49
44		

```
## pretermExtreme preterm:heart.rate.circulation.issues1          0.63
76
## pretermPreterm:heart.rate.circulation.issues1          0.85
50
## pretermExtreme preterm:respiratory1          1.84
25
## pretermPreterm:respiratory1          0.93
92
## pretermExtreme preterm:sepsis1          0.61
52
## pretermPreterm:sepsis1          0.75
90
## pretermExtreme preterm:jaundice1          0.84
85
## pretermPreterm:jaundice1          0.89
86
##                                         lower .9
5
## GENDERM          0.8824
6
## pretermExtreme preterm          0.0240
1
## pretermPreterm          0.0869
7
## cardiac.hemorrhage1          0.2292
0
## circulatory.system.malformation1          0.1648
0
## heart.rate.circulation.issues1          0.3912
4
## respiratory1          0.1753
1
## sepsis1          0.1728
4
## jaundice1          0.3959
7
## pretermExtreme preterm:cardiac.hemorrhage1          0.8096
7
## pretermPreterm:cardiac.hemorrhage1          0.5503
0
```

```
## pretermExtreme preterm:circulatory.system.malformation1 1.3085
0
## pretermPreterm:circulatory.system.malformation1 0.9540
3
## pretermExtreme preterm:heart.rate.circulation.issues1 0.7136
1
## pretermPreterm:heart.rate.circulation.issues1 0.5467
9
## pretermExtreme preterm:respiratory1 0.0693
2
## pretermPreterm:respiratory1 0.2231
1
## pretermExtreme preterm:sepsis1 0.6259
8
## pretermPreterm:sepsis1 0.4990
5
## pretermExtreme preterm:jaundice1 0.4126
9
## pretermPreterm:jaundice1 0.4523
3
## upper .9
5
## GENDERM 1.301
4
## pretermExtreme preterm 1.610
9
## pretermPreterm 2.974
1
## cardiac.hemorrhage1 1.456
8
## circulatory.system.malformation1 0.674
6
## heart.rate.circulation.issues1 1.601
9
## respiratory1 2.566
9
## sepsis1 1.034
3
## jaundice1 1.801
5
```

```

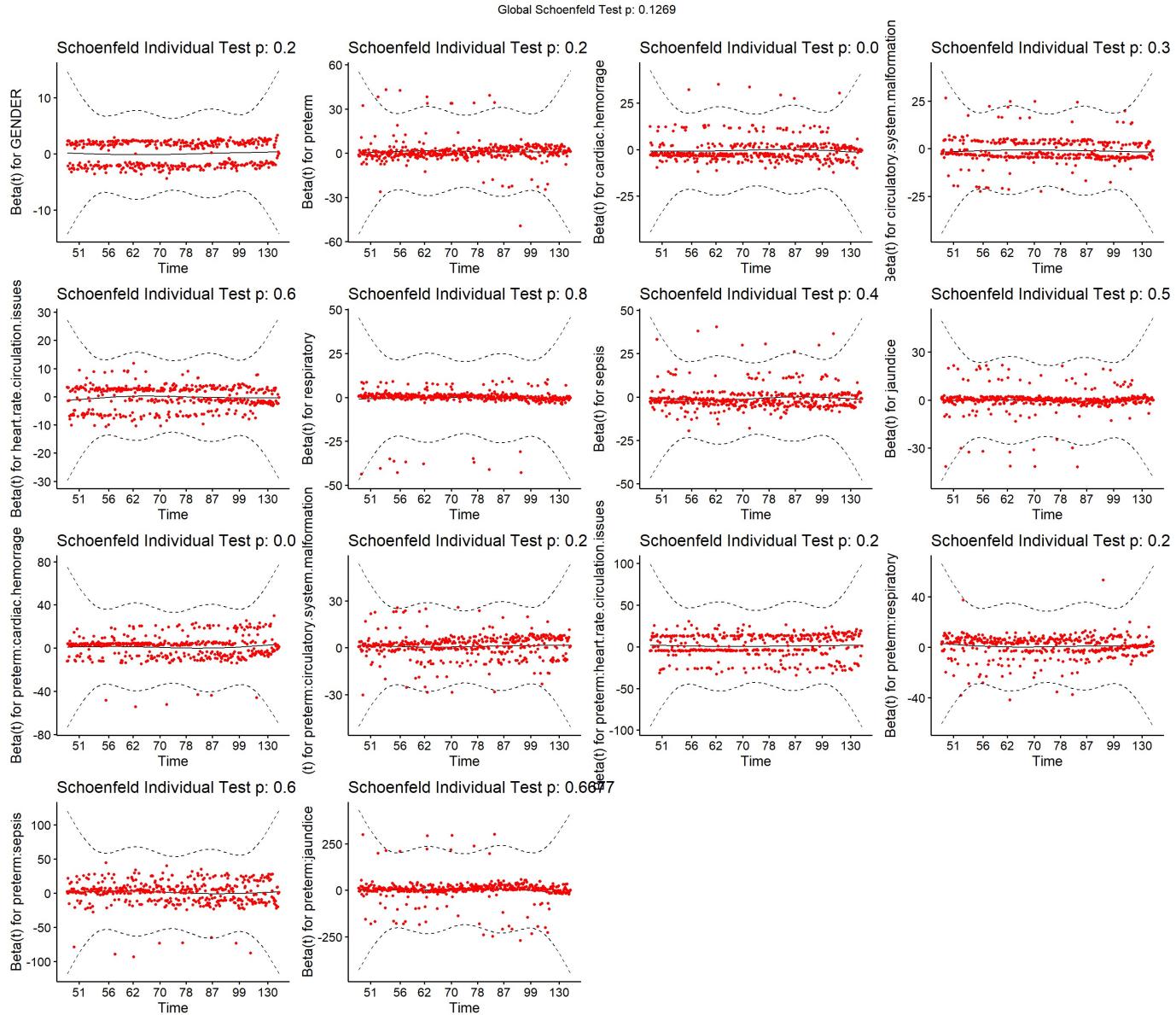
## pretermExtreme preterm:cardiac.hemorragel           5.884
4
## pretermPreterm:cardiac.hemorragel                 3.979
4
## pretermExtreme preterm:circulatory.system.malformation1 6.174
9
## pretermPreterm:circulatory.system.malformation1    4.287
4
## pretermExtreme preterm:heart.rate.circulation.issues1 3.447
3
## pretermPreterm:heart.rate.circulation.issues1       2.501
6
## pretermExtreme preterm:respiratory1                4.249
5
## pretermPreterm:respiratory1                        5.081
5
## pretermExtreme preterm:sepsis1                   4.221
2
## pretermPreterm:sepsis1                          3.478
7
## pretermExtreme preterm:jaundice1                3.365
5
## pretermPreterm:jaundice1                      2.737
7
##
## Concordance= 0.684  (se = 0.013 )
## Likelihood ratio test= 158.2  on 21 df,   p=<0.0000000000000002
## Wald test              = 153.2  on 21 df,   p=<0.0000000000000002
## Score (logrank) test = 170  on 21 df,   p=<0.0000000000000002

```

```

test1 <- cox.zph(full_model)
ggcoxzph(test1)

```



- Global tests show full model is significant
- The global p value from the Schoenfeld plot indicates the model agrees with the proportional hazards assumption. However, cardiac hemorrhage and its interaction have local p values that indicate a violation of this assumption. For the time being, we will keep these in the model and see whether these become significant. If so, we will attempt to remedy this violation, e.g. by stratification of this variable. This caveat will apply to subsequent model tests.
- Many of the interaction terms are not significant. We will proceed by removing these 1 by 1, in descending order of their distance from our significance threshold of 0.05.

**Remove interaction term furthest from significance (preterm\*jaundice):**

```
reduced_modelA <-coxph(Surv(LOS, delta) ~ GENDER + preterm*cardiac.h  
emorrhage + preterm*circulatory.system.malformation + preterm*heart.  
rate.circulation.issues + preterm*respiratory + preterm*sepsis + ja  
undice, method=c("breslow"), data = data)  
summary(reduced_modelA)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ GENDER + preterm * cardiac.hemorrhage +
##         preterm * circulatory.system.malformation + preterm * heart.rate.circulation.issues +
##         preterm * respiratory + preterm * sepsis + jaundice, data =
##         data,
##         method = c("breslow"))

##
## n= 450, number of events= 449

## exp(coef)
##                               coef
## GENDERM                      0.06868
## 1.07109
## pretermExtreme preterm        -1.59781
## 0.20234
## pretermPreterm                 -0.64532
## 0.52449
## cardiac.hemorrhage1          -0.57551
## 0.56242
## circulatory.system.malformation1 -1.11929
## 0.32651
## heart.rate.circulation.issues1 -0.22866
## 0.79560
## respiratory1                  -0.43284
## 0.64866
## sepsis1                       -0.86758
## 0.41996
## jaundice1                     -0.07090
## 0.93156
## pretermExtreme preterm:cardiac.hemorrhage1 0.80458
## 2.23575
## pretermPreterm:cardiac.hemorrhage1           0.41813
## 1.51912
## pretermExtreme preterm:circulatory.system.malformation1 1.06815
## 2.91000
## pretermPreterm:circulatory.system.malformation1       0.72498
## 2.06469
## pretermExtreme preterm:heart.rate.circulation.issues1 0.44932

```

1.56725	
## pretermPreterm:heart.rate.circulation.issues1	0.15130
1.16335	
## pretermExtreme preterm:respiratory1	-0.52175
0.59348	
## pretermPreterm:respiratory1	0.10091
1.10618	
## pretermExtreme preterm:sepsis1	0.49586
1.64191	
## pretermPreterm:sepsis1	0.28175
1.32545	
##	se(coef)
z	
## GENDERM	0.09896
0.694	
## pretermExtreme preterm	1.06739
-1.497	
## pretermPreterm	0.88634
-0.728	
## cardiac.hemorrhage1	0.46328
-1.242	
## circulatory.system.malformation1	0.35063
-3.192	
## heart.rate.circulation.issues1	0.35831
-0.638	
## respiratory1	0.67366
-0.643	
## sepsis1	0.45406
-1.911	
## jaundice1	0.18064
-0.392	
## pretermExtreme preterm:cardiac.hemorrhage1	0.49884
1.613	
## pretermPreterm:cardiac.hemorrhage1	0.49836
0.839	
## pretermExtreme preterm:circulatory.system.malformation1	0.38690
2.761	
## pretermPreterm:circulatory.system.malformation1	0.37521
1.932	
## pretermExtreme preterm:heart.rate.circulation.issues1	0.40036

```

1.122
## pretermPreterm:heart.rate.circulation.issues1          0.38658
0.391
## pretermExtreme preterm:respiratory1                  0.99707
-0.523
## pretermPreterm:respiratory1                         0.77569
0.130
## pretermExtreme preterm:sepsis1                      0.48425
1.024
## pretermPreterm:sepsis1                            0.49330
0.571
##
## GENDERM                                         Pr(>|z|)
## pretermExtreme preterm                           0.48769
## pretermPreterm                                0.13441
## pretermPreterm                                0.46656
## cardiac.hemorrhage1                          0.21415
## circulatory.system.malformation1            0.00141
**
## heart.rate.circulation.issues1                0.52336
## respiratory1                                 0.52053
## sepsis1                                      0.05604
.
## jaundice1                                    0.69471
## pretermExtreme preterm:cardiac.hemorrhage1    0.10677
## pretermPreterm:cardiac.hemorrhage1           0.40146
## pretermExtreme preterm:circulatory.system.malformation1 0.00577
**
## pretermPreterm:circulatory.system.malformation1   0.05333
.
## pretermExtreme preterm:heart.rate.circulation.issues1 0.26174
## pretermPreterm:heart.rate.circulation.issues1        0.69551
## pretermExtreme preterm:respiratory1                 0.60077
## pretermPreterm:respiratory1                        0.89649
## pretermExtreme preterm:sepsis1                   0.30585
## pretermPreterm:sepsis1                           0.56789
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coe
f)

```

## GENDERM	1.071
1	
## pretermExtreme preterm	0.202
3	
## pretermPreterm	0.524
5	
## cardiac.hemorrhage1	0.562
4	
## circulatory.system.malformation1	0.326
5	
## heart.rate.circulation.issues1	0.795
6	
## respiratory1	0.648
7	
## sepsis1	0.420
0	
## jaundice1	0.931
6	
## pretermExtreme preterm:cardiac.hemorrhage1	2.235
8	
## pretermPreterm:cardiac.hemorrhage1	1.519
1	
## pretermExtreme preterm:circulatory.system.malformation1	2.910
0	
## pretermPreterm:circulatory.system.malformation1	2.064
7	
## pretermExtreme preterm:heart.rate.circulation.issues1	1.567
2	
## pretermPreterm:heart.rate.circulation.issues1	1.163
3	
## pretermExtreme preterm:respiratory1	0.593
5	
## pretermPreterm:respiratory1	1.106
2	
## pretermExtreme preterm:sepsis1	1.641
9	
## pretermPreterm:sepsis1	1.325
5	
##	exp (-coe
f)	

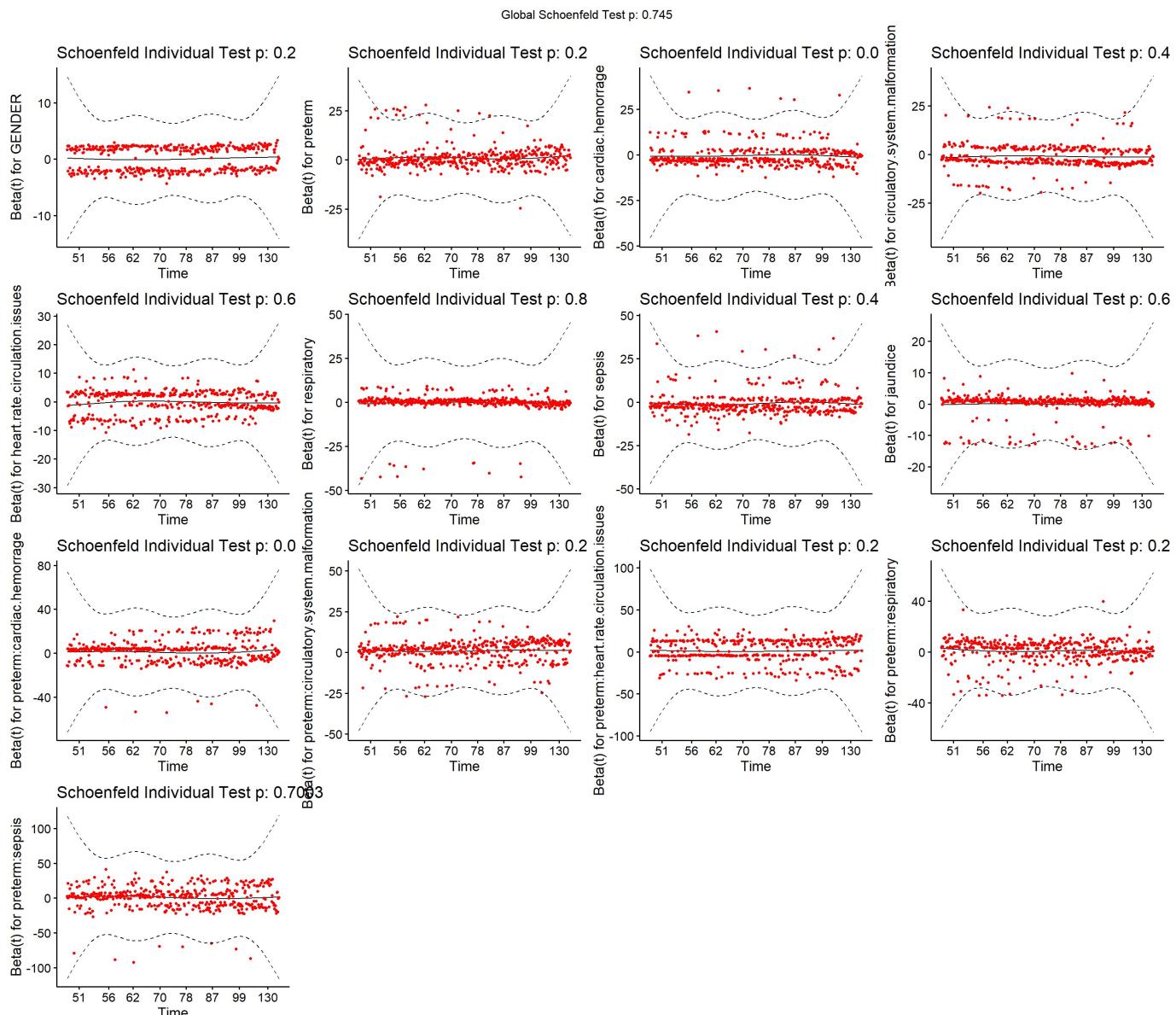
## GENDERM	0.93
36	
## pretermExtreme preterm	4.94
22	
## pretermPreterm	1.90
66	
## cardiac.hemorrhage1	1.77
80	
## circulatory.system.malformation1	3.06
27	
## heart.rate.circulation.issues1	1.25
69	
## respiratory1	1.54
16	
## sepsis1	2.38
12	
## jaundice1	1.07
35	
## pretermExtreme preterm:cardiac.hemorrhage1	0.44
73	
## pretermPreterm:cardiac.hemorrhage1	0.65
83	
## pretermExtreme preterm:circulatory.system.malformation1	0.34
36	
## pretermPreterm:circulatory.system.malformation1	0.48
43	
## pretermExtreme preterm:heart.rate.circulation.issues1	0.63
81	
## pretermPreterm:heart.rate.circulation.issues1	0.85
96	
## pretermExtreme preterm:respiratory1	1.68
50	
## pretermPreterm:respiratory1	0.90
40	
## pretermExtreme preterm:sepsis1	0.60
90	
## pretermPreterm:sepsis1	0.75
45	
##	lower .9
5	

## GENDERM	0.8822
4	
## pretermExtreme preterm	0.0249
8	
## pretermPreterm	0.0923
2	
## cardiac.hemorrhage1	0.2268
4	
## circulatory.system.malformation1	0.1642
3	
## heart.rate.circulation.issues1	0.3941
9	
## respiratory1	0.1732
2	
## sepsis1	0.1724
7	
## jaundice1	0.6538
1	
## pretermExtreme preterm:cardiac.hemorrhage1	0.8410
3	
## pretermPreterm:cardiac.hemorrhage1	0.5719
9	
## pretermExtreme preterm:circulatory.system.malformation1	1.3632
2	
## pretermPreterm:circulatory.system.malformation1	0.9896
5	
## pretermExtreme preterm:heart.rate.circulation.issues1	0.7150
7	
## pretermPreterm:heart.rate.circulation.issues1	0.5453
3	
## pretermExtreme preterm:respiratory1	0.0840
8	
## pretermPreterm:respiratory1	0.2418
6	
## pretermExtreme preterm:sepsis1	0.6355
6	
## pretermPreterm:sepsis1	0.5040
4	
##	upper .9
5	

## GENDERM	1.300
4	
## pretermExtreme preterm	1.639
2	
## pretermPreterm	2.979
8	
## cardiac.hemorrhage1	1.394
5	
## circulatory.system.malformation1	0.649
2	
## heart.rate.circulation.issues1	1.605
8	
## respiratory1	2.429
1	
## sepsis1	1.022
6	
## jaundice1	1.327
3	
## pretermExtreme preterm:cardiac.hemorrhage1	5.943
4	
## pretermPreterm:cardiac.hemorrhage1	4.034
5	
## pretermExtreme preterm:circulatory.system.malformation1	6.211
8	
## pretermPreterm:circulatory.system.malformation1	4.307
6	
## pretermExtreme preterm:heart.rate.circulation.issues1	3.435
0	
## pretermPreterm:heart.rate.circulation.issues1	2.481
8	
## pretermExtreme preterm:respiratory1	4.189
0	
## pretermPreterm:respiratory1	5.059
3	
## pretermExtreme preterm:sepsis1	4.241
7	
## pretermPreterm:sepsis1	3.485
4	
##	
## Concordance= 0.684 (se = 0.013 )	

```
## Likelihood ratio test= 158.1 on 19 df, p=<0.0000000000000002
## Wald test = 153.1 on 19 df, p=<0.0000000000000002
## Score (logrank) test = 169.8 on 19 df, p=<0.0000000000000002
```

```
test2 <- cox.zph(reduced_modelA)
ggcoxzph(test2)
```



- Global tests show reduced model is significant
- Schoenfeld plot global p value shows model meets proportional hazards assumption.

```
anova(full_model, reduced_modelA)
```

	<b>loglik</b> <dbl>	<b>Chisq</b> <dbl>	<b>Df</b> <int>	<b>P(&gt; Chi )</b> <dbl>
1	-2220.260	NA	NA	NA
2	-2220.309	0.09754741	2	0.9523966
2 rows				

- The full model is not significantly better than the reduced model, thus we will remove this interaction and proceed.

## Remove interaction term furthest from significance (preterm\*respiratory):

```
reduced_modelB <- coxph(Surv(LOS, delta) ~ GENDER+preterm*cardiac.hem
orragic + preterm*circulatory.system.malformation + preterm*heart.ra
te.circulation.issues + respiratory + preterm*sepsis + jaundice, me
thod=c("breslow"), data = data)
summary(reduced_modelB)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ GENDER + preterm * cardiac.hemorrhage +
##         preterm * circulatory.system.malformation + preterm * heart.rate.circulation.issues +
##         respiratory + preterm * sepsis + jaundice, data = data, method = c("breslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## GENDERM                                0.06592
1.06814
## pretermExtreme preterm                  -2.09527
0.12304
## pretermPreterm                           -0.54382
0.58053
## cardiac.hemorrhage1                   -0.57395
0.56330
## circulatory.system.malformation1       -1.11720
0.32719
## heart.rate.circulation.issues1        -0.22952
0.79491
## respiratory1                            -0.43512
0.64719
## sepsis1                                 -0.86891
0.41941
## jaundice1                              -0.07933
0.92373
## pretermExtreme preterm:cardiac.hemorrhage1    0.79865
2.22253
## pretermPreterm:cardiac.hemorrhage1          0.41816
1.51916
## pretermExtreme preterm:circulatory.system.malformation1 1.05787
2.88022
## pretermPreterm:circulatory.system.malformation1       0.71949
2.05338
## pretermExtreme preterm:heart.rate.circulation.issues1 0.43653
1.54733

```

## pretermPreterm:heart.rate.circulation.issues1	0.15024
1.16212	
## pretermExtreme preterm:sepsis1	0.49617
1.64242	
## pretermPreterm:sepsis1	0.28629
1.33148	
##	se(coef)
z	
## GENDERM	0.09866
0.668	
## pretermExtreme preterm	0.38433
-5.452	
## pretermPreterm	0.36419
-1.493	
## cardiac.hemorragel	0.45278
-1.268	
## circulatory.system.malformation1	0.35064
-3.186	
## heart.rate.circulation.issues1	0.34645
-0.663	
## respiratory1	0.31987
-1.360	
## sepsis1	0.43608
-1.993	
## jaundice1	0.17895
-0.443	
## pretermExtreme preterm:cardiac.hemorragel	0.48924
1.632	
## pretermPreterm:cardiac.hemorragel	0.48925
0.855	
## pretermExtreme preterm:circulatory.system.malformation1	0.38644
2.737	
## pretermPreterm:circulatory.system.malformation1	0.37503
1.918	
## pretermExtreme preterm:heart.rate.circulation.issues1	0.38972
1.120	
## pretermPreterm:heart.rate.circulation.issues1	0.37544
0.400	
## pretermExtreme preterm:sepsis1	0.46735
1.062	

	Pr
## pretermPreterm:sepsis1	0.47860
0.598	
##	
(>   z   )	
## GENDERM	0.5
0402	
## pretermExtreme preterm	0.000000
0499 ***	
## pretermPreterm	0.1
3538	
## cardiac.hemorrhage1	0.2
0494	
## circulatory.system.malformation1	0.0
0144 **	
## heart.rate.circulation.issues1	0.5
0765	
## respiratory1	0.1
7374	
## sepsis1	0.0
4631 *	
## jaundice1	0.6
5753	
## pretermExtreme preterm:cardiac.hemorrhage1	0.1
0259	
## pretermPreterm:cardiac.hemorrhage1	0.3
9273	
## pretermExtreme preterm:circulatory.system.malformation1	0.0
0619 **	
## pretermPreterm:circulatory.system.malformation1	0.0
5505 .	
## pretermExtreme preterm:heart.rate.circulation.issues1	0.2
6267	
## pretermPreterm:heart.rate.circulation.issues1	0.6
8902	
## pretermExtreme preterm:sepsis1	0.2
8839	
## pretermPreterm:sepsis1	0.5
4971	
## ---	
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	

```

## exp (coe
##
f)
## GENDERM 1.068
1
## pretermExtreme preterm 0.123
0
## pretermPreterm 0.580
5
## cardiac.hemorrhage1 0.563
3
## circulatory.system.malformation1 0.327
2
## heart.rate.circulation.issues1 0.794
9
## respiratory1 0.647
2
## sepsis1 0.419
4
## jaundice1 0.923
7
## pretermExtreme preterm:cardiac.hemorrhage1 2.222
5
## pretermPreterm:cardiac.hemorrhage1 1.519
2
## pretermExtreme preterm:circulatory.system.malformation1 2.880
2
## pretermPreterm:circulatory.system.malformation1 2.053
4
## pretermExtreme preterm:heart.rate.circulation.issues1 1.547
3
## pretermPreterm:heart.rate.circulation.issues1 1.162
1
## pretermExtreme preterm:sepsis1 1.642
4
## pretermPreterm:sepsis1 1.331
5
## exp (-coe
f)
## GENDERM 0.93

```

62		
## pretermExtreme	preterm	8.12
77		
## pretermPreterm		1.72
26		
## cardiac.hemorrhage1		1.77
53		
## circulatory.system.malformation1		3.05
63		
## heart.rate.circulation.issues1		1.25
80		
## respiratory1		1.54
51		
## sepsis1		2.38
43		
## jaundice1		1.08
26		
## pretermExtreme	preterm:cardiac.hemorrhage1	0.44
99		
## pretermPreterm:cardiac.hemorrhage1		0.65
83		
## pretermExtreme	preterm:circulatory.system.malformation1	0.34
72		
## pretermPreterm:circulatory.system.malformation1		0.48
70		
## pretermExtreme	preterm:heart.rate.circulation.issues1	0.64
63		
## pretermPreterm:heart.rate.circulation.issues1		0.86
05		
## pretermExtreme	preterm:sepsis1	0.60
89		
## pretermPreterm:sepsis1		0.75
10		
##		lower .9
5		
## GENDER	M	0.8803
4		
## pretermExtreme	preterm	0.0579
3		
## pretermPreterm		0.2843

3		
## cardiac.hemorrhage1		0.2319
1		
## circulatory.system.malformation1		0.1645
7		
## heart.rate.circulation.issues1		0.4031
1		
## respiratory1		0.3457
4		
## sepsis1		0.1784
2		
## jaundice1		0.6504
6		
## pretermExtreme preterm:cardiac.hemorrhage1		0.8519
3		
## pretermPreterm:cardiac.hemorrhage1		0.5823
0		
## pretermExtreme preterm:circulatory.system.malformation1		1.3504
8		
## pretermPreterm:circulatory.system.malformation1		0.9845
6		
## pretermExtreme preterm:heart.rate.circulation.issues1		0.7208
6		
## pretermPreterm:heart.rate.circulation.issues1		0.5567
7		
## pretermExtreme preterm:sepsis1		0.6571
6		
## pretermPreterm:sepsis1		0.5211
4		
##	upper .9	
5		
## GENDER.M		1.296
0		
## pretermExtreme preterm		0.261
3		
## pretermPreterm		1.185
3		
## cardiac.hemorrhage1		1.368
2		
## circulatory.system.malformation1		0.650

```

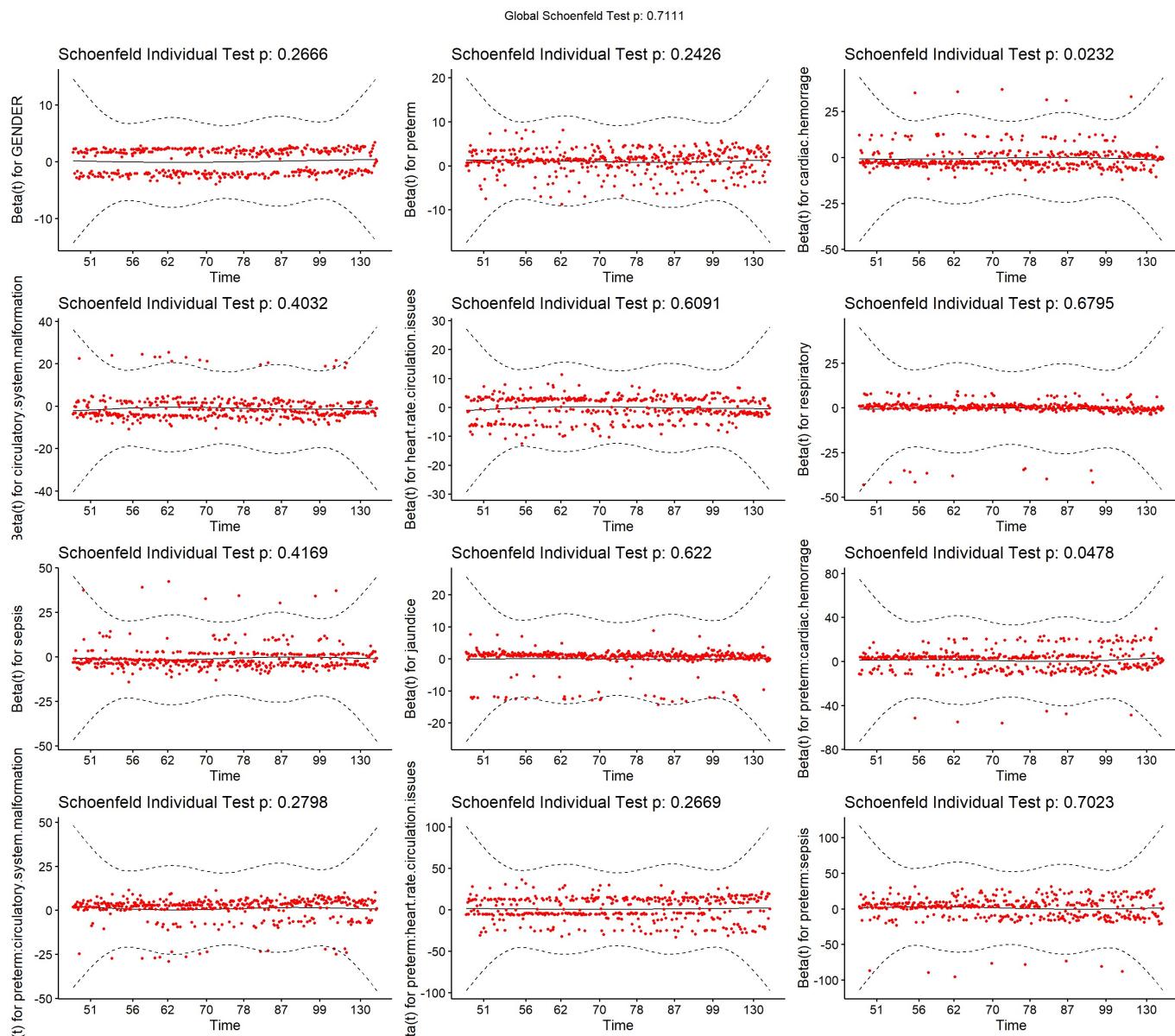
5
## heart.rate.circulation.issues1          1.567
5
## respiratory1                          1.211
5
## sepsis1                             0.985
9
## jaundice1                           1.311
8
## pretermExtreme preterm:cardiac.hemorrhage1 5.798
2
## pretermPreterm:cardiac.hemorrhage1      3.963
3
## pretermExtreme preterm:circulatory.system.malformation1 6.142
8
## pretermPreterm:circulatory.system.malformation1      4.282
5
## pretermExtreme preterm:heart.rate.circulation.issues1 3.321
4
## pretermPreterm:heart.rate.circulation.issues1        2.425
6
## pretermExtreme preterm:sepsis1                4.104
8
## pretermPreterm:sepsis1                      3.401
9
##
## Concordance= 0.683  (se = 0.014 )
## Likelihood ratio test= 157.6  on 17 df,    p=<0.0000000000000002
## Wald test           = 153.2  on 17 df,    p=<0.0000000000000002
## Score (logrank) test = 169.7  on 17 df,    p=<0.0000000000000002

```

```

test3 <- cox.zph(reduced_modelB)
ggcoxzph(test3)

```



- Global tests show reduced model is significant
- Schoenfeld plot global p value shows model meets proportional hazards assumption.

```
anova(reduced_modelA, reduced_modelB)
```

	loglik <dbl>	Chisq <dbl>	Df <int>	P(> Chi ) <dbl>
1	-2220.309	NA	NA	NA
2	-2220.563	0.5083886	2	0.7755411
2 rows				

- The larger model is not significantly better than the reduced model, thus we will remove this interaction and proceed.

## Remove interaction term furthest from significance (preterm\*heart.rate.circulation.issues):

```
reduced_modelC <- coxph(Surv(LOS, delta) ~ GENDER + preterm*cardiac.h  
emorrhage + preterm*circulatory.system.malformation + heart.rate.cir  
culation.issues + respiratory + preterm*sepsis + jaundice, method=c  
("breslow"), data = data)  
summary(reduced_modelC)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ GENDER + preterm * cardiac.hemorrhage +
##         preterm * circulatory.system.malformation + heart.rate.circulation.issues +
##         respiratory + preterm * sepsis + jaundice, data = data, method = c("breslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## GENDERM                                0.06310
1.06514
## pretermExtreme preterm                  -1.84475
0.15807
## pretermPreterm                           -0.46125
0.63050
## cardiac.hemorrhage1                   -0.59027
0.55418
## circulatory.system.malformation1      -1.07477
0.34138
## heart.rate.circulation.issues1        0.01407
1.01417
## respiratory1                            -0.39536
0.67344
## sepsis1                                 -0.77548
0.46048
## jaundice1                              -0.05995
0.94181
## pretermExtreme preterm:cardiac.hemorrhage1   0.83083
2.29522
## pretermPreterm:cardiac.hemorrhage1        0.44634
1.56258
## pretermExtreme preterm:circulatory.system.malformation1  1.04047
2.83054
## pretermPreterm:circulatory.system.malformation1       0.66393
1.94241
## pretermExtreme preterm:sepsis1            0.43123
1.53915

```

## pretermPreterm:sepsis1	0.19722
1.21801	
##	se (coef)
z	
## GENDER.M	0.09828
0.642	
## pretermExtreme preterm	0.30986
-5.953	
## pretermPreterm	0.28548
-1.616	
## cardiac.hemorrhage1	0.45185
-1.306	
## circulatory.system.malformation1	0.34529
-3.113	
## heart.rate.circulation.issues1	0.10590
0.133	
## respiratory1	0.31827
-1.242	
## sepsis1	0.42058
-1.844	
## jaundice1	0.17868
-0.335	
## pretermExtreme preterm:cardiac.hemorrhage1	0.48793
1.703	
## pretermPreterm:cardiac.hemorrhage1	0.48829
0.914	
## pretermExtreme preterm:circulatory.system.malformation1	0.38211
2.723	
## pretermPreterm:circulatory.system.malformation1	0.37092
1.790	
## pretermExtreme preterm:sepsis1	0.45407
0.950	
## pretermPreterm:sepsis1	0.46475
0.424	
##	Pr
(>   z   )	
## GENDER.M	0.
52080	
## pretermExtreme preterm	0.000000
00263 ***	

```
## pretermPreterm 0.
10616
## cardiac.hemorragel 0.
19144
## circulatory.system.malformation1 0.
00185 **
## heart.rate.circulation.issues1 0.
89427
## respiratory1 0.
21416
## sepsis1 0.
06521 .
## jaundice1 0.
73725
## pretermExtreme preterm:cardiac.hemorragel 0.
08861 .
## pretermPreterm:cardiac.hemorragel 0.
36067
## pretermExtreme preterm:circulatory.system.malformation1 0.
00647 **
## pretermPreterm:circulatory.system.malformation1 0.
07346 .
## pretermExtreme preterm:sepsis1 0.
34227
## pretermPreterm:sepsis1 0.
67131
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coef)
f)
## GENDERM 1.065
1
## pretermExtreme preterm 0.158
1
## pretermPreterm 0.630
5
## cardiac.hemorragel 0.554
2
## circulatory.system.malformation1 0.341
```

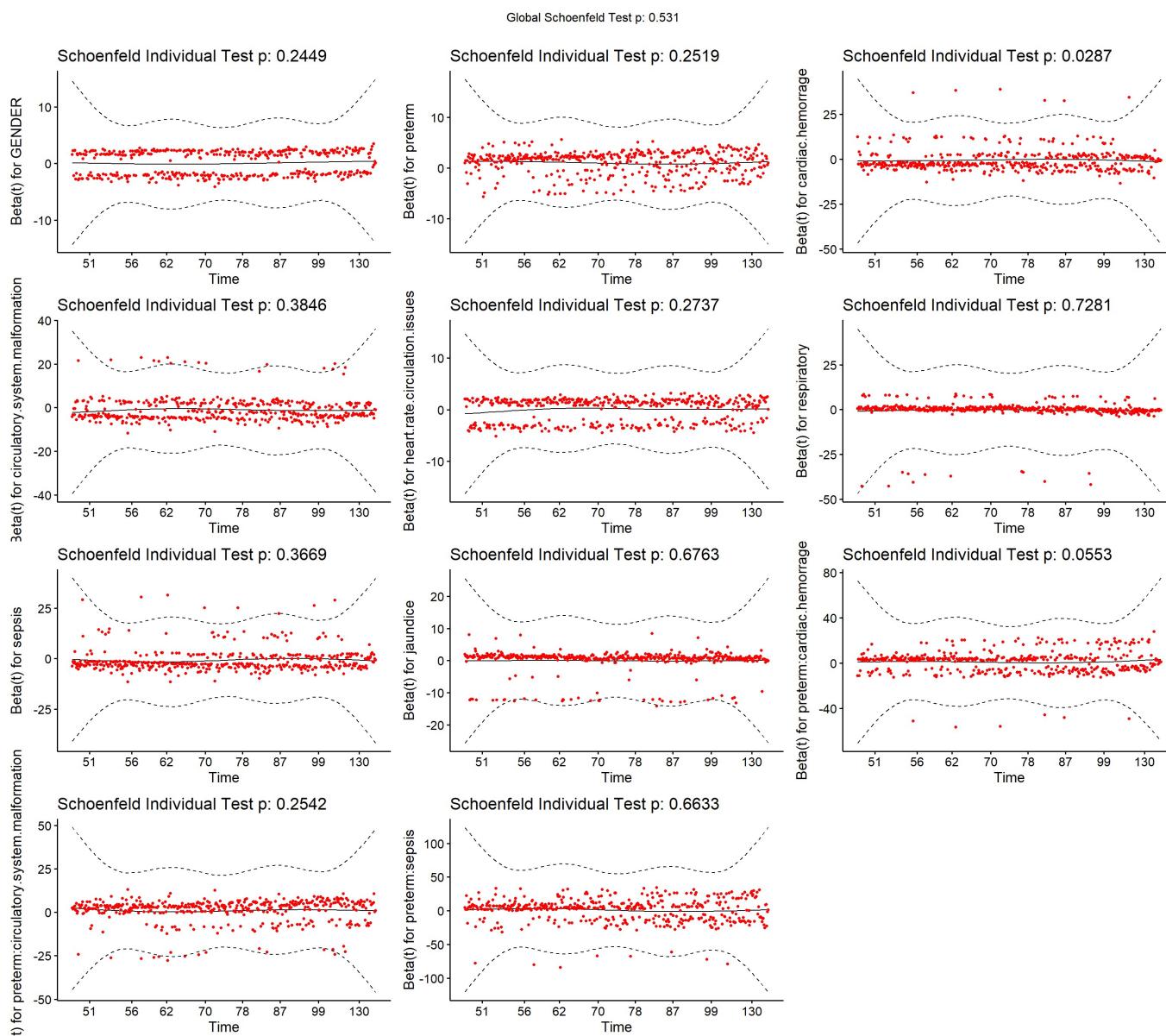
```
4
## heart.rate.circulation.issues1          1.014
2
## respiratory1                          0.673
4
## sepsis1                             0.460
5
## jaundice1                           0.941
8
## pretermExtreme preterm:cardiac.hemorrhage1 2.295
2
## pretermPreterm:cardiac.hemorrhage1      1.562
6
## pretermExtreme preterm:circulatory.system.malformation1 2.830
5
## pretermPreterm:circulatory.system.malformation1      1.942
4
## pretermExtreme preterm:sepsis1          1.539
2
## pretermPreterm:sepsis1                 1.218
0
##                                         exp (-coe
f)
## GENDERM                            0.93
88
## pretermExtreme preterm               6.32
65
## pretermPreterm                      1.58
60
## cardiac.hemorrhage1                1.80
45
## circulatory.system.malformation1   2.92
93
## heart.rate.circulation.issues1      0.98
60
## respiratory1                       1.48
49
## sepsis1                            2.17
16
## jaundice1                          1.06
```

```
18  
## pretermExtreme preterm:cardiac.hemorrhage1 0.43  
57  
## pretermPreterm:cardiac.hemorrhage1 0.64  
00  
## pretermExtreme preterm:circulatory.system.malformation1 0.35  
33  
## pretermPreterm:circulatory.system.malformation1 0.51  
48  
## pretermExtreme preterm:sepsis1 0.64  
97  
## pretermPreterm:sepsis1 0.82  
10  
## lower .9  
5  
## GENDERM 0.8785  
2  
## pretermExtreme preterm 0.0861  
2  
## pretermPreterm 0.3603  
2  
## cardiac.hemorrhage1 0.2285  
8  
## circulatory.system.malformation1 0.1735  
1  
## heart.rate.circulation.issues1 0.8240  
8  
## respiratory1 0.3609  
0  
## sepsis1 0.2019  
4  
## jaundice1 0.6635  
5  
## pretermExtreme preterm:cardiac.hemorrhage1 0.8820  
7  
## pretermPreterm:cardiac.hemorrhage1 0.6000  
8  
## pretermExtreme preterm:circulatory.system.malformation1 1.3384  
9  
## pretermPreterm:circulatory.system.malformation1 0.9388
```

```
8  
## pretermExtreme preterm:sepsis1 0.6320  
8  
## pretermPreterm:sepsis1 0.4898  
4  
## upper .9  
5  
## GENDERM 1.291  
4  
## pretermExtreme preterm 0.290  
1  
## pretermPreterm 1.103  
3  
## cardiac.hemorrhage1 1.343  
6  
## circulatory.system.malformation1 0.671  
6  
## heart.rate.circulation.issues1 1.248  
1  
## respiratory1 1.256  
6  
## sepsis1 1.050  
1  
## jaundice1 1.336  
8  
## pretermExtreme preterm:cardiac.hemorrhage1 5.972  
4  
## pretermPreterm:cardiac.hemorrhage1 4.068  
9  
## pretermExtreme preterm:circulatory.system.malformation1 5.985  
9  
## pretermPreterm:circulatory.system.malformation1 4.018  
6  
## pretermExtreme preterm:sepsis1 3.747  
9  
## pretermPreterm:sepsis1 3.028  
6  
##  
## Concordance= 0.682 (se = 0.014 )  
## Likelihood ratio test= 155.4 on 15 df, p=<0.0000000000000002
```

```
## Wald test = 151.9 on 15 df, p=<0.0000000000000002
## Score (logrank) test = 168.2 on 15 df, p=<0.0000000000000002
```

```
test4 <- cox.zph(reduced_modelC)
ggcoxzph(test4)
```



- Global tests show reduced model is significant
- Schoenfeld plot global p value shows model meets proportional hazards assumption.

```
anova(reduced_modelB, reduced_modelC)
```

	<b>loglik</b> <dbl>	<b>Chisq</b> <dbl>	<b>Df</b> <int>	<b>P(&gt; Chi )</b> <dbl>
1	-2220.563	NA	NA	NA
2	-2221.633	2.139556	2	0.3430847
2 rows				

- The larger model is not significantly better than the reduced model, thus we will remove this interaction and proceed.

## Remove interaction term furthest from significance (preterm\*cardiac.hemorrhage):

```
reduced_modelD <- coxph(Surv(LOS, delta) ~ GENDER + cardiac.hemorrhage
+ preterm*circulatory.system.malformation + heart.rate.circulation.
issues + respiratory + preterm*sepsis + jaundice, method=c("breslow"),
data = data)
summary(reduced_modelD)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ GENDER + cardiac.hemorrhage +
##        preterm * circulatory.system.malformation + heart.rate.circu-
##        lation.issues +
##        respiratory + preterm * sepsis + jaundice, data = data, meth-
##        od = c("breslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## GENDERM                               0.05103
1.05235
## cardiac.hemorrhage1                  -0.01976
0.98043
## pretermExtreme preterm               -1.67739
0.18686
## pretermPreterm                         -0.35320
0.70244
## circulatory.system.malformation1     -1.06695
0.34406
## heart.rate.circulation.issues1       0.02357
1.02385
## respiratory1                          -0.32779
0.72051
## sepsis1                                -0.73619
0.47893
## jaundice1                             -0.08622
0.91740
## pretermExtreme preterm:circulatory.system.malformation1  1.05623
2.87551
## pretermPreterm:circulatory.system.malformation1          0.65098
1.91742
## pretermExtreme preterm:sepsis1           0.42931
1.53619
## pretermPreterm:sepsis1                  0.11421
1.12099
##
##                                         se(coef)
z
## GENDERM                               0.09811

```

0.520	
## cardiac.hemorrhage1	0.12129
-0.163	
## pretermExtreme preterm	0.29887
-5.612	
## pretermPreterm	0.27447
-1.287	
## circulatory.system.malformation1	0.34707
-3.074	
## heart.rate.circulation.issues1	0.10529
0.224	
## respiratory1	0.31139
-1.053	
## sepsis1	0.42043
-1.751	
## jaundice1	0.17537
-0.492	
## pretermExtreme preterm:circulatory.system.malformation1	0.38285
2.759	
## pretermPreterm:circulatory.system.malformation1	0.37218
1.749	
## pretermExtreme preterm:sepsis1	0.45376
0.946	
## pretermPreterm:sepsis1	0.46331
0.247	
## z   )	Pr (>
## GENDERM	0.602
97	
## cardiac.hemorrhage1	0.870
58	
## pretermExtreme preterm	0.000000
02 ***	
## pretermPreterm	0.198
16	
## circulatory.system.malformation1	0.002
11 **	
## heart.rate.circulation.issues1	0.822
88	
## respiratory1	0.292

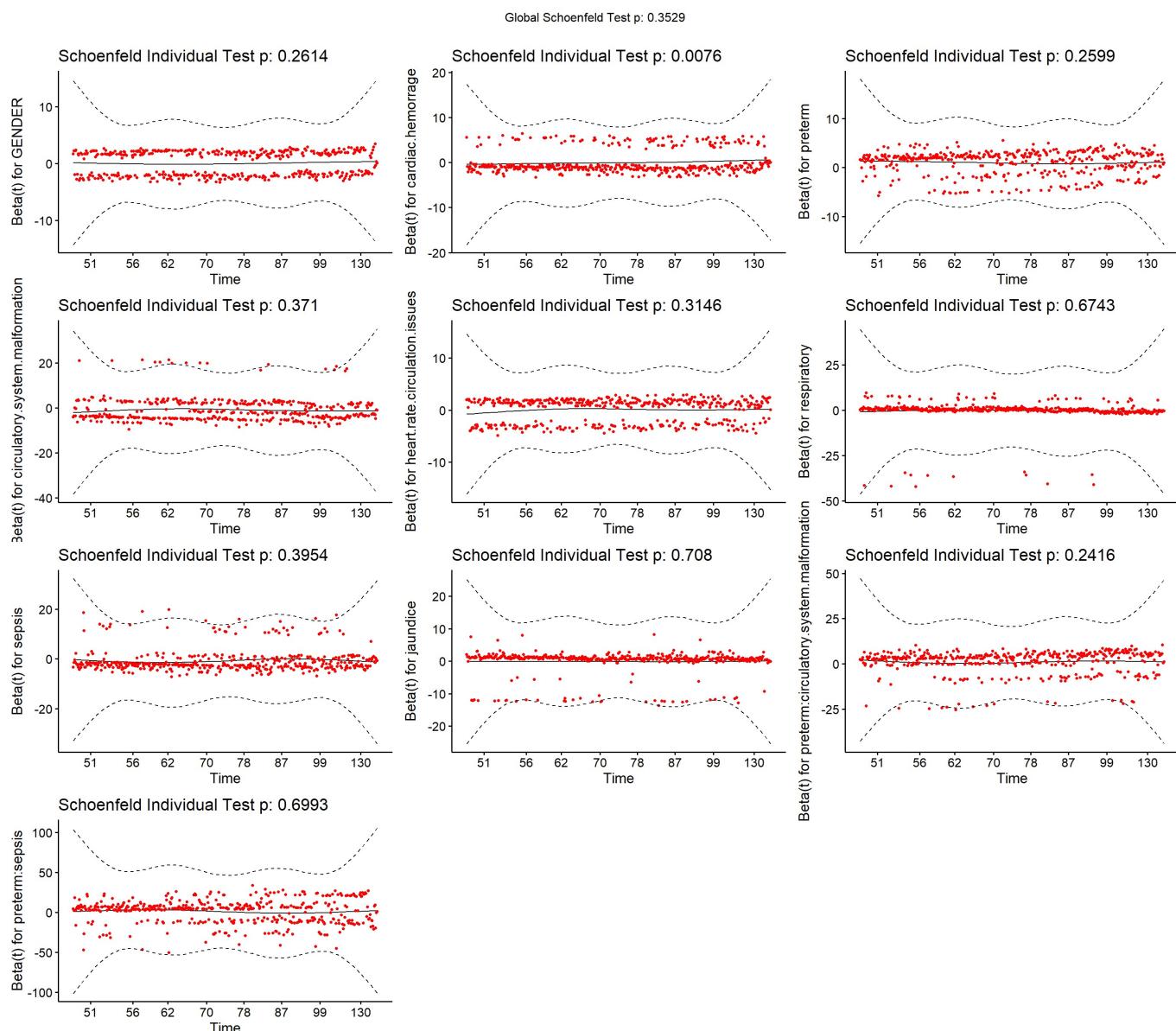
```
49
## sepsis1                               0.079
93 .
## jaundice1                             0.622
98
## pretermExtreme preterm:circulatory.system.malformation1 0.005
80 **
## pretermPreterm:circulatory.system.malformation1      0.080
27 .
## pretermExtreme preterm:sepsis1           0.344
09
## pretermPreterm:sepsis1                 0.805
28
## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                         exp(coe
f)
## GENDERM                                1.052
4
## cardiac.hemorragel                     0.980
4
## pretermExtreme preterm                  0.186
9
## pretermPreterm                           0.702
4
## circulatory.system.malformation1       0.344
1
## heart.rate.circulation.issues1         1.023
8
## respiratory1                            0.720
5
## sepsis1                                 0.478
9
## jaundice1                             0.917
4
## pretermExtreme preterm:circulatory.system.malformation1 2.875
5
## pretermPreterm:circulatory.system.malformation1      1.917
4
```

## pretermExtreme	1.536
2	
## pretermPreterm:sepsis1	1.121
0	
##	exp (-coe
f)	
## GENDERM	0.95
03	
## cardiac.hemorrhage1	1.02
00	
## pretermExtreme preterm	5.35
16	
## pretermPreterm	1.42
36	
## circulatory.system.malformation1	2.90
65	
## heart.rate.circulation.issues1	0.97
67	
## respiratory1	1.38
79	
## sepsis1	2.08
80	
## jaundice1	1.09
00	
## pretermExtreme preterm:circulatory.system.malformation1	0.34
78	
## pretermPreterm:circulatory.system.malformation1	0.52
15	
## pretermExtreme preterm:sepsis1	0.65
10	
## pretermPreterm:sepsis1	0.89
21	
##	lower .9
5	
## GENDERM	0.868
3	
## cardiac.hemorrhage1	0.773
0	
## pretermExtreme preterm	0.104
0	

## pretermPreterm	0.410
2	
## circulatory.system.malformation1	0.174
3	
## heart.rate.circulation.issues1	0.832
9	
## respiratory1	0.391
4	
## sepsis1	0.210
1	
## jaundice1	0.650
6	
## pretermExtreme preterm:circulatory.system.malformation1	1.357
8	
## pretermPreterm:circulatory.system.malformation1	0.924
5	
## pretermExtreme preterm:sepsis1	0.631
3	
## pretermPreterm:sepsis1	0.452
1	
##	upper .9
5	
## GENDERM	1.275
5	
## cardiac.hemorrhage1	1.243
5	
## pretermExtreme preterm	0.335
7	
## pretermPreterm	1.202
9	
## circulatory.system.malformation1	0.679
3	
## heart.rate.circulation.issues1	1.258
5	
## respiratory1	1.326
5	
## sepsis1	1.091
8	
## jaundice1	1.293
7	

```
## pretermExtreme preterm:circulatory.system.malformation1      6.089
8
## pretermPreterm:circulatory.system.malformation1              3.976
6
## pretermExtreme preterm:sepsis1                            3.738
4
## pretermPreterm:sepsis1                                2.779
5
##
## Concordance= 0.677  (se = 0.013 )
## Likelihood ratio test= 151.2  on 13 df,  p=<0.0000000000000002
## Wald test          = 147.5  on 13 df,  p=<0.0000000000000002
## Score (logrank) test = 162.9  on 13 df,  p=<0.0000000000000002
```

```
test5 <- cox.zph(reduced_modelD)
ggcoxzph(test5)
```



- Global tests show reduced model is significant
- Schoenfeld plot global p value shows model meets proportional hazards assumption.

```
anova(reduced_modelD, reduced_modelC)
```

	loglik	Chisq	Df	P(> Chi )
	<dbl>	<dbl>	<int>	<dbl>
1	-2223.777	NA	NA	NA
2	-2221.633	4.288776	2	0.1171397
2 rows				

- The larger model is not significantly better than the reduced model, thus we will remove this interaction and proceed.

## Remove interaction term furthest from significance (preterm\*sepsis):

```
reduced_modelE <- coxph(Surv(LOS, delta) ~ GENDER + cardiac.hemorrhage  
+ preterm*circulatory.system.malformation + heart.rate.circulation.  
issues + respiratory + sepsis + jaundice, method=c("breslow"), data  
= data)  
summary(reduced_modelE)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ GENDER + cardiac.hemorrhage +
##        preterm * circulatory.system.malformation + heart.rate.circu-
##        lation.issues +
##        respiratory + sepsis + jaundice, data = data, method = c("br-
##        eslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## GENDERM                               0.04746
1.04861
## cardiac.hemorrhage1                  -0.02670
0.97365
## pretermExtreme preterm               -1.55479
0.21123
## pretermPreterm                         -0.28948
0.74865
## circulatory.system.malformation1     -1.00448
0.36624
## heart.rate.circulation.issues1       0.04177
1.04266
## respiratory1                          -0.32626
0.72162
## sepsis1                                -0.47297
0.62315
## jaundice1                             -0.06464
0.93741
## pretermExtreme preterm:circulatory.system.malformation1  1.02010
2.77348
## pretermPreterm:circulatory.system.malformation1          0.57342
1.77433
##
##                                         se(coef)
z
## GENDERM                               0.09781
0.485
## cardiac.hemorrhage1                  0.12070
-0.221
## pretermExtreme preterm                0.26570

```

```

-5.852
## pretermPreterm 0.24240
-1.194
## circulatory.system.malformation1 0.33468
-3.001
## heart.rate.circulation.issues1 0.10426
0.401
## respiratory1 0.30937
-1.055
## sepsis1 0.11987
-3.946
## jaundice1 0.17418
-0.371
## pretermExtreme preterm:circulatory.system.malformation1 0.37156
2.745
## pretermPreterm:circulatory.system.malformation1 0.36125
1.587
##
Pr
(>|z|)
## GENDERM 0.
62750
## cardiac.hemorrhage1 0.
82492
## pretermExtreme preterm 0.000000
00487 ***
## pretermPreterm 0.
23239
## circulatory.system.malformation1 0.
00269 **
## heart.rate.circulation.issues1 0.
68867
## respiratory1 0.
29161
## sepsis1 0.000079
59035 ***
## jaundice1 0.
71056
## pretermExtreme preterm:circulatory.system.malformation1 0.
00604 **
## pretermPreterm:circulatory.system.malformation1 0.

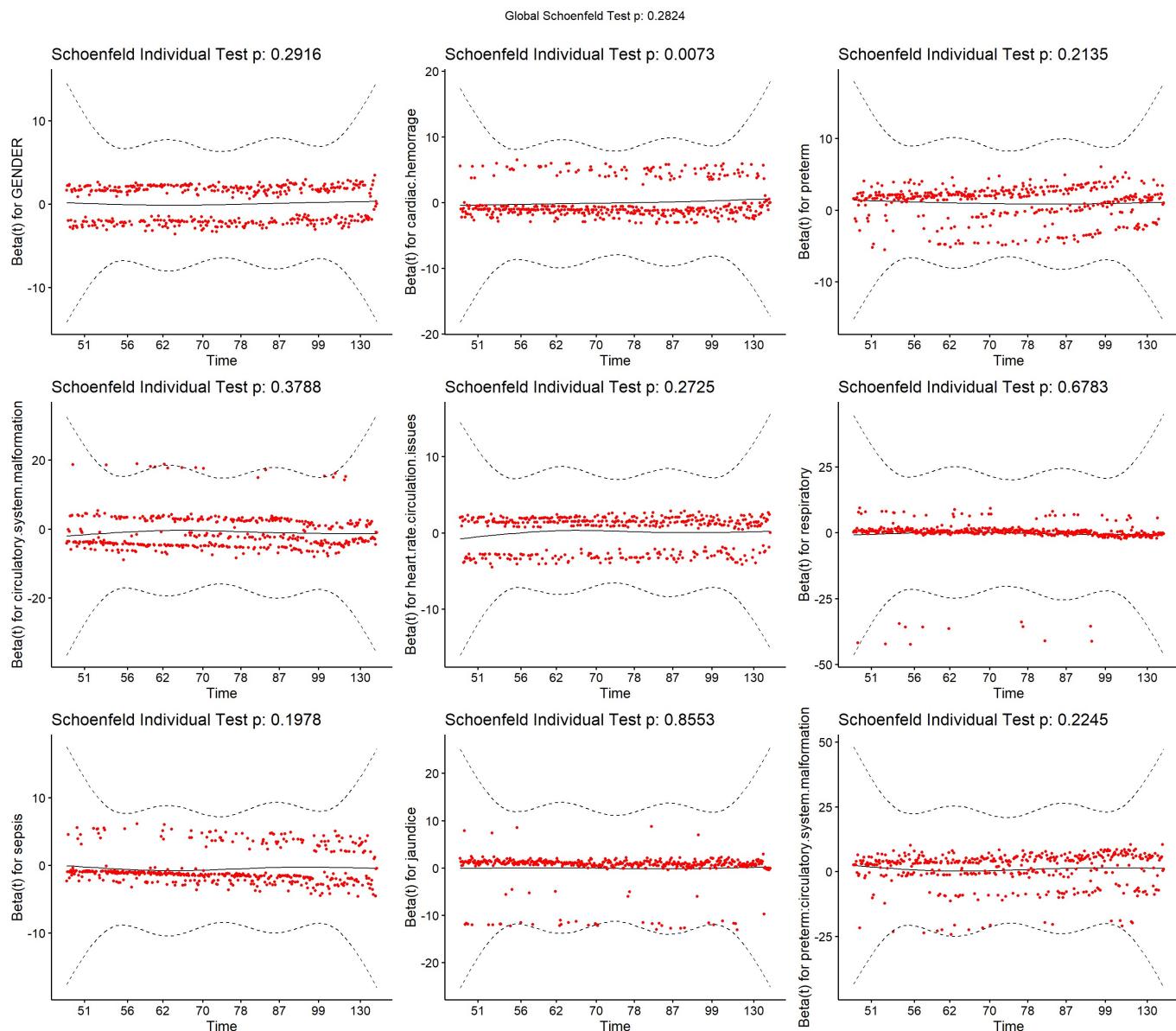
```

```
11244
## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                         exp (coe
f)
## GENDERM                                1.048
6
## cardiac.hemorrhage1                     0.973
7
## pretermExtreme preterm                  0.211
2
## pretermPreterm                           0.748
7
## circulatory.system.malformation1       0.366
2
## heart.rate.circulation.issues1         1.042
7
## respiratory1                            0.721
6
## sepsis1                                 0.623
1
## jaundice1                               0.937
4
## pretermExtreme preterm:circulatory.system.malformation1 2.773
5
## pretermPreterm:circulatory.system.malformation1        1.774
3
##
##                                         exp (-coe
f)
## GENDERM                                0.95
36
## cardiac.hemorrhage1                     1.02
71
## pretermExtreme preterm                  4.73
41
## pretermPreterm                           1.33
57
## circulatory.system.malformation1       2.73
05
```

## heart.rate.circulation.issues1	0.95
91	
## respiratory1	1.38
58	
## sepsis1	1.60
48	
## jaundice1	1.06
68	
## pretermExtreme preterm:circulatory.system.malformation1	0.36
06	
## pretermPreterm:circulatory.system.malformation1	0.56
36	
##	lower .9
5	
## GENDERM	0.865
7	
## cardiac.hemorrhage1	0.768
5	
## pretermExtreme preterm	0.125
5	
## pretermPreterm	0.465
5	
## circulatory.system.malformation1	0.190
1	
## heart.rate.circulation.issues1	0.850
0	
## respiratory1	0.393
5	
## sepsis1	0.492
7	
## jaundice1	0.666
3	
## pretermExtreme preterm:circulatory.system.malformation1	1.338
9	
## pretermPreterm:circulatory.system.malformation1	0.874
1	
##	upper .9
5	
## GENDERM	1.270
2	

```
## cardiac.hemorragel 1.233
5
## pretermExtreme preterm 0.355
6
## pretermPreterm 1.204
0
## circulatory.system.malformation1 0.705
7
## heart.rate.circulation.issues1 1.279
0
## respiratory1 1.323
3
## sepsis1 0.788
2
## jaundice1 1.318
8
## pretermExtreme preterm:circulatory.system.malformation1 5.745
1
## pretermPreterm:circulatory.system.malformation1 3.601
9
##
## Concordance= 0.678 (se = 0.013 )
## Likelihood ratio test= 149.1 on 11 df, p=<0.0000000000000002
## Wald test = 142.1 on 11 df, p=<0.0000000000000002
## Score (logrank) test = 155.1 on 11 df, p=<0.0000000000000002
```

```
test6 <- cox.zph(reduced_modelE)
ggcoxzph(test6)
```



- Global tests show reduced model is significant
- Schoenfeld plot global p value shows model meets proportional hazards assumption.

```
anova(reduced_modelD, reduced_modelE)
```

	loglik <dbl>	Chisq <dbl>	Df <int>	P(> Chi ) <dbl>
1	-2223.777	NA	NA	NA
2	-2224.789	2.023127	2	0.3636499

2 rows

- The larger model is not significantly better than the reduced model, thus we will remove this interaction and proceed.
- Now we just have a model including the main effects and one significant interaction term. We will proceed by removing non-significant main effects 1 by 1 in descending order of their distance from our significance threshold of 0.05.

## Remove main effect furthest from significance (heart.rate.circulation.issues):

```
reduced_modelF <- coxph(Surv(LOS, delta) ~ GENDER + cardiac.hemorrhage  
+ preterm*circulatory.system.malformation + respiratory + sepsis +  
jaundice, method=c("breslow"), data = data)  
summary(reduced_modelF)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ GENDER + cardiac.hemorrhage +
##        preterm * circulatory.system.malformation + respiratory +
##        sepsis + jaundice, data = data, method = c("breslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## GENDERM                                0.04285
1.04378
## cardiac.hemorrhage1                   -0.02612
0.97422
## pretermExtreme preterm                -1.55234
0.21175
## pretermPreterm                         -0.28251
0.75389
## circulatory.system.malformation1     -1.00317
0.36671
## respiratory1                           -0.33124
0.71803
## sepsis1                                -0.47172
0.62393
## jaundice1                             -0.06315
0.93881
## pretermExtreme preterm:circulatory.system.malformation1 1.02628
2.79066
## pretermPreterm:circulatory.system.malformation1          0.57861
1.78355
##
##                                         se(coef)
z
## GENDERM                                0.09714
0.441
## cardiac.hemorrhage1                   0.12058
-0.217
## pretermExtreme preterm                0.26548
-5.847
## pretermPreterm                         0.24167
-1.169
## circulatory.system.malformation1     0.33451

```

```

-2.999
## respiratory1 0.30910
-1.072
## sepsis1 0.11974
-3.940
## jaundice1 0.17400
-0.363
## pretermExtreme preterm:circulatory.system.malformation1 0.37114
2.765
## pretermPreterm:circulatory.system.malformation1 0.36095
1.603
## Pr
(>|z|)
## GENDERM 0.
65914
## cardiac.hemorrhage1 0.
82854
## pretermExtreme preterm 0.000000
00499 ***
## pretermPreterm 0.
24241
## circulatory.system.malformation1 0.
00271 **
## respiratory1 0.
28389
## sepsis1 0.000081
58168 ***
## jaundice1 0.
71667
## pretermExtreme preterm:circulatory.system.malformation1 0.
00569 **
## pretermPreterm:circulatory.system.malformation1 0.
10894
## ---
## Signif. codes: 0 '****' 0.001 '***' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coe
f)
## GENDERM 1.043
8

```

## cardiac.hemorragel	0.974
2	
## pretermExtreme preterm	0.211
8	
## pretermPreterm	0.753
9	
## circulatory.system.malformation1	0.366
7	
## respiratory1	0.718
0	
## sepsis1	0.623
9	
## jaundice1	0.938
8	
## pretermExtreme preterm:circulatory.system.malformation1	2.790
7	
## pretermPreterm:circulatory.system.malformation1	1.783
6	
## exp (-coe f)	
## GENDER.M	0.95
81	
## cardiac.hemorragel	1.02
65	
## pretermExtreme preterm	4.72
25	
## pretermPreterm	1.32
65	
## circulatory.system.malformation1	2.72
69	
## respiratory1	1.39
27	
## sepsis1	1.60
28	
## jaundice1	1.06
52	
## pretermExtreme preterm:circulatory.system.malformation1	0.35
83	
## pretermPreterm:circulatory.system.malformation1	0.56
07	

	lower .9	upper .9
##		
5		
## GENDERM	0.862	
8		
## cardiac.hemorrhage1	0.769	
2		
## pretermExtreme preterm	0.125	
9		
## pretermPreterm	0.469	
5		
## circulatory.system.malformation1	0.190	
4		
## respiratory1	0.391	
8		
## sepsis1	0.493	
4		
## jaundice1	0.667	
5		
## pretermExtreme preterm:circulatory.system.malformation1	1.348	
3		
## pretermPreterm:circulatory.system.malformation1	0.879	
1		
##		upper .9
5		
## GENDERM	1.262	
7		
## cardiac.hemorrhage1	1.233	
9		
## pretermExtreme preterm	0.356	
3		
## pretermPreterm	1.210	
6		
## circulatory.system.malformation1	0.706	
4		
## respiratory1	1.316	
0		
## sepsis1	0.789	
0		
## jaundice1	1.320	
3		

```

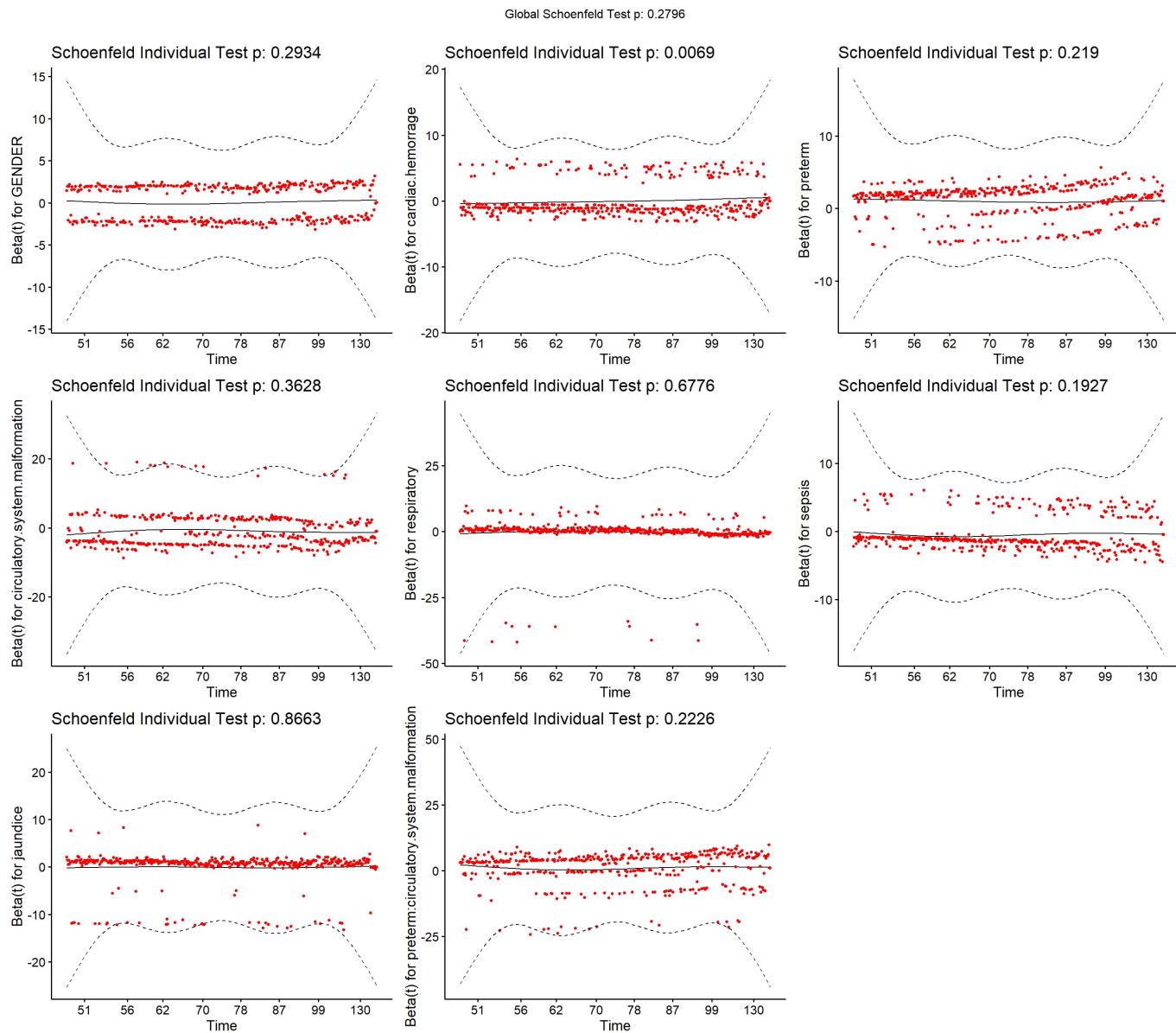
## pretermExtreme preterm:circulatory.system.malformation1      5.775
9
## pretermPreterm:circulatory.system.malformation1            3.618
5
##
## Concordance= 0.678  (se = 0.013 )
## Likelihood ratio test= 149  on 10 df,  p=<0.0000000000000002
## Wald test             = 141.7  on 10 df,  p=<0.0000000000000002
## Score (logrank) test = 154.7  on 10 df,  p=<0.0000000000000002

```

```

test7 <- cox.zph(reduced_modelF)
ggcoxzph(test7)

```



- Global tests show reduced model is significant

- Schoenfeld plot global p value shows model meets proportional hazards assumption.

```
anova(reduced_modelE, reduced_modelF)
```

	<b>loglik</b> <dbl>	<b>Chisq</b> <dbl>	<b>Df</b> <int>	<b>P(&gt; Chi )</b> <dbl>
1	-2224.789	NA	NA	NA
2	-2224.870	0.1611968	1	0.688057
2 rows				

- The larger model is not significantly better than the reduced model, and this main effect is not a confounder (< 10% change in other betas), thus we will remove this main effect and proceed.

## Remove main effect furthest from significance (cardiac hemorrhage):

```
reduced_modelG <- coxph(Surv(LOS, delta) ~ GENDER + preterm*circulatory.system.malformation + respiratory + sepsis + jaundice, method=c("breslow"), data = data)
summary(reduced_modelG)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ GENDER + preterm * circulator
y.system.malformation +
##      respiratory + sepsis + jaundice, data = data, method = c("br
eslow"))
##
##     n= 450, number of events= 449
##
##                                         coef
exp(coef)
## GENDERM                               0.04074
1.04158
## pretermExtreme preterm                -1.55244
0.21173
## pretermPreterm                         -0.28264
0.75379
## circulatory.system.malformation1    -1.00505
0.36603
## respiratory1                           -0.33283
0.71689
## sepsis1                                -0.47708
0.62059
## jaundice1                             -0.06025
0.94153
## pretermExtreme preterm:circulatory.system.malformation1  1.02660
2.79157
## pretermPreterm:circulatory.system.malformation1          0.57849
1.78334
##
##                                         se(coef)
z
## GENDERM                               0.09667
0.421
## pretermExtreme preterm                0.26544
-5.849
## pretermPreterm                         0.24161
-1.170
## circulatory.system.malformation1    0.33441
-3.005
## respiratory1                          0.30892
-1.077

```

```

## sepsis1                               0.11718
-4.071
## jaundice1                            0.17338
-0.348
## pretermExtreme preterm:circulatory.system.malformation1 0.37115
2.766
## pretermPreterm:circulatory.system.malformation1          0.36099
1.602
##                                         Pr
(> | z | )
## GENDERM                                0.
67341
## pretermExtreme preterm                  0.000000
00496 ***
## pretermPreterm                           0.
24206
## circulatory.system.malformation1       0.
00265 **
## respiratory1                            0.
28130
## sepsis1                                 0.000046
74066 ***
## jaundice1                             0.
72820
## pretermExtreme preterm:circulatory.system.malformation1 0.
00568 **
## pretermPreterm:circulatory.system.malformation1          0.
10905
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                         exp(coe
f)
## GENDERM                                1.041
6
## pretermExtreme preterm                  0.211
7
## pretermPreterm                           0.753
8
## circulatory.system.malformation1       0.366

```

0		
## respiratory1		0.716
9		
## sepsis1		0.620
6		
## jaundice1		0.941
5		
## pretermExtreme preterm:circulatory.system.malformation1		2.791
6		
## pretermPreterm:circulatory.system.malformation1		1.783
3		
##	exp (-coe	
f)		
## GENDERM		0.96
01		
## pretermExtreme preterm		4.72
30		
## pretermPreterm		1.32
66		
## circulatory.system.malformation1		2.73
20		
## respiratory1		1.39
49		
## sepsis1		1.61
14		
## jaundice1		1.06
21		
## pretermExtreme preterm:circulatory.system.malformation1		0.35
82		
## pretermPreterm:circulatory.system.malformation1		0.56
07		
##	lower .9	
5		
## GENDERM		0.861
8		
## pretermExtreme preterm		0.125
8		
## pretermPreterm		0.469
5		
## circulatory.system.malformation1		0.190

```

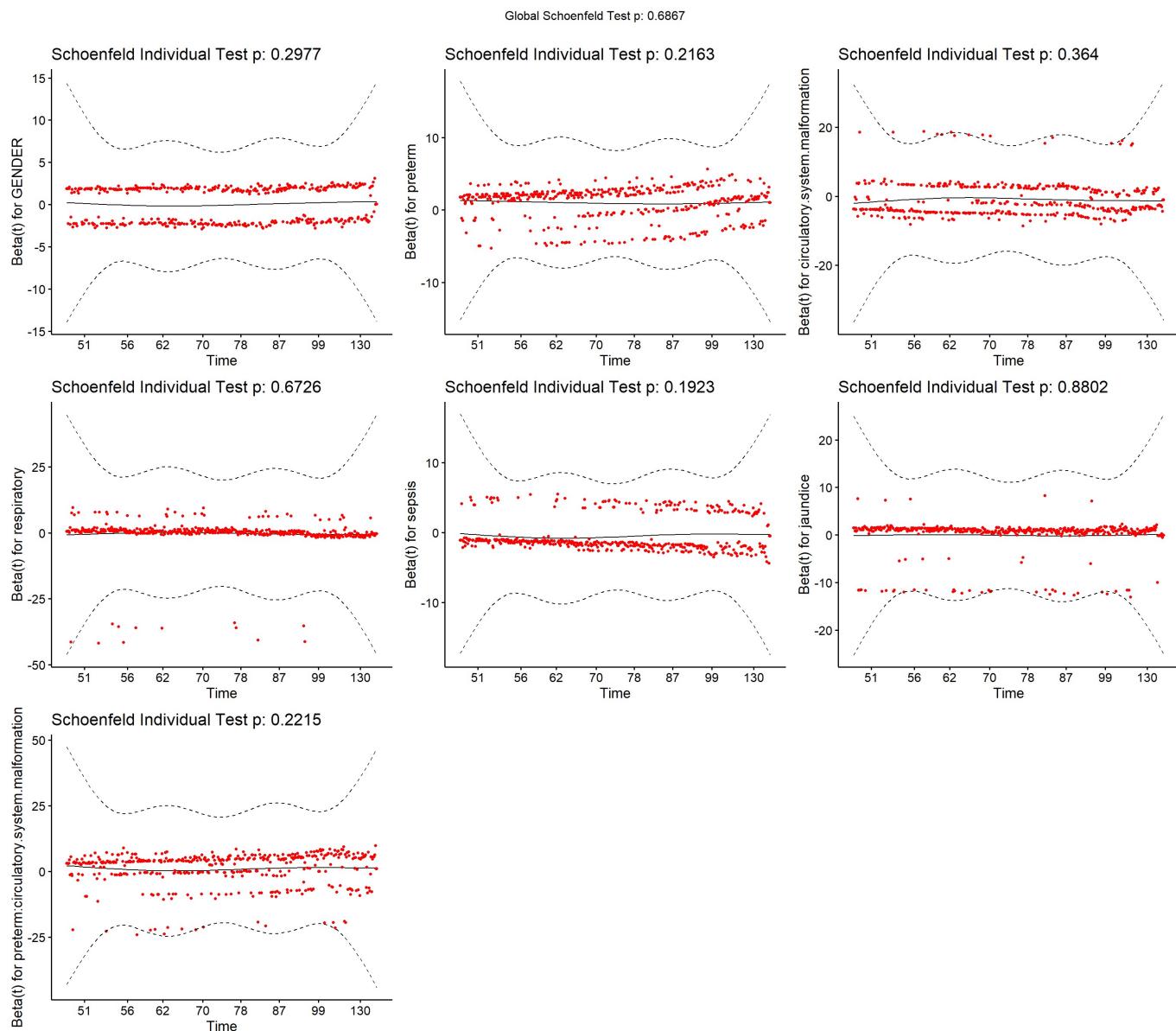
0
## respiratory1          0.391
3
## sepsis1              0.493
2
## jaundice1            0.670
3
## pretermExtreme preterm:circulatory.system.malformation1 1.348
7
## pretermPreterm:circulatory.system.malformation1         0.878
9
##                                         upper .9
5
## GENDERM               1.258
9
## pretermExtreme preterm           0.356
2
## pretermPreterm           1.210
3
## circulatory.system.malformation1 0.705
0
## respiratory1           1.313
4
## sepsis1                0.780
8
## jaundice1              1.322
6
## pretermExtreme preterm:circulatory.system.malformation1 5.778
0
## pretermPreterm:circulatory.system.malformation1         3.618
4
##
## Concordance= 0.678  (se = 0.013 )
## Likelihood ratio test= 148.9  on 9 df,    p=<0.0000000000000002
## Wald test             = 141.7  on 9 df,    p=<0.0000000000000002
## Score (logrank) test = 154.6  on 9 df,    p=<0.0000000000000002

```

```

test8 <- cox.zph(reduced_modelG)
ggcoxzph(test8)

```



- Global tests show reduced model is significant
- Schoenfeld plot p values shows model meets proportional hazards assumption.

```
anova(reduced_modelF, reduced_modelG)
```

	<b>loglik</b> <dbl>	<b>Chisq</b> <dbl>	<b>Df</b> <int>	<b>P(&gt; Chisq )</b> <dbl>
1	-2224.870	NA	NA	NA
2	-2224.893	0.04711173	1	0.8281675

2 rows

- The larger model is not significantly better than the reduced model, and this main

effect is not a confounder (< 10% change in other betas), thus we will remove this main effect and proceed.

## Remove main effect furthest from significance (jaundice):

```
reduced_modelH <- coxph(Surv(LOS, delta) ~ GENDER + preterm*circulatory.system.malformation + respiratory + sepsis, method=c("breslow"),  
data = data)  
summary(reduced_modelH)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ GENDER + preterm * circulator
y.system.malformation +
##      respiratory + sepsis, data = data, method = c("breslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## GENDERM                               0.04285
1.04378
## pretermExtreme preterm                -1.56419
0.20926
## pretermPreterm                         -0.29700
0.74304
## circulatory.system.malformation1    -1.01840
0.36117
## respiratory1                           -0.36042
0.69738
## sepsis1                                -0.47582
0.62137
## pretermExtreme preterm:circulatory.system.malformation1 1.03583
2.81743
## pretermPreterm:circulatory.system.malformation1          0.59475
1.81259
##                                         se(coef)
z
## GENDERM                               0.09649
0.444
## pretermExtreme preterm                0.26323
-5.942
## pretermPreterm                         0.23796
-1.248
## circulatory.system.malformation1    0.33218
-3.066
## respiratory1                          0.29849
-1.207
## sepsis1                                0.11711
-4.063
## pretermExtreme preterm:circulatory.system.malformation1 0.37021

```

2.798	
## pretermPreterm:circulatory.system.malformation1	0.35786
1.662	
##	Pr
(> z )	
## GENDERM	0.
65701	
## pretermExtreme preterm	0.000000
00281 ***	
## pretermPreterm	0.
21198	
## circulatory.system.malformation1	0.
00217 **	
## respiratory1	0.
22725	
## sepsis1	0.000048
46728 ***	
## pretermExtreme preterm:circulatory.system.malformation1	0.
00514 **	
## pretermPreterm:circulatory.system.malformation1	0.
09652 .	
## ---	
## Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	
##	
##	exp(coe
f)	
## GENDERM	1.043
8	
## pretermExtreme preterm	0.209
3	
## pretermPreterm	0.743
0	
## circulatory.system.malformation1	0.361
2	
## respiratory1	0.697
4	
## sepsis1	0.621
4	
## pretermExtreme preterm:circulatory.system.malformation1	2.817
4	

## pretermPreterm:circulatory.system.malformation1	1.812
6	
##	exp (-coe
f)	
## GENDERM	0.95
81	
## pretermExtreme preterm	4.77
88	
## pretermPreterm	1.34
58	
## circulatory.system.malformation1	2.76
88	
## respiratory1	1.43
39	
## sepsis1	1.60
93	
## pretermExtreme preterm:circulatory.system.malformation1	0.35
49	
## pretermPreterm:circulatory.system.malformation1	0.55
17	
##	lower .9
5	
## GENDERM	0.863
9	
## pretermExtreme preterm	0.124
9	
## pretermPreterm	0.466
1	
## circulatory.system.malformation1	0.188
3	
## respiratory1	0.388
5	
## sepsis1	0.493
9	
## pretermExtreme preterm:circulatory.system.malformation1	1.363
7	
## pretermPreterm:circulatory.system.malformation1	0.898
9	
##	upper .9
5	

```

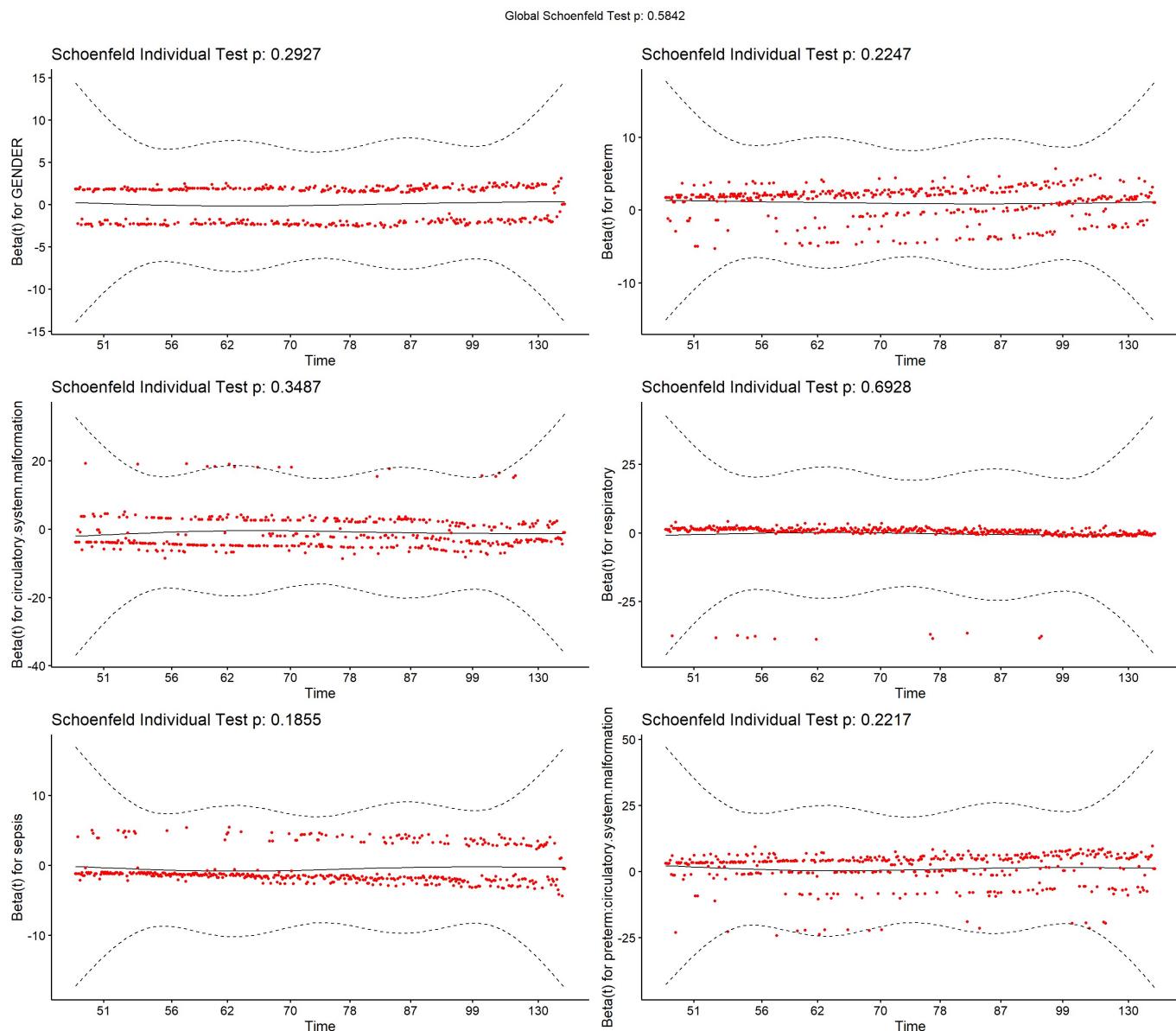
## GENDERM                                1.261
1
## pretermExtreme preterm                0.350
5
## pretermPreterm                         1.184
6
## circulatory.system.malformation1    0.692
6
## respiratory1                           1.251
8
## sepsis1                                0.781
7
## pretermExtreme preterm:circulatory.system.malformation1 5.820
8
## pretermPreterm:circulatory.system.malformation1      3.655
2
##
## Concordance= 0.678 (se = 0.013 )
## Likelihood ratio test= 148.8 on 8 df,   p=<0.0000000000000002
## Wald test                  = 141.6 on 8 df,   p=<0.0000000000000002
## Score (logrank) test = 154.5 on 8 df,   p=<0.0000000000000002

```

```

test9 <- cox.zph(reduced_modelH)
ggcoxzph(test9)

```



- Global tests show reduced model is significant
- Schoenfeld plot p values shows model meets proportional hazards assumption.

```
anova(reduced_modelG, reduced_modelH)
```

	loglik <dbl>	Chisq <dbl>	Df <int>	P(> Chisq ) <dbl>
1	-2224.893	NA	NA	NA
2	-2224.953	0.1191078	1	0.7300042

2 rows

- The larger model is not significantly better than the reduced model, and this main

effect is not a confounder (< 10% change in other betas), thus we will remove this main effect and proceed.

## Remove main effect furthest from significance (gender):

```
reduced_modelI <- coxph(Surv(LOS, delta) ~ preterm*circulatory.system.malformation + respiratory + sepsis, method=c("breslow"), data = data)
summary(reduced_modelI)
```

```
## Call:  
## coxph(formula = Surv(LOS, delta) ~ preterm * circulatory.system.  
malformation +  
##      respiratory + sepsis, data = data, method = c("breslow"))  
##  
##    n= 450, number of events= 449  
##  
##                                         coef  
exp(coef)  
## pretermExtreme preterm                         -1.5730  
0.2074  
## pretermPreterm                                -0.2987  
0.7418  
## circulatory.system.malformation1             -1.0132  
0.3630  
## respiratory1                                  -0.3546  
0.7015  
## sepsis1                                         -0.4740  
0.6225  
## pretermExtreme preterm:circulatory.system.malformation1 1.0367  
2.8199  
## pretermPreterm:circulatory.system.malformation1          0.5873  
1.7991  
##                                         se(coef)  
z  
## pretermExtreme preterm                         0.2625  
-5.993  
## pretermPreterm                                0.2379  
-1.255  
## circulatory.system.malformation1             0.3320  
-3.052  
## respiratory1                                 0.2982  
-1.189  
## sepsis1                                       0.1171  
-4.050  
## pretermExtreme preterm:circulatory.system.malformation1 0.3702  
2.800  
## pretermPreterm:circulatory.system.malformation1          0.3575  
1.643  
##                                         Pr
```

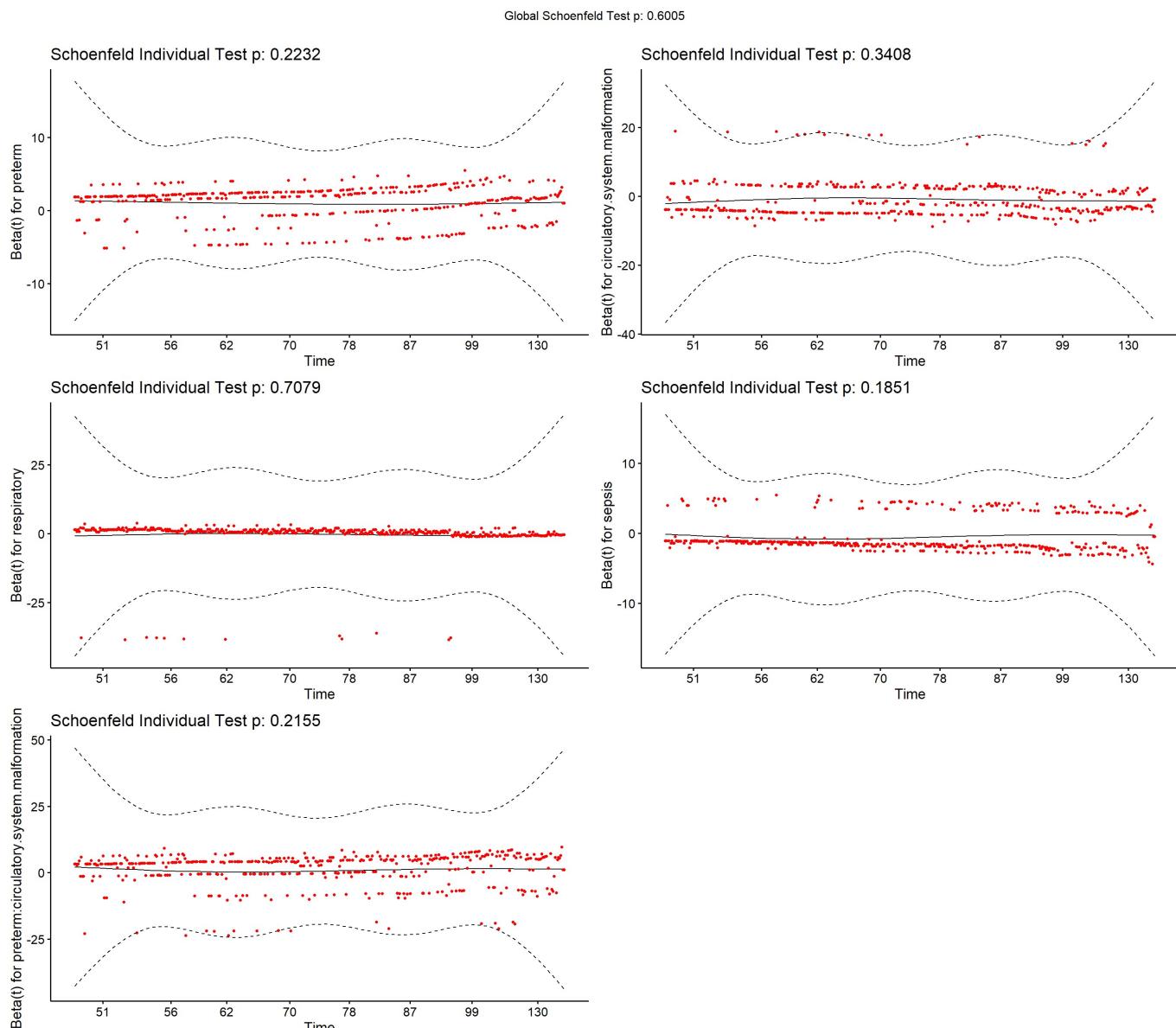
```
(> | z |)

## pretermExtreme preterm 0.000000
00206 ***
## pretermPreterm 0.
20933
## circulatory.system.malformation1 0.
00227 **
## respiratory1 0.
23437
## sepsis1 0.000051
23492 ***
## pretermExtreme preterm:circulatory.system.malformation1 0.
00510 **
## pretermPreterm:circulatory.system.malformation1 0.
10041
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coe
f)
## pretermExtreme preterm 0.207
4
## pretermPreterm 0.741
8
## circulatory.system.malformation1 0.363
0
## respiratory1 0.701
5
## sepsis1 0.622
5
## pretermExtreme preterm:circulatory.system.malformation1 2.819
9
## pretermPreterm:circulatory.system.malformation1 1.799
1
## exp(-coe
f)
## pretermExtreme preterm 4.82
09
## pretermPreterm 1.34
81
```

## circulatory.system.malformation1	2.75
45	
## respiratory1	1.42
56	
## sepsis1	1.60
65	
## pretermExtreme preterm:circulatory.system.malformation1	0.35
46	
## pretermPreterm:circulatory.system.malformation1	0.55
58	
##	lower .9
5	
## pretermExtreme preterm	0.124
0	
## pretermPreterm	0.465
3	
## circulatory.system.malformation1	0.189
4	
## respiratory1	0.391
0	
## sepsis1	0.494
9	
## pretermExtreme preterm:circulatory.system.malformation1	1.364
9	
## pretermPreterm:circulatory.system.malformation1	0.892
8	
##	upper .9
5	
## pretermExtreme preterm	0.347
0	
## pretermPreterm	1.182
5	
## circulatory.system.malformation1	0.695
9	
## respiratory1	1.258
4	
## sepsis1	0.783
0	
## pretermExtreme preterm:circulatory.system.malformation1	5.825
7	

```
## pretermPreterm:circulatory.system.malformation1 3.625
1
##
## Concordance= 0.676 (se = 0.013 )
## Likelihood ratio test= 148.6 on 7 df, p=<0.0000000000000002
## Wald test = 141.5 on 7 df, p=<0.0000000000000002
## Score (logrank) test = 154.4 on 7 df, p=<0.0000000000000002
```

```
test10 <- cox.zph(reduced_modelI)
ggcoxzph(test10)
```



- Global tests show reduced model is significant
- Schoenfeld plot p values shows model meets proportional hazards assumption.

```
anova(reduced_modelH, reduced_modelI)
```

	<b>loglik</b> <dbl>	<b>Chisq</b> <dbl>	<b>Df</b> <int>	<b>P(&gt; Chi )</b> <dbl>
1	-2224.953	NA	NA	NA
2	-2225.051	0.1974027	1	0.6568255
2 rows				

- The larger model is not significantly better than the reduced model, and this main effect is not a confounder (< 10% change in other betas), thus we will remove this main effect and proceed.

## Remove main effect furthest from significance (respiratory):

```
reduced_modelJ <- coxph(Surv(LOS, delta) ~ preterm*circulatory.system.malformation + sepsis, method=c("breslow"), data = data)
summary(reduced_modelJ)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ preterm * circulatory.system.
malformation +
##       sepsis, data = data, method = c("breslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## pretermExtreme preterm                         -1.5991
0.2021
## pretermPreterm                                -0.3296
0.7192
## circulatory.system.malformation1            -1.0228
0.3596
## sepsis1                                         -0.4761
0.6212
## pretermExtreme preterm:circulatory.system.malformation1 1.0409
2.8317
## pretermPreterm:circulatory.system.malformation1          0.6074
1.8357
##
##                                         se(coef)
z
## pretermExtreme preterm                         0.2610
-6.128
## pretermPreterm                                0.2358
-1.398
## circulatory.system.malformation1            0.3317
-3.083
## sepsis1                                         0.1169
-4.071
## pretermExtreme preterm:circulatory.system.malformation1 0.3699
2.814
## pretermPreterm:circulatory.system.malformation1          0.3568
1.702
##
##                                         Pr
(>| z | )
## pretermExtreme preterm                         0.000000
000892 ***
## pretermPreterm

```

```
0.16208
## circulatory.system.malformation1
0.00205 **
## sepsis1                                         0.000046
725442 ***
## pretermExtreme preterm:circulatory.system.malformation1
0.00490 **
## pretermPreterm:circulatory.system.malformation1
0.08866 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coe
f)
## pretermExtreme preterm                           0.202
1
## pretermPreterm                                 0.719
2
## circulatory.system.malformation1             0.359
6
## sepsis1                                       0.621
2
## pretermExtreme preterm:circulatory.system.malformation1   2.831
7
## pretermPreterm:circulatory.system.malformation1           1.835
7
## exp(-coe
f)
## pretermExtreme preterm                           4.94
88
## pretermPreterm                                 1.39
04
## circulatory.system.malformation1             2.78
10
## sepsis1                                       1.60
97
## pretermExtreme preterm:circulatory.system.malformation1   0.35
31
## pretermPreterm:circulatory.system.malformation1           0.54
48
```

```

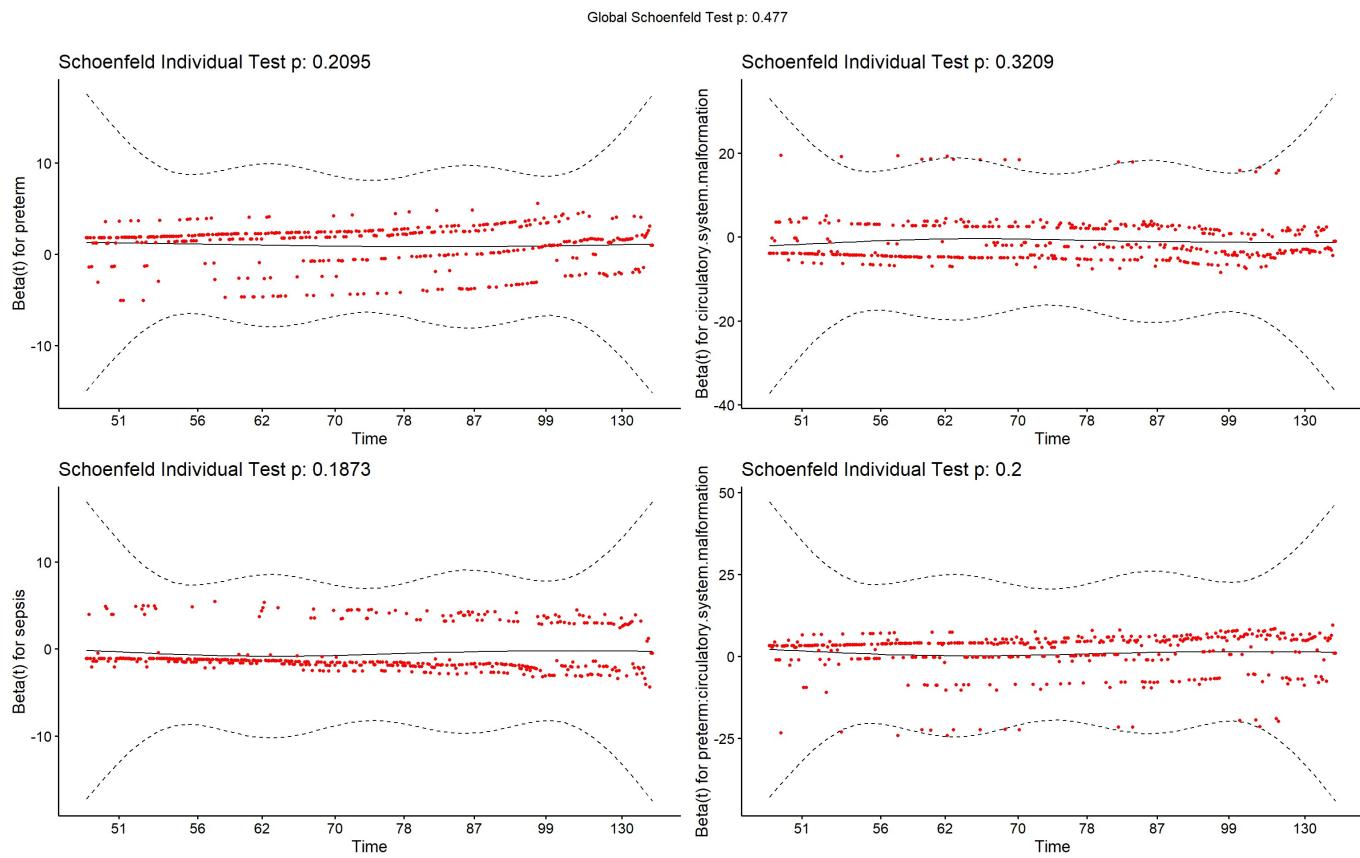
##                                         lower .9
5
## pretermExtreme preterm          0.121
2
## pretermPreterm           0.453
1
## circulatory.system.malformation1 0.187
7
## sepsis1                  0.494
0
## pretermExtreme preterm:circulatory.system.malformation1 1.371
4
## pretermPreterm:circulatory.system.malformation1      0.912
2
##                                         upper .9
5
## pretermExtreme preterm          0.337
0
## pretermPreterm           1.141
6
## circulatory.system.malformation1 0.688
9
## sepsis1                  0.781
2
## pretermExtreme preterm:circulatory.system.malformation1 5.847
1
## pretermPreterm:circulatory.system.malformation1      3.694
0
##
## Concordance= 0.676  (se = 0.013 )
## Likelihood ratio test= 147.3  on 6 df,    p=<0.0000000000000002
## Wald test              = 140.3  on 6 df,    p=<0.0000000000000002
## Score (logrank) test = 152.9  on 6 df,    p=<0.0000000000000002

```

```

test11 <- cox.zph(reduced_modelJ)
ggcoxzph(test11)

```



- Global tests show reduced model is significant
- Schoenfeld plot p values shows model meets proportional hazards assumption.

```
anova(reduced_modelI, reduced_modelJ)
```

	loglik <dbl>	Chisq <dbl>	Df <int>	P(> Chisq ) <dbl>
1	-2225.051	NA	NA	NA
2	-2225.690	1.27695	1	0.2584669
2 rows				

- The larger model is not significantly better than the reduced model, and this main effect is not a confounder (< 10% change in other betas), thus we will remove this main effect and proceed.
- Now all main effects and all interaction terms remaining in our model are significant. Thus we have our final model:

## FINAL MODEL:

```
final_model <-coxph(Surv(LOS, delta) ~ preterm*circulatory.system.ma  
lformation + sepsis, method=c("breslow"), data = data)  
summary(final_model)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ preterm * circulatory.system.
malformation +
##       sepsis, data = data, method = c("breslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## pretermExtreme preterm                         -1.5991
0.2021
## pretermPreterm                                -0.3296
0.7192
## circulatory.system.malformation1            -1.0228
0.3596
## sepsis1                                         -0.4761
0.6212
## pretermExtreme preterm:circulatory.system.malformation1 1.0409
2.8317
## pretermPreterm:circulatory.system.malformation1          0.6074
1.8357
##
##                                         se(coef)
z
## pretermExtreme preterm                         0.2610
-6.128
## pretermPreterm                                0.2358
-1.398
## circulatory.system.malformation1            0.3317
-3.083
## sepsis1                                         0.1169
-4.071
## pretermExtreme preterm:circulatory.system.malformation1 0.3699
2.814
## pretermPreterm:circulatory.system.malformation1          0.3568
1.702
##
##                                         Pr
(>| z | )
## pretermExtreme preterm                         0.000000
000892 ***
## pretermPreterm

```

```
0.16208
## circulatory.system.malformation1
0.00205 **
## sepsis1                                         0.000046
725442 ***
## pretermExtreme preterm:circulatory.system.malformation1
0.00490 **
## pretermPreterm:circulatory.system.malformation1
0.08866 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coe
f)
## pretermExtreme preterm                           0.202
1
## pretermPreterm                                  0.719
2
## circulatory.system.malformation1               0.359
6
## sepsis1                                         0.621
2
## pretermExtreme preterm:circulatory.system.malformation1   2.831
7
## pretermPreterm:circulatory.system.malformation1           1.835
7
## exp(-coe
f)
## pretermExtreme preterm                           4.94
88
## pretermPreterm                                  1.39
04
## circulatory.system.malformation1               2.78
10
## sepsis1                                         1.60
97
## pretermExtreme preterm:circulatory.system.malformation1   0.35
31
## pretermPreterm:circulatory.system.malformation1           0.54
48
```

```

##                                         lower .9
5
## pretermExtreme preterm               0.121
2
## pretermPreterm                      0.453
1
## circulatory.system.malformation1   0.187
7
## sepsis1                            0.494
0
## pretermExtreme preterm:circulatory.system.malformation1 1.371
4
## pretermPreterm:circulatory.system.malformation1        0.912
2
##                                         upper .9
5
## pretermExtreme preterm              0.337
0
## pretermPreterm                      1.141
6
## circulatory.system.malformation1   0.688
9
## sepsis1                            0.781
2
## pretermExtreme preterm:circulatory.system.malformation1 5.847
1
## pretermPreterm:circulatory.system.malformation1        3.694
0
##
## Concordance= 0.676  (se = 0.013 )
## Likelihood ratio test= 147.3  on 6 df,    p=<0.0000000000000002
## Wald test                  = 140.3  on 6 df,    p=<0.0000000000000002
## Score (logrank) test = 152.9  on 6 df,    p=<0.0000000000000002

```

**Double check that this interaction improves our model over just a model with these main effects:**

```
final_noInt <-coxph(Surv(LOS, delta) ~ preterm+circulatory.system.ma  
lformation + sepsis, method=c("breslow"), data = data)
```

```
anova(final_model, final_noInt)
```

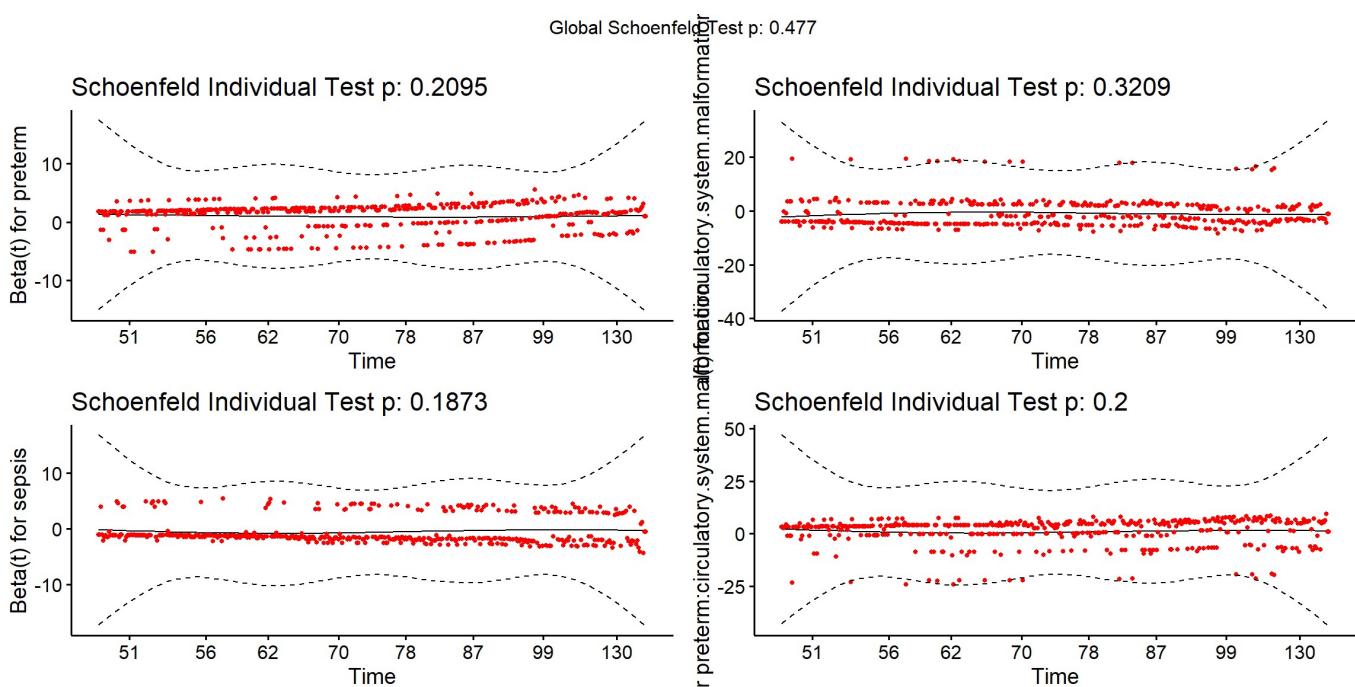
	<b>loglik</b> <dbl>	<b>Chisq</b> <dbl>	<b>Df</b> <int>	<b>P(&gt; Chi )</b> <dbl>
1	-2225.690	NA	NA	NA
2	-2230.412	9.444	2	0.008897364

2 rows

- The model with the interaction included is significantly better. Thus, our final model remains as it is.

Test proportional hazards assumption for final model with schoenfeld test:

```
test12 <- cox.zph(final_model)
ggcoxzph(test12)
```



- Schoenfeld plot p values shows that our final model meets proportional hazards assumption.

## Final Cox Proportional Hazards Model

```
summary(final_model)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ preterm * circulatory.system.
malformation +
##       sepsis, data = data, method = c("breslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## pretermExtreme preterm                         -1.5991
0.2021
## pretermPreterm                                -0.3296
0.7192
## circulatory.system.malformation1            -1.0228
0.3596
## sepsis1                                         -0.4761
0.6212
## pretermExtreme preterm:circulatory.system.malformation1 1.0409
2.8317
## pretermPreterm:circulatory.system.malformation1          0.6074
1.8357
##
##                                         se(coef)
z
## pretermExtreme preterm                         0.2610
-6.128
## pretermPreterm                                0.2358
-1.398
## circulatory.system.malformation1            0.3317
-3.083
## sepsis1                                         0.1169
-4.071
## pretermExtreme preterm:circulatory.system.malformation1 0.3699
2.814
## pretermPreterm:circulatory.system.malformation1          0.3568
1.702
##
##                                         Pr
(>| z | )
## pretermExtreme preterm                         0.000000
000892 ***
## pretermPreterm

```

```
0.16208
## circulatory.system.malformation1
0.00205 **
## sepsis1                                         0.000046
725442 ***
## pretermExtreme preterm:circulatory.system.malformation1
0.00490 **
## pretermPreterm:circulatory.system.malformation1
0.08866 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coe
f)
## pretermExtreme preterm                           0.202
1
## pretermPreterm                                  0.719
2
## circulatory.system.malformation1               0.359
6
## sepsis1                                         0.621
2
## pretermExtreme preterm:circulatory.system.malformation1   2.831
7
## pretermPreterm:circulatory.system.malformation1           1.835
7
## exp(-coe
f)
## pretermExtreme preterm                           4.94
88
## pretermPreterm                                  1.39
04
## circulatory.system.malformation1               2.78
10
## sepsis1                                         1.60
97
## pretermExtreme preterm:circulatory.system.malformation1   0.35
31
## pretermPreterm:circulatory.system.malformation1           0.54
48
```

```

##                                         lower .9
5
## pretermExtreme preterm                         0.121
2
## pretermPreterm                           0.453
1
## circulatory.system.malformation1          0.187
7
## sepsis1                                 0.494
0
## pretermExtreme preterm:circulatory.system.malformation1 1.371
4
## pretermPreterm:circulatory.system.malformation1        0.912
2
##                                         upper .9
5
## pretermExtreme preterm                      0.337
0
## pretermPreterm                           1.141
6
## circulatory.system.malformation1          0.688
9
## sepsis1                                 0.781
2
## pretermExtreme preterm:circulatory.system.malformation1 5.847
1
## pretermPreterm:circulatory.system.malformation1        3.694
0
##
## Concordance= 0.676  (se = 0.013 )
## Likelihood ratio test= 147.3  on 6 df,    p=<0.0000000000000002
## Wald test              = 140.3  on 6 df,    p=<0.0000000000000002
## Score (logrank) test = 152.9  on 6 df,    p=<0.0000000000000002

```

## Addressing Aim 2): Tests for proportional differences between preterm levels in having each category of diagnosis

Based on me coding medical conditions as 0 for absence, and 1 for presence, the 0 column appears as the first column in the table and is read as the success column. Thus, the proportions presented in the prop tests below should be read as the proportion of infants in each group who DO NOT suffer from that condition. Additionally, based on the leveling of the three preterm labels; proportion 1 = Full-term, proportion 2 = Extreme preterm, and proportion 3 = Preterm.

## Prop test for difference in prevalence of cardiac hemorrhage:

```
prop.test(cardtab, alternative = c("two.sided"))
```

```
## 
## 3-sample test for equality of proportions without continuity
## correction
##
## data: cardtab
## X-squared = 4.6061, df = 2, p-value = 0.09995
## alternative hypothesis: two.sided
## sample estimates:
##      prop 1      prop 2      prop 3
## 0.8421053 0.7368421 0.8174274
```

- Not a significant difference between groups.

## Prop test for difference in prevalence of circulatory malformations:

```
prop.test(circtab, alternative = c("two.sided"))
```

```

## 
## 3-sample test for equality of proportions without continuity
## correction
## 
## data: circtab
## X-squared = 20.628, df = 2, p-value = 0.00003316
## alternative hypothesis: two.sided
## sample estimates:
##      prop 1      prop 2      prop 3
## 0.5526316 0.3801170 0.6058091

```

- Significant difference in prevalence between groups ; we will do post-hoc pairwise comparisons to see which groups are different:

```

pairwise.prop.test(circtab, alternative = c("two.sided"), p.adjust.method = 'bonferroni') #adjusts for multiple comparisons

```

```

## 
## Pairwise comparisons using Pairwise comparison of proportions
## 
## data: circtab
## 
##          Full term Extreme preterm
## Extreme preterm 0.23      -
## Preterm         1.00      0.00003
## 
## P value adjustment method: bonferroni

```

- We can see that the significant difference was between preterm and extreme preterm, where extreme preterm had a higher prevalence of circulatory system malformations (lower proportion of healthy kids)

## Prop test for difference in prevalence of heart rate anomalies:

```

prop.test(hrtab, alternative = c("two.sided"))

```

```
## 
## 3-sample test for equality of proportions without continuity
## correction
## 
## data: hrtab
## X-squared = 9.6389, df = 2, p-value = 0.008071
## alternative hypothesis: two.sided
## sample estimates:
##      prop 1      prop 2      prop 3
## 0.5526316 0.3040936 0.3070539
```

- Significant difference in prevalence between groups ; we will do post-hoc pairwise comparisons to see which groups are different:

```
pairwise.prop.test(hrtab, alternative = c("two.sided"), p.adjust.method = 'bonferroni') #adjusts for multiple comparisons
```

```
## 
## Pairwise comparisons using Pairwise comparison of proportions
## 
## data: hrtab
## 
##          Full term Extreme preterm
## Extreme preterm 0.020      -
## Preterm         0.016      1.000
## 
## P value adjustment method: bonferroni
```

- We can see that there were significant differences between full-term and preterm, and full-term and extreme preterm, where full-term had a lower prevalence of heart rate anomalies than the other two groups (higher proportion of healthy kids)

## Prop test for difference in prevalence of respiratory distress:

```
prop.test(resptab, alternative = c("two.sided"))
```

```

## 
## 3-sample test for equality of proportions without continuity
## correction
## 
## data: resptab
## X-squared = 5.5307, df = 2, p-value = 0.06295
## alternative hypothesis: two.sided
## sample estimates:
##      prop 1      prop 2      prop 3
## 0.07894737 0.01169591 0.02904564

```

- Not a significant difference between groups. As this was close to significant, I inspected with pairwise tests, but all pairwise tests were non-significant too.

## Prop test for difference in prevalence of sepsis:

```
prop.test(septab, alternative = c("two.sided"))
```

```

## 
## 3-sample test for equality of proportions without continuity
## correction
## 
## data: septab
## X-squared = 29.123, df = 2, p-value = 0.0000004743
## alternative hypothesis: two.sided
## sample estimates:
##      prop 1      prop 2      prop 3
## 0.7894737 0.6257310 0.8547718

```

- Significant difference in prevalence between groups ; we will do post-hoc pairwise comparisons to see which groups are different:

```
pairwise.prop.test(septab, alternative = c("two.sided"), p.adjust.method = 'bonferroni') #adjusts for multiple comparisons
```

```
## 
##  Pairwise comparisons using Pairwise comparison of proportions
## 
##  data:  septab
## 
##          Full term Extreme preterm
## Extreme preterm 0.25      -
## Preterm         1.00      0.00000047
## 
## P value adjustment method: bonferroni
```

- We can see that the significant difference was between preterm and extreme preterm, where extreme preterm had a higher prevalence of sepsis (lower proportion of healthy kids)

## Prop test for difference in prevalence of jaundice:

```
prop.test(jauntab, alternative = c("two.sided"))
```

```
## 
##  3-sample test for equality of proportions without continuity
##  correction
## 
##  data:  jauntab
##  X-squared = 19.84, df = 2, p-value = 0.00004919
##  alternative hypothesis: two.sided
##  sample estimates:
##    prop 1    prop 2    prop 3
##  0.28947368 0.05847953 0.08713693
```

- Significant difference in prevalence between groups ; we will do post-hoc pairwise comparisons to see which groups are different:

```
pairwise.prop.test(jauntab, alternative = c("two.sided"), p.adjust.method = 'bonferroni') #adjusts for multiple comparisons
```

```
##  
##  Pairwise comparisons using Pairwise comparison of proportions  
##  
## data: jauntab  
##  
##           Full term Extreme preterm  
## Extreme preterm 0.0002 -  
## Preterm        0.0023 1.0000  
##  
## P value adjustment method: bonferroni
```

- We can see that there were significant differences between full-term and preterm, and full-term and extreme preterm, where full-term had a lower prevalence of jaundice than the other two groups (higher proportion of healthy kids)

## Addressing Aim 3): Multinomial Modelling of Discharge Locations

To examine the influence of diagnoses on the initial location to which neonate patients will be discharged, multinomial modelling conducted using the VGAM package (Yee, 2019; Yee, Stoklosa, & Huggins, 2015), with the same categorical variables used in the above analysis, though delta was removed.

```

dat = read.csv("datadfpretermNew.csv")

dat$HADM_ID = factor(dat$HADM_ID)
dat$cardiac.hemorrhage = factor(dat$cardiac.hemorrhage)
dat$circulatory.system.malformation = factor(dat$circulatory.system.malformation)
dat$heart.rate.circulation.issues = factor(dat$heart.rate.circulation.issues)
dat$preterm = factor(dat$preterm) #0 means not preterm, 1 means preterm, and 2 is extreme preterm.
dat$respiratory = factor(dat$respiratory)
dat$jaundice = factor(dat$jaundice)
dat$sepsis = factor(dat$sepsis)

```

```
table(dat$DISCHARGE_LOCATION)
```

	DEAD/EXPIRED	DISC-TRAN CANCER/CHLDRN H
##	1	1
##	HOME	HOME HEALTH CARE
##	103	284
##	LONG TERM CARE HOSPITAL	SHORT TERM HOSPITAL
##	1	60

Because there was only one individual in each of dead, discharge transfer to cancer or childrens' hospital, and long term care hospital categories, we needed to remove those patients as the groups were too small to make comparisons nor could we meaningfully combine them.

```
dat<-dat[!(dat$DISCHARGE_LOCATION=="DEAD/EXPIRED" | dat$DISCHARGE_LOCATION=="DISC-TRAN CANCER/CHLDRN H" | dat$DISCHARGE_LOCATION=="LONG TERM CARE HOSPITAL"),]
```

```
(multi.wvgam <- vglm(formula = DISCHARGE_LOCATION ~ cardiac.hemorrhage + circulatory.system.malformation +
  heart.rate.circulation.issues + preterm +
  respiratory + sepsis + jaundice + GENDER,
  data=dat,
  family=multinomial))
```

```

## 
## Call:
## vglm(formula = DISCHARGE_LOCATION ~ cardiac.hemorrhage + circulatory.system.malformation +
##       heart.rate.circulation.issues + preterm + respiratory + sepsis +
##       jaundice + GENDER, family = multinomial, data = dat)
## 
## 
## Coefficients:
##                               (Intercept) :1          (Intercept) :2
##                               16.24550526           15.25057627
##       cardiac.hemorrhage1:1          cardiac.hemorrhage1:2
##                               -0.40511634           -0.11294309
##       circulatory.system.malformation1:1 circulatory.system.malformation1:2
##                               -0.45281023            0.05448176
##       heart.rate.circulation.issues1:1   heart.rate.circulation.issues1:2
##                               -0.02706367            0.20760301
##       preterm1:1                      preterm1:2
##                               0.47919789           -0.03099562
##       preterm2:1                      preterm2:2
##                               -1.10957320           -1.67576661
##       respiratory1:1                  respiratory1:2
##                               -15.44533265          -13.78155105
##       sepsis1:1                      sepsis1:2
##
```

```
## -0.14096245 0.800
53480 jaundice1:1 jaundi
ce1:2
## 0.33656623 0.878
07209 GENDER:1 GEND
ERM:2
## 0.01579849 -0.358
57510
##
## Degrees of Freedom: 894 Total; 874 Residual
## Residual deviance: 732.1778
## Log-likelihood: -366.0889
##
## This is a multinomial logit model with 3 levels
```

```
summary(multi.wvgam)
```

```

## 
## Call:
## vglm(formula = DISCHARGE_LOCATION ~ cardiac.hemorrhage + circulatory.system.malformation +
##       heart.rate.circulation.issues + preterm + respiratory + sepsis +
##       jaundice + GENDER, family = multinomial, data = dat)
##
## Pearson residuals:
##                               Min      1Q   Median      3Q     Max
## log(mu[,1]/mu[,3]) -2.870 -0.3408 -0.2409 -0.1096 3.854
## log(mu[,2]/mu[,3]) -4.213 -0.7095  0.5001  0.6768 1.610
##
## Coefficients:
##                                     Estimate Std. Error z value
Pr(>|z|)
## (Intercept):1                  16.24551  655.30486      NA
NA
## (Intercept):2                  15.25058  655.30492      0.023
0.9814
## cardiac.hemorrhage1:1          -0.40512    0.43261     -0.936
0.3490
## cardiac.hemorrhage1:2          -0.11294    0.35553     -0.318
0.7507
## circulatory.system.malformation1:1 -0.45281    0.35616     -1.271
0.2036
## circulatory.system.malformation1:2  0.05448    0.31027      0.176
0.8606
## heart.rate.circulation.issues1:1 -0.02706    0.36301     -0.075
0.9406
## heart.rate.circulation.issues1:2  0.20760    0.32226      0.644
0.5194
## preterm1:1                      0.47920    0.76779      0.624
0.5325
## preterm1:2                      -0.03100    0.69251     -0.045
0.9643
## preterm2:1                      -1.10957    0.76090     -1.458
0.1448
## preterm2:2                      -1.67577    0.67736     -2.474
0.0134 *

```

```

## respiratory1:1           -15.44533  655.30468 -0.024
0.9812
## respiratory1:2           -13.78155  655.30477 -0.021
0.9832
## sepsis1:1                -0.14096   0.47385 -0.297
0.7661
## sepsis1:2                 0.80053   0.37293  2.147
0.0318 *
## jaundice1:1              0.33657   0.55997  0.601
0.5478
## jaundice1:2              0.87807   0.51636  1.701
0.0890 .
## GENDERM:1                0.01580   0.34933  0.045
0.9639
## GENDERM:2                -0.35858  0.30367 -1.181
0.2377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 732.1778 on 874 degrees of freedom
##
## Log-likelihood: -366.0889 on 874 degrees of freedom
##
## Number of Fisher scoring iterations: 15
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1'
##
##
## Reference group is level 3 of the response

```

## Testing for an initial interaction for prematurity

As there is likely a clinical interplay between how premature a new born is and the complexity of their case, we felt that it was prudent to investigate interaction models before eliminating factors.

```
anova(multi.wvgam, update(multi.wvgam, .~.+preterm * cardiac.hemorrag  
e), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	870	727.5762	4	4.601643	0.3306647
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.+preterm * circulatory.syst  
em.malformation), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	870	727.1378	4	5.04005	0.2832128
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.+preterm * heart.rate.circu  
lation.issues), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	870	726.6451	4	5.532784	0.2368627
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.+preterm * respiratory), typ  
e=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	870	726.1352	4	6.042649	0.1959858
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.+preterm * sepsis), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	870	727.1033	4	5.074508	0.2797376
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.+preterm * jaundice), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	870	726.3874	4	5.790435	0.2153549
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.+preterm * GENDER), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	870	725.0668	4	7.111013	0.1301367

2 rows

However, probing these interactions did not yield findings that would give us cause to choose an interaction model. This is to say that we would fail to reject the null hypothesis of all of the above likelihood ratio tests ( $p>0.05$ ) and conclude that the larger interaction models do not account for additional variability in the data.

## Assumptions of Multinomial models

The multinomial logistic regression assumes:

- Case specificity (that each participant is in the data set once)
- collinearity is relatively low
- the independence of irrelevant alternatives

However, these tests are not accessible for the types of data being analysed in this study. Therefore, we will proceed with satisfying the first assumption: all rows of data were retrieved from unique patient identifiers in the MIMIC study.

## Making a more parsimonious model

Above, we see variables that do not effectively model the data, so we will perform a likelihood ratio test when each of the above variables has been removed to determine which should be taken out of the model. From the findings of the Wald Z tests for the main effects model shown above, we would not expect sepsis or prematurity to indicate that we should reject the null hypothesis of the likelihood ratio test, as we would already fail to reject the null hypothesis of the Wald tests ( $p<0.05$ , for at least one level of the factor)

```
anova(multi.wvgam, update(multi.wvgam, .~.-cardiac.hemorrhage), type=1)
```

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	874	732.1778	NA	NA
2	876	733.2637	-2	-1.085844

2 rows

```
anova(multi.wvgam, update(multi.wvgam, .~.-circulatory.system.malformation), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	876	736.3465	-2	-4.168625	0.1243926
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-heart.rate.circulation.issues), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	876	733.2240	-2	-1.046172	0.5926888
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-respiratory), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	876	741.4737	-2	-9.295846	0.009581482
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-GENDER), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	876	735.3283	-2	-3.150491	0.2069567
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-jaundice), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	874	732.1778	NA	NA	NA
2	876	735.6853	-2	-3.507445	0.1731283
2 rows					

Based on these findings, we cannot remove the respiratory variables ( $\chi^2(2)=9.296$ ,  $p = 0.0096$ ). Therefore, we will remove gender: it is less clinically important to the research question we asked regarding the possible link between the reason for admission and the discharge location, and statistically, we are correct to remove it.

## Removing Gender from the Model

```
multi.wvgam = update(multi.wvgam, .~.-GENDER)
summary(multi.wvgam)
```

```

## 
## Call:
## vglm(formula = DISCHARGE_LOCATION ~ cardiac.hemorrhage + circulatory.system.malformation +
##       heart.rate.circulation.issues + preterm + respiratory + sepsis +
##       jaundice, family = multinomial, data = dat)
##
## Pearson residuals:
##                               Min      1Q   Median      3Q     Max
## log(mu[,1]/mu[,3]) -2.638 -0.3740 -0.2321 -0.1017 3.537
## log(mu[,2]/mu[,3]) -3.784 -0.7816  0.5519  0.6593 1.442
##
## Coefficients:
##                                     Estimate Std. Error z value
Pr(>|z|)
## (Intercept):1                  16.26718  653.47160      NA
NA
## (Intercept):2                  15.10100  653.47166      0.023
0.9816
## cardiac.hemorrhage1:1          -0.38361   0.43130    -0.889
0.3738
## cardiac.hemorrhage1:2          -0.12661   0.35490    -0.357
0.7213
## circulatory.system.malformation1:1 -0.46626   0.35469    -1.315
0.1887
## circulatory.system.malformation1:2  0.03648   0.30845     0.118
0.9059
## heart.rate.circulation.issues1:1 -0.01253   0.36083    -0.035
0.9723
## heart.rate.circulation.issues1:2  0.25289   0.31923     0.792
0.4283
## preterm1:1                      0.45985   0.76826     0.599
0.5495
## preterm1:2                      -0.04662   0.69307    -0.067
0.9464
## preterm2:1                      -1.13080   0.76171    -1.485
0.1377
## preterm2:2                      -1.65940   0.67800    -2.447
0.0144 *

```

```

## respiratory1:1           -15.44063   653.47141  -0.024
0.9811
## respiratory1:2           -13.85888   653.47150  -0.021
0.9831
## sepsis1:1                -0.15061    0.47182  -0.319
0.7496
## sepsis1:2                 0.78267    0.37071   2.111
0.0347 *
## jaundice1:1              0.33219    0.55944   0.594
0.5527
## jaundice1:2              0.89486    0.51465   1.739
0.0821 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 735.3283 on 876 degrees of freedom
##
## Log-likelihood: -367.6642 on 876 degrees of freedom
##
## Number of Fisher scoring iterations: 15
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1'
##
##
## Reference group is level 3 of the response

```

Again, from the Wald tests, sepsis and prematurity (and potentially jaundice) seem most promising, so we will not remove them, but we must investigate other factors to make a more parsimonious models.

```
anova(multi.wvgam, update(multi.wvgam, .~.-cardiac.hemorrhage), type=1)
```

<b>Resid. Df</b>	<b>Resid. Dev</b>	<b>Df</b>	<b>Deviance</b>	<b>Pr(&gt;Chi)</b>
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	876	735.3283	NA	NA	NA
2	878	736.2496	-2	-0.9212618	0.6308855
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-circulatory.system.malformation), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	876	735.3283	NA	NA	NA
2	878	739.4880	-2	-4.159716	0.1249479
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-heart.rate.circulation.issues), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	876	735.3283	NA	NA	NA
2	878	736.7493	-2	-1.420991	0.4914006
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-respiratory), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	876	735.3283	NA	NA	NA
2	878	744.2807	-2	-8.952346	0.01137687

2 rows

```
anova(multi.wvgam, update(multi.wvgam, .~.-jaundice), type=1)
```

<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	876	735.3283	NA	NA
2	878	739.0644	-2	-3.736034 0.1544296

2 rows

## Removing Cardiac from the Model

cardiac.hemorrhage was removed from the model due to the high p value of the likelihood ratio test.

```
multi.wvgam = update(multi.wvgam, .~.-cardiac.hemorrhage)
summary(multi.wvgam)
```

```

## 
## Call:
## vglm(formula = DISCHARGE_LOCATION ~ circulatory.system.malformat
## ion +
##       heart.rate.circulation.issues + preterm + respiratory + seps
## is +
##       jaundice, family = multinomial, data = dat)
##
## Pearson residuals:
##                               Min      1Q   Median      3Q     Max
## log(mu[,1]/mu[,3]) -2.574 -0.3657 -0.2577 -0.1174 3.212
## log(mu[,2]/mu[,3]) -3.760 -0.7898  0.5472  0.6539 1.451
##
## Coefficients:
##                                         Estimate Std. Error z value
Pr(>|z|)
## (Intercept):1                  16.21707  650.15032      NA
NA
## (Intercept):2                  15.10740  650.15038      0.023
0.9815
## circulatory.system.malformation1:1 -0.48132    0.35406 -1.359
0.1740
## circulatory.system.malformation1:2  0.03283    0.30828  0.106
0.9152
## heart.rate.circulation.issues1:1 -0.03172    0.35945 -0.088
0.9297
## heart.rate.circulation.issues1:2  0.23917    0.31777  0.753
0.4517
## preterm1:1                     0.44703    0.76882  0.581
0.5609
## preterm1:2                     -0.06165    0.69311 -0.089
0.9291
## preterm2:1                     -1.15802    0.76165 -1.520
0.1284
## preterm2:2                     -1.67988    0.67704 -2.481
0.0131 *
## respiratory1:1                 -15.43629  650.15012 -0.024
0.9811
## respiratory1:2                 -13.87788  650.15022 -0.021
0.9830

```

```

## sepsis1:1           -0.21071   0.46606  -0.452
0.6512
## sepsis1:2           0.76208   0.36564   2.084
0.0371 *
## jaundice1:1         0.35309   0.55687   0.634
0.5260
## jaundice1:2         0.91230   0.51358   1.776
0.0757 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 736.2496 on 878 degrees of freedom
##
## Log-likelihood: -368.1248 on 878 degrees of freedom
##
## Number of Fisher scoring iterations: 15
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1'
##
## Reference group is level 3 of the response

```

Once more, sepsis and prematurity do not need to be tested with the likelihood ratio test, but the remaining factors should be examined.

```
anova(multi.wvgam, update(multi.wvgam, .~.-circulatory.system.malformation), type=1)
```

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	736.2496	NA	NA	NA
2	740.6366	-2	-4.386973	0.1115272
<b>2 rows</b>				

```
anova(multi.wvgam, update(multi.wvgam, .~.-heart.rate.circulation.issues), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	878	736.2496	NA	NA	NA
2	880	737.6679	-2	-1.418278	0.4920676
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-jaundice), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	878	736.2496	NA	NA	NA
2	880	740.0683	-2	-3.818722	0.1481751
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-respiratory), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	878	736.2496	NA	NA	NA
2	880	745.1048	-2	-8.855219	0.011943
2 rows					

## Removing Heart Rate

heart.rate.circulation.issues was removed because the anova showed greatest p-value, indicating the lowest probability of the larger model accounting for additional variability.

```
multi.wvgam = update(multi.wvgam, .~.-heart.rate.circulation.issues)
summary(multi.wvgam)
```

```

## 
## Call:
## vglm(formula = DISCHARGE_LOCATION ~ circulatory.system.malformat
## ion +
##     preterm + respiratory + sepsis + jaundice, family = multinom
## ial,
##     data = dat)
## 
## Pearson residuals:
##                               Min      1Q   Median      3Q      Max
## log(mu[,1]/mu[,3]) -2.702 -0.3820 -0.2361 -0.1224  3.107
## log(mu[,2]/mu[,3]) -4.049 -0.7527  0.5677  0.6840  1.454
## 
## Coefficients:
##                               Estimate Std. Error z value
## Pr(>|z|)                                 
## (Intercept):1          16.181612 650.936583      NA
## NA
## (Intercept):2          15.191569 650.936640      0.023
## 0.9814
## circulatory.system.malformation1:1 -0.487763  0.353582 -1.379
## 0.1677
## circulatory.system.malformation1:2  0.042861  0.307729  0.139
## 0.8892
## preterm1:1              0.445543  0.763832  0.583
## 0.5597
## preterm1:2              -0.005171  0.687914 -0.008
## 0.9940
## preterm2:1              -1.157664  0.758689 -1.526
## 0.1270
## preterm2:2              -1.627385  0.672567 -2.420
## 0.0155
## respiratory1:1          -15.437051 650.936396 -0.024
## 0.9811
## respiratory1:2          -13.869172 650.936483 -0.021
## 0.9830
## sepsis1:1                -0.201768  0.465882 -0.433
## 0.6650
## sepsis1:2                0.773990  0.365385  2.118
## 0.0342

```

```

## jaundice1:1          0.373563   0.556398   0.671
0.5020
## jaundice1:2          0.924644   0.511161   1.809
0.0705
##
## (Intercept):1
## (Intercept):2
## circulatory.system.malformation1:1
## circulatory.system.malformation1:2
## preterm1:1
## preterm1:2
## preterm2:1
## preterm2:2          *
## respiratory1:1
## respiratory1:2
## sepsis1:1
## sepsis1:2          *
## jaundice1:1
## jaundice1:2          .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 737.6679 on 880 degrees of freedom
##
## Log-likelihood: -368.8339 on 880 degrees of freedom
##
## Number of Fisher scoring iterations: 15
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1'
##
##
## Reference group is level 3 of the response

```

```

anova(multi.wvgam, update(multi.wvgam, .~.-circulatory.system.malformation), type=1)

```

<b>Resid. Df</b>	<b>Resid. Dev</b>	<b>Df</b>	<b>Deviance</b>	<b>Pr(&gt;Chi)</b>
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	880	737.6679	NA	NA
2	882	742.3513	-2	-4.683418
<b>2 rows</b>				

```
anova(multi.wvgam, update(multi.wvgam, .~.-jaundice), type=1)
```

<b>Resid. Df</b>	<b>Resid. Dev</b>	<b>Df</b>	<b>Deviance</b>	<b>Pr(&gt;Chi)</b>
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	880	737.6679	NA	NA
2	882	741.5434	-2	-3.875485
<b>2 rows</b>				

```
anova(multi.wvgam, update(multi.wvgam, .~.-respiratory), type=1)
```

<b>Resid. Df</b>	<b>Resid. Dev</b>	<b>Df</b>	<b>Deviance</b>	<b>Pr(&gt;Chi)</b>
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	880	737.6679	NA	NA
2	882	746.5667	-2	-8.898803
<b>2 rows</b>				

## Removing Circulatory System Malformation from the Model

In the Wald Z tests, both sepsis and preterm have levels which would indicate that we should reject the null hypothesis ( $p<0.05$ ), and jaundice is close in the wald test. In the Wald test, respiratory causes are not significant, however, the likelihood ratio tests indicates that we should conclude that a significant amount of variability is accounted for by including this factor. Therefore, the only variable that both tests agree upon is circulatory malformations.

```
multi.wvgam = update(multi.wvgam, .~.-circulatory.system.malformatio  
n)  
summary(multi.wvgam)
```

```

## 
## Call:
## vglm(formula = DISCHARGE_LOCATION ~ preterm + respiratory + sepsis +
##       jaundice, family = multinomial, data = dat)
## 
## Pearson residuals:
##                               Min      1Q   Median      3Q     Max
## log(mu[,1]/mu[,3]) -2.814 -0.3428 -0.2355 -0.1375 2.981
## log(mu[,2]/mu[,3]) -4.037 -0.8140  0.5975  0.6389 1.398
## 
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1      16.06379  646.47400      NA      NA
## (Intercept):2      15.23486  646.47406      0.024  0.9812
## preterm1:1        0.46824   0.76196      0.615  0.5389
## preterm1:2       -0.00435   0.68787     -0.006  0.9950
## preterm2:1        -1.21597   0.75619     -1.608  0.1078
## preterm2:2        -1.61580   0.67119     -2.407  0.0161 *
## respiratory1:1    -15.47033  646.47382     -0.024  0.9809
## respiratory1:2    -13.89519  646.47391     -0.021  0.9829
## sepsis1:1         -0.28609   0.45989     -0.622  0.5339
## sepsis1:2          0.77758   0.36075      2.155  0.0311 *
## jaundice1:1       0.33167   0.55461      0.598  0.5498
## jaundice1:2       0.92281   0.51105      1.806  0.0710 .
## --- 
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
## 
## Residual deviance: 742.3513 on 882 degrees of freedom
## 
## Log-likelihood: -371.1756 on 882 degrees of freedom
## 
## Number of Fisher scoring iterations: 15
## 
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1'

```

```
##  
##  
## Reference group is level 3 of the response
```

```
anova(multi.wvgam, update(multi.wvgam, .~.-respiratory), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	882	742.3513	NA	NA	NA
2	884	751.5161	-2	-9.164858	0.01023002
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.-jaundice), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	882	742.3513	NA	NA	NA
2	884	746.4526	-2	-4.101287	0.1286521
2 rows					

## Removing Jaundice from the Model

The likelihood ratio test explores the null hypothesis as to whether the additional argument in the model accounts for more variability within the data, therefore, while the wald test shows that the estimate for jaundice is approaching our alpha, we must conclude in alignment with the likelihood ratio test. Therefore, removing jaundice from the model did not affect it's accounting of the data's variability, and we would reduce the model as such.

```
multi.wvgam = update(multi.wvgam, .~.-jaundice)
summary(multi.wvgam)
```

```

## 
## Call:
## vglm(formula = DISCHARGE_LOCATION ~ preterm + respiratory + seps
is,
##       family = multinomial, data = dat)
##
## Pearson residuals:
##                               Min      1Q Median      3Q     Max
## log(mu[,1]/mu[,3]) -2.746 -0.3485 -0.2368 -0.1380 2.84
## log(mu[,2]/mu[,3]) -3.892 -0.7894  0.6066  0.6584 1.17
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept):1 16.0918   644.9393     NA      NA
## (Intercept):2 15.4580   644.9393     NA      NA
## preterm1:1     0.5857    0.7470     0.784    0.4330
## preterm1:2     0.2125    0.6668     0.319    0.7499
## preterm2:1    -1.0933    0.7367    -1.484    0.1378
## preterm2:2    -1.3807    0.6442    -2.143    0.0321 *
## respiratory1:1 -15.3190   644.9391    -0.024   0.9810
## respiratory1:2 -13.4944   644.9392    -0.021   0.9833
## sepsis1:1      -0.2733    0.4592    -0.595   0.5517
## sepsis1:2       0.8059    0.3593     2.243   0.0249 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 746.4526 on 884 degrees of freedom
##
## Log-likelihood: -373.2263 on 884 degrees of freedom
##
## Number of Fisher scoring iterations: 15
##
## Warning: Hauck-Donner effect detected in the following estimate
## (s):
## '(Intercept):1', '(Intercept):2'
##
## 
```

```
## Reference group is level 3 of the response
```

```
anova(multi.wvgam, update(multi.wvgam, .~.-respiratory), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	884	746.4526	NA	NA	NA
2	886	757.6726	-2	-11.21999	0.003661095
2 rows					

Because the likelihood ratio test above ( $F(2, )=9.165$ ,  $p=0.0102$ ) indicates that the model accounts for more variability in the data with respiratory variables included, we will proceed with this variable in the model.

## Interaction model with most relevant factors

```
anova(multi.wvgam, update(multi.wvgam, .~.+sepsis * preterm), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	884	746.4526	NA	NA	NA
2	880	741.3563	4	5.096305	0.2775579
2 rows					

```
anova(multi.wvgam, update(multi.wvgam, .~.+respiratory * preterm), type=1)
```

	<b>Resid. Df</b> <dbl>	<b>Resid. Dev</b> <dbl>	<b>Df</b> <dbl>	<b>Deviance</b> <dbl>	<b>Pr(&gt;Chi)</b> <dbl>
1	884	746.4526	NA	NA	NA
2	880	741.5326	4	4.92001	0.2956039

2 rows

```
anova(multi.wvgam, update(multi.wvgam, .~.+sepsis * respiratory), type =1)
```

<b>Resid. Df</b>	<b>Resid. Dev</b>	<b>Df</b>	<b>Deviance</b>	<b>Pr(&gt;Chi)</b>
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	884	746.4526	NA	NA
2	882	746.3935	2	0.05904004
2 rows				

It appears that there are no interactions that need to be included in the model: we would fail to reject the null hypothesis on all likelihood ratio tests ( $p < 0.05$ )

## Final Multinomial Model

```
summary(multi.wvgam)
```

```

## 
## Call:
## vglm(formula = DISCHARGE_LOCATION ~ preterm + respiratory + seps
is,
##       family = multinomial, data = dat)
##
## Pearson residuals:
##                               Min      1Q Median      3Q     Max
## log(mu[,1]/mu[,3]) -2.746 -0.3485 -0.2368 -0.1380 2.84
## log(mu[,2]/mu[,3]) -3.892 -0.7894  0.6066  0.6584 1.17
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1    16.0918   644.9393     NA      NA
## (Intercept):2    15.4580   644.9393     NA      NA
## preterm1:1        0.5857    0.7470     0.784    0.4330
## preterm1:2        0.2125    0.6668     0.319    0.7499
## preterm2:1       -1.0933    0.7367    -1.484    0.1378
## preterm2:2       -1.3807    0.6442    -2.143    0.0321 *
## respiratory1:1   -15.3190   644.9391    -0.024   0.9810
## respiratory1:2   -13.4944   644.9392    -0.021   0.9833
## sepsis1:1        -0.2733    0.4592    -0.595   0.5517
## sepsis1:2         0.8059    0.3593     2.243   0.0249 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 746.4526 on 884 degrees of freedom
##
## Log-likelihood: -373.2263 on 884 degrees of freedom
##
## Number of Fisher scoring iterations: 15
##
## Warning: Hauck-Donner effect detected in the following estimate
## (s):
## '(Intercept):1', '(Intercept):2'
##
## 
```

```
## Reference group is level 3 of the response
```

# Results

## Aim 1): Cox Proportional Hazards Model

Model Interpretation:

```
summary(final_model)
```

```

## Call:
## coxph(formula = Surv(LOS, delta) ~ preterm * circulatory.system.
malformation +
##       sepsis, data = data, method = c("breslow"))
##
##      n= 450, number of events= 449
##
##                                         coef
exp(coef)
## pretermExtreme preterm                         -1.5991
0.2021
## pretermPreterm                                -0.3296
0.7192
## circulatory.system.malformation1            -1.0228
0.3596
## sepsis1                                         -0.4761
0.6212
## pretermExtreme preterm:circulatory.system.malformation1 1.0409
2.8317
## pretermPreterm:circulatory.system.malformation1          0.6074
1.8357
##
##                                         se(coef)
z
## pretermExtreme preterm                         0.2610
-6.128
## pretermPreterm                                0.2358
-1.398
## circulatory.system.malformation1            0.3317
-3.083
## sepsis1                                         0.1169
-4.071
## pretermExtreme preterm:circulatory.system.malformation1 0.3699
2.814
## pretermPreterm:circulatory.system.malformation1          0.3568
1.702
##
##                                         Pr
(>| z | )
## pretermExtreme preterm                         0.000000
000892 ***
## pretermPreterm

```

```
0.16208
## circulatory.system.malformation1
0.00205 **
## sepsis1                                0.000046
725442 ***
## pretermExtreme preterm:circulatory.system.malformation1
0.00490 **
## pretermPreterm:circulatory.system.malformation1
0.08866 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coe
f)
## pretermExtreme preterm                  0.202
1
## pretermPreterm                           0.719
2
## circulatory.system.malformation1       0.359
6
## sepsis1                                 0.621
2
## pretermExtreme preterm:circulatory.system.malformation1   2.831
7
## pretermPreterm:circulatory.system.malformation1          1.835
7
## exp(-coe
f)
## pretermExtreme preterm                  4.94
88
## pretermPreterm                           1.39
04
## circulatory.system.malformation1       2.78
10
## sepsis1                                 1.60
97
## pretermExtreme preterm:circulatory.system.malformation1   0.35
31
## pretermPreterm:circulatory.system.malformation1          0.54
48
```

```

##                                         lower .9
5
## pretermExtreme preterm                         0.121
2
## pretermPreterm                           0.453
1
## circulatory.system.malformation1          0.187
7
## sepsis1                                 0.494
0
## pretermExtreme preterm:circulatory.system.malformation1 1.371
4
## pretermPreterm:circulatory.system.malformation1        0.912
2
##                                         upper .9
5
## pretermExtreme preterm                      0.337
0
## pretermPreterm                           1.141
6
## circulatory.system.malformation1          0.688
9
## sepsis1                                 0.781
2
## pretermExtreme preterm:circulatory.system.malformation1 5.847
1
## pretermPreterm:circulatory.system.malformation1        3.694
0
##
## Concordance= 0.676  (se = 0.013 )
## Likelihood ratio test= 147.3  on 6 df,    p=<0.0000000000000002
## Wald test           = 140.3  on 6 df,    p=<0.0000000000000002
## Score (logrank) test = 152.9  on 6 df,    p=<0.0000000000000002

```

$$\begin{aligned}
\ln \left( \frac{h_{Discharge}(t)}{h_{Discharge0}(t)} \right) = & \hat{\beta}_{Preterm1} \times X_{Preterm1} + \hat{\beta}_{Preterm2} \times X_{Preterm2} + \hat{\beta}_{Ciculate} \\
& \hat{\beta}_{Sepsis} \times X_{Sepsis} + \hat{\beta}_{Preterm1*CirculatoryMalformation} \times \\
& \hat{\beta}_{Preterm2*CirculatoryMalformation} \times X_{Preterm2*Circulatory}.
\end{aligned}$$

- Extreme preterm infants were significantly less likely to be discharged from NICU than full-term infants ( $p < 0.0001$ ), while there was not a significant difference between full-term and preterm infants ( $p = 0.16$ ).  $\exp(\beta)$  for extreme preterm relative to full-term is 0.2021, indicating that the infants diagnosed as having sepsis were 0.2 times as likely to be discharged from NICU than were infants who were full-term.
- A diagnosis of sepsis significantly affects an infant's likelihood of being discharged from NICU ( $p = 0.00005$ ).  $\exp(\beta)$  for sepsis is 0.6212, indicating that the infants diagnosed as having sepsis were 0.62 times as likely to be discharged from NICU than were infants who had not received this diagnosis. i.e. having sepsis decreased an infant's likelihood of being discharged.
- For full-term infants, being diagnosed with a circulatory system malformation significantly decreased an infant's likelihood of being discharged from NICU ( $P = 0.002$ ), and there was an interaction between prematurity level and having a circulatory system malformation on likelihood of discharge.
- The interaction between an infant being extreme preterm and having a circulatory system malformation is significant ( $p = 0.0049$ ), while the interaction between an infant being preterm and having a circulatory system malformation is just above the threshold of significance ( $p = 0.089$ ). This indicates that the effect on likelihood of being discharged from NICU of having, or not having, a circulatory system malformation differed depending on which level of preterm the infant is.
- Specifically, this indicates that having a circulatory system malformation has a lesser affect on likelihood of being discharged from NICU for extreme preterm infants than for full-term infants. Additionally, the near-significant interaction between being preterm and having circulatory system malformation hints that having a circulatory system malformation may have had a lesser effect on likelihood of being discharged from NICU for preterm infants than for non-preterm infants.

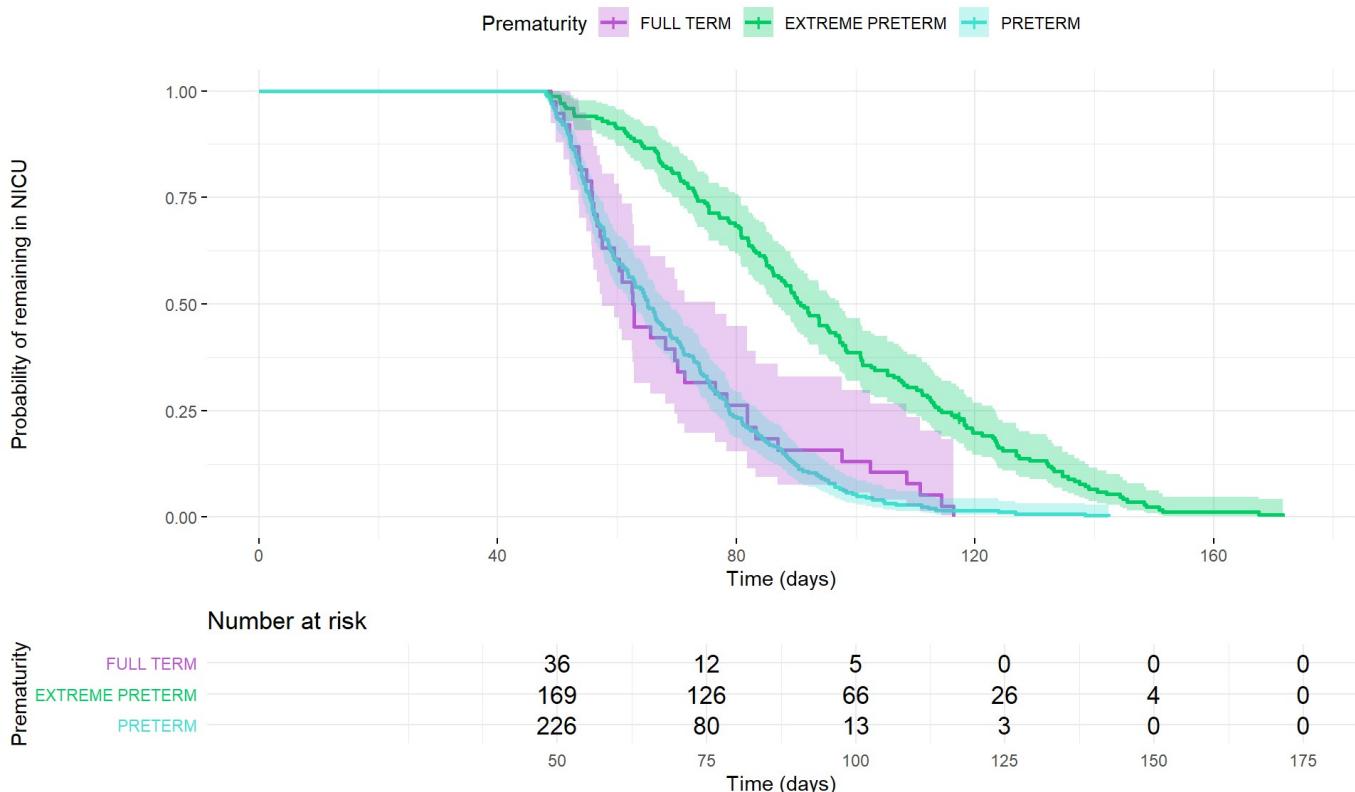
## Plotting Kaplan-Meier Curves for Variables Significant in Coxph Model

The Kaplan-Meier Curves plotted here are mainly intended as visual illustrations of the results of Cox PH model detailed above.

# Plot survival curves for each level of prematurity:

```
pretermSurv <- survfit(Surv(LOS, delta) ~ preterm, data=data)
```

```
ggsurvplot(
  pretermSurv,
  data = data,
  risk.table = TRUE,
  conf.int = TRUE,
  ggtheme = theme_minimal(),
  risk.table.y.text.col = T,
  palette=c("mediumorchid", "springgreen3", 'turquoise'),
  legend.labs=c('FULL TERM', 'EXTREME PRETERM', 'PRETERM'),
  legend.title="Prematurity",
  ylab='Probability of remaining in NICU',
  xlab='Time (days)'
)
```



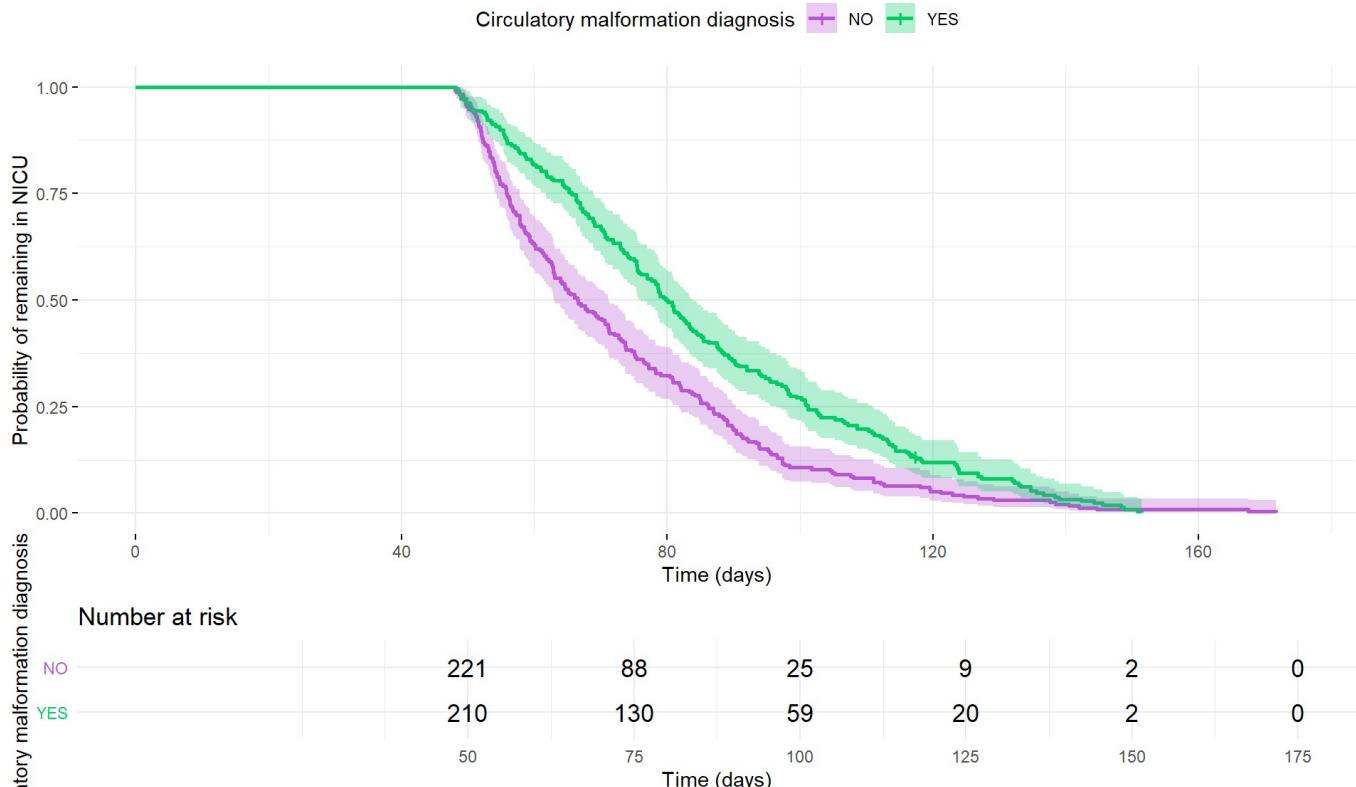
- You can see that the confidence intervals for full-term and preterm overlap, and their curves cross one another. As such, we can't say that one group showed higher likelihood of being discharged from NICU than did the other. However, the

curve for extreme preterm is separate from these other 2 curves, and shows a decreased likelihood to be discharged from NICU for these infants compared to the other 2 groups.

## Plot survival curves for presence/absence of circulatory system malformation:

```
circSurv <- survfit(Surv(LOS, delta) ~ circulatory.system.malformat  
ion, data=data)
```

```
ggsurvplot(  
  circSurv,  
  data = data,  
  risk.table = TRUE,  
  conf.int = TRUE,  
  ggtheme = theme_minimal(),  
  risk.table.y.text.col = T,  
  palette=c("mediumorchid", "springgreen3"),  
  legend.labs=c('NO', 'YES'),  
  legend.title="Circulatory malformation diagnosis",  
  ylab='Probability of remaining in NICU',  
  xlab='Time (days)'  
)
```



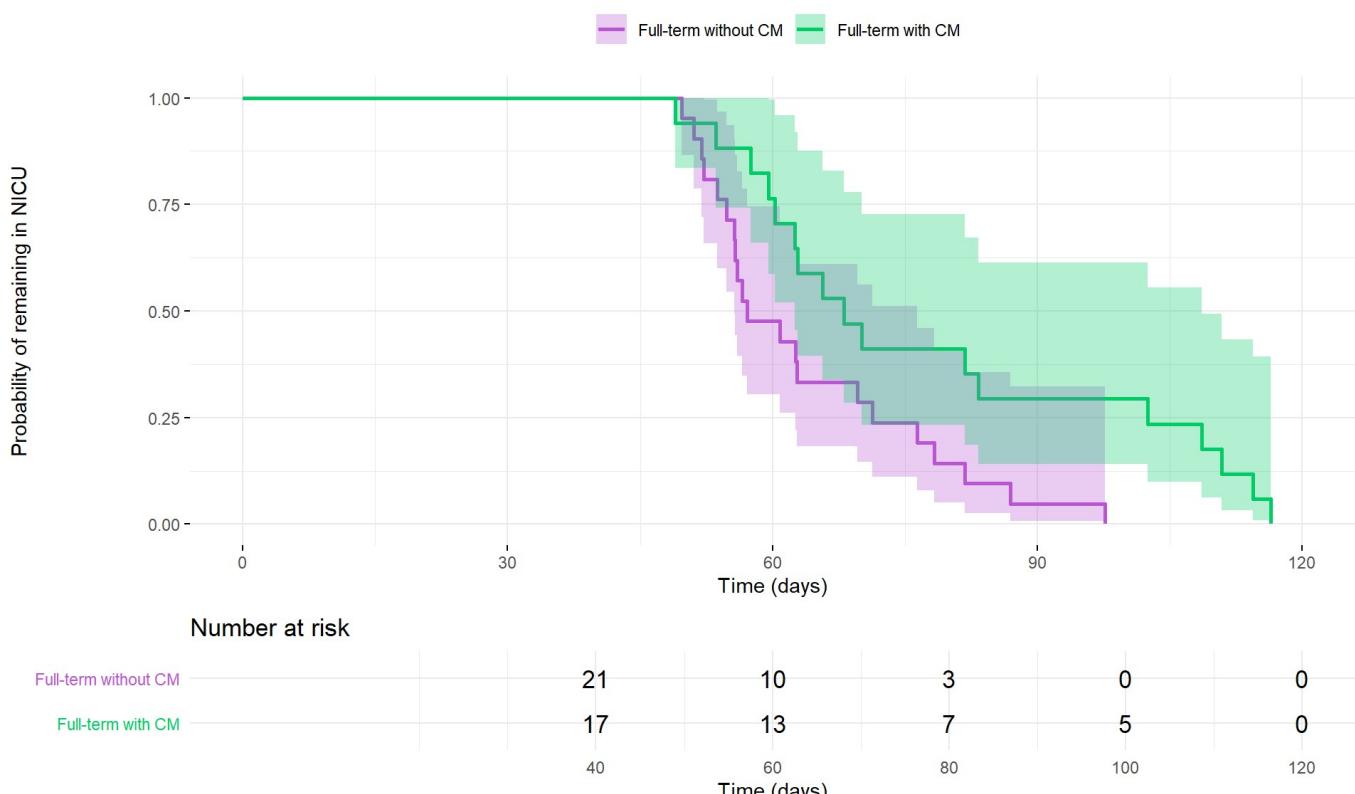
- Having been diagnosed with a circulatory system malformation decreased an infant's likelihood of being discharged from NICU.

**Plot interaction between level of prematurity and circulatory system malformation:**

**Full-term with/without circulatory system malformation:**

```
p0Surv <- survfit(Surv(LOS, delta) ~ Group, data=p0)
```

```
ggsurvplot(
  p0Surv,
  data = data,
  risk.table = TRUE,
  conf.int = TRUE,
  ggtheme = theme_minimal(),
  risk.table.y.text.col = T,
  palette=c("mediumorchid", "springgreen3"),
  legend.labs=c('Full-term without CM', 'Full-term with CM'),
  legend.title="",
  ylab='Probability of remaining in NICU',
  xlab='Time (days)'
)
```

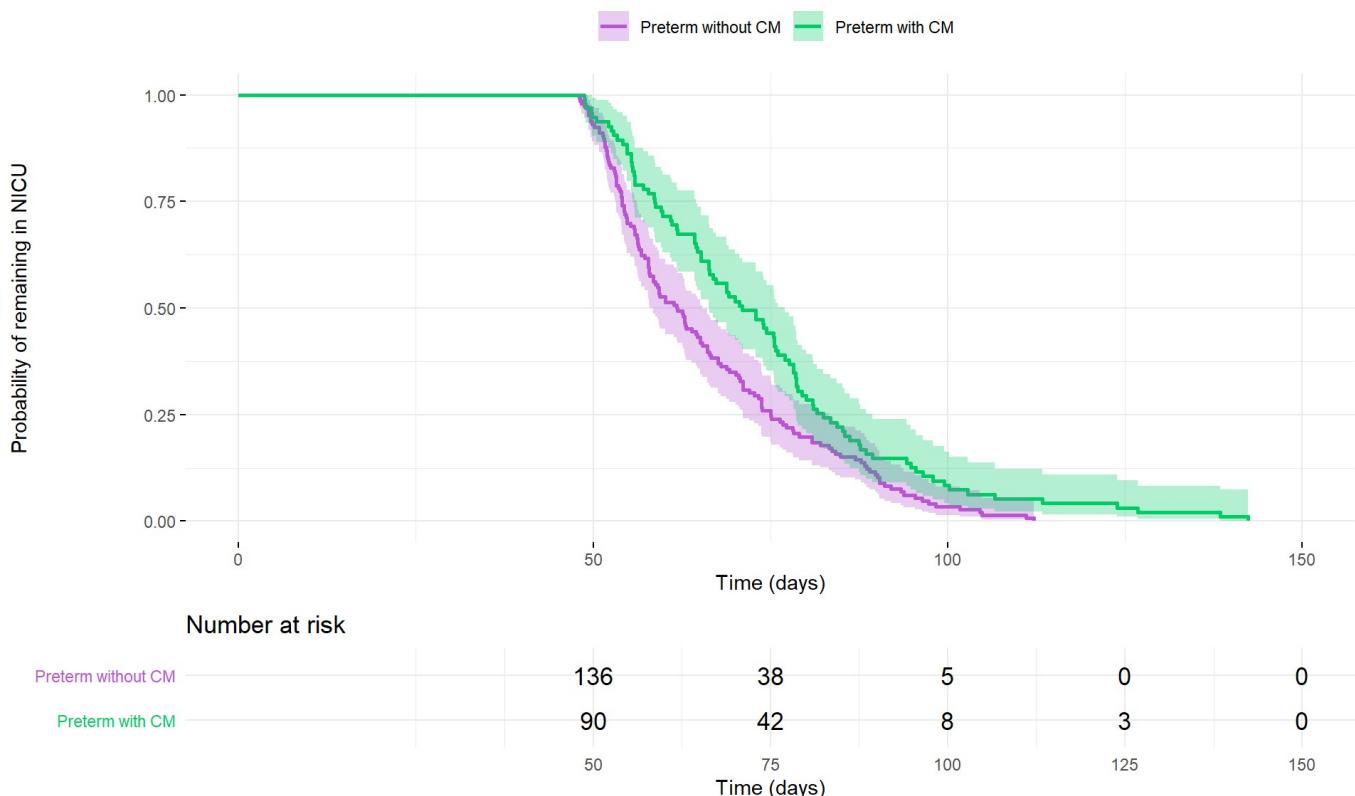


- We can see that curves do not cross, though the CI's overlap. For full-term infants, having a circulatory system malformation decreased the likelihood of being discharged from NICU.

## Preterm with/without circulatory system malformation:

```
p1Surv <- survfit(Surv(LOS, delta) ~ Group, data=p1)
```

```
ggsurvplot(
  p1Surv,
  data = data,
  risk.table = TRUE,
  conf.int = TRUE,
  ggtheme = theme_minimal(),
  risk.table.y.text.col = T,
  palette=c("mediumorchid", "springgreen3"),
  legend.labs=c('Preterm without CM', 'Preterm with CM'),
  legend.title="",
  ylab='Probability of remaining in NICU',
  xlab='Time (days)'
)
```



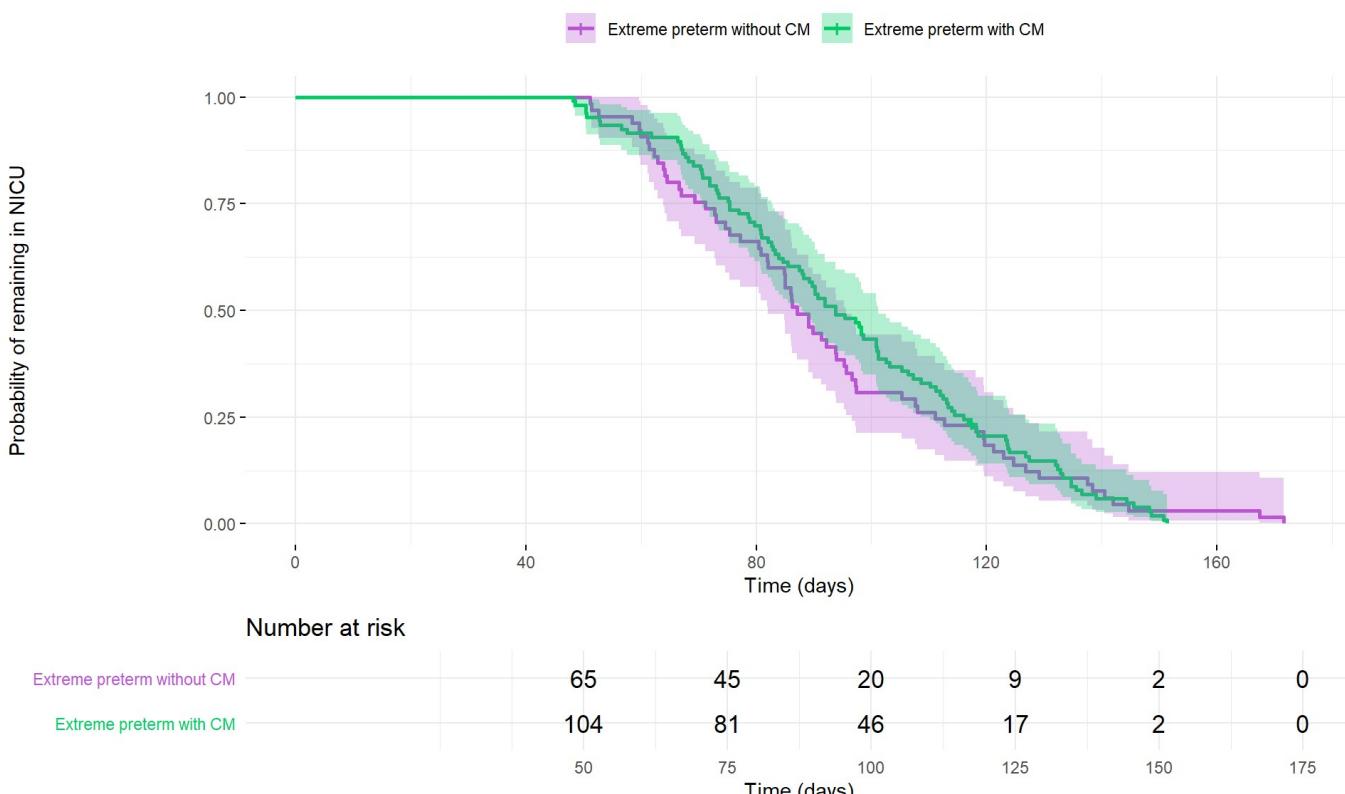
- Can see that curves do not cross, though CI's overlap in some places. For preterm infants, having a circulatory system malformation decreased likelihood of being discharged from NICU.

## Extreme preterm with/without circulatory system

# malformation:

```
p2Surv <- survfit(Surv(LOS, delta) ~ Group, data=p2)
```

```
ggsurvplot(
  p2Surv,
  data = data,
  risk.table = TRUE,
  conf.int = TRUE,
  ggtheme = theme_minimal(),
  risk.table.y.text.col = T,
  palette=c("mediumorchid", "springgreen3"),
  legend.labs=c('Extreme preterm without CM', 'Extreme preterm with CM'),
  legend.title="",
  ylab='Probability of remaining in NICU',
  xlab='Time (days)'
)
```



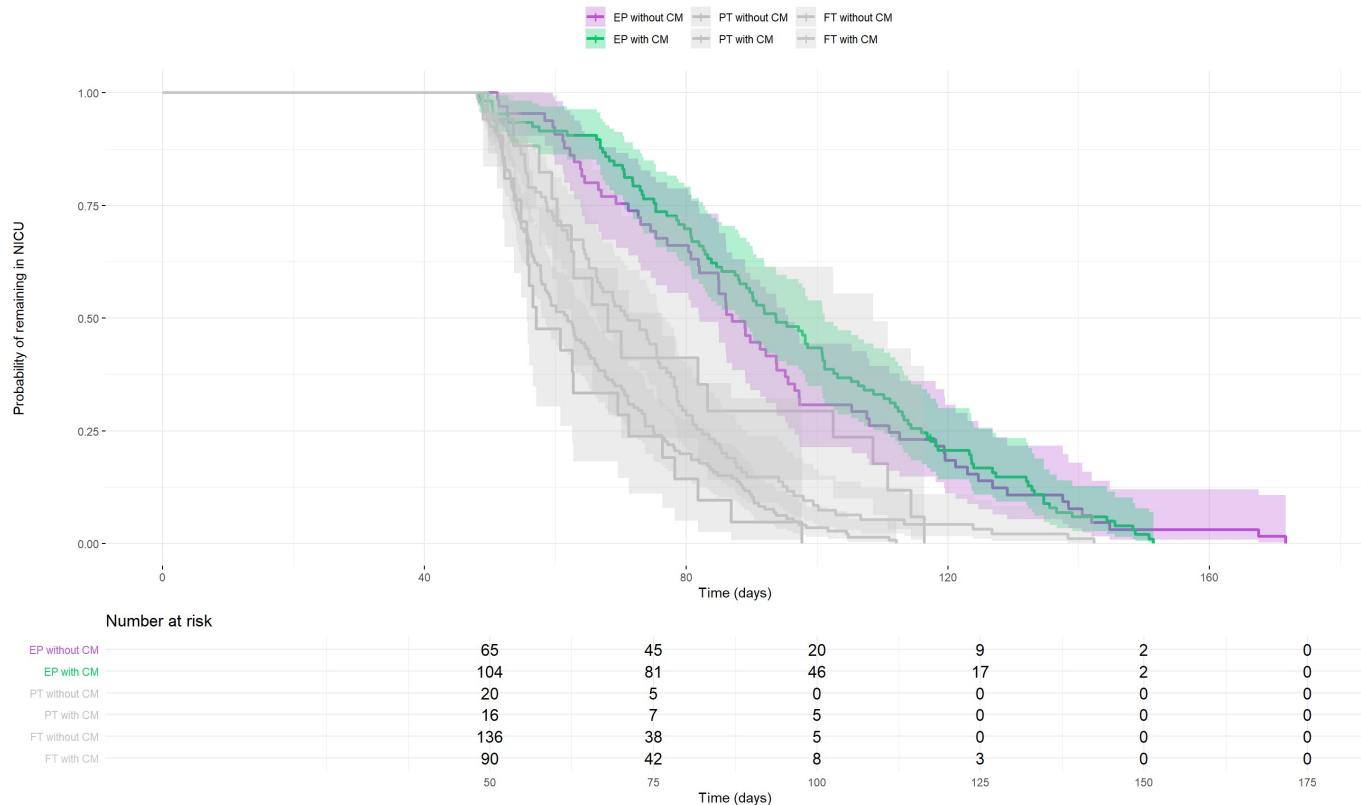
- Can see that the curves are very close to one another, the CI's overlap extensively, and the curves cross at multiple points. Thus, we can not say that

having a circulatory system malformation decreased or increased an extreme preterm infant's likelihood of being discharged from NICU.

## Plot interaction of preterm level and presence/absence of circulatory system malformation:

```
groupSurv <- survfit(Surv(LOS, delta) ~ Group, data=data)
```

```
ggsurvplot(  
  groupSurv,  
  data = data,  
  risk.table = TRUE,  
  conf.int = TRUE,  
  ggtheme = theme_minimal(),  
  risk.table.y.text.col = T,  
  palette=c("mediumorchid", "springgreen3", "gray75", "gray76", 'gray77', 'gray78'),  
  legend.labs=c('EP without CM', 'EP with CM', 'PT without CM', 'PT with CM', 'FT without CM', 'FT with CM'),  
  legend.title="",  
  ylab='Probability of remaining in NICU',  
  xlab='Time (days)',  
  size = 1.2  
)
```

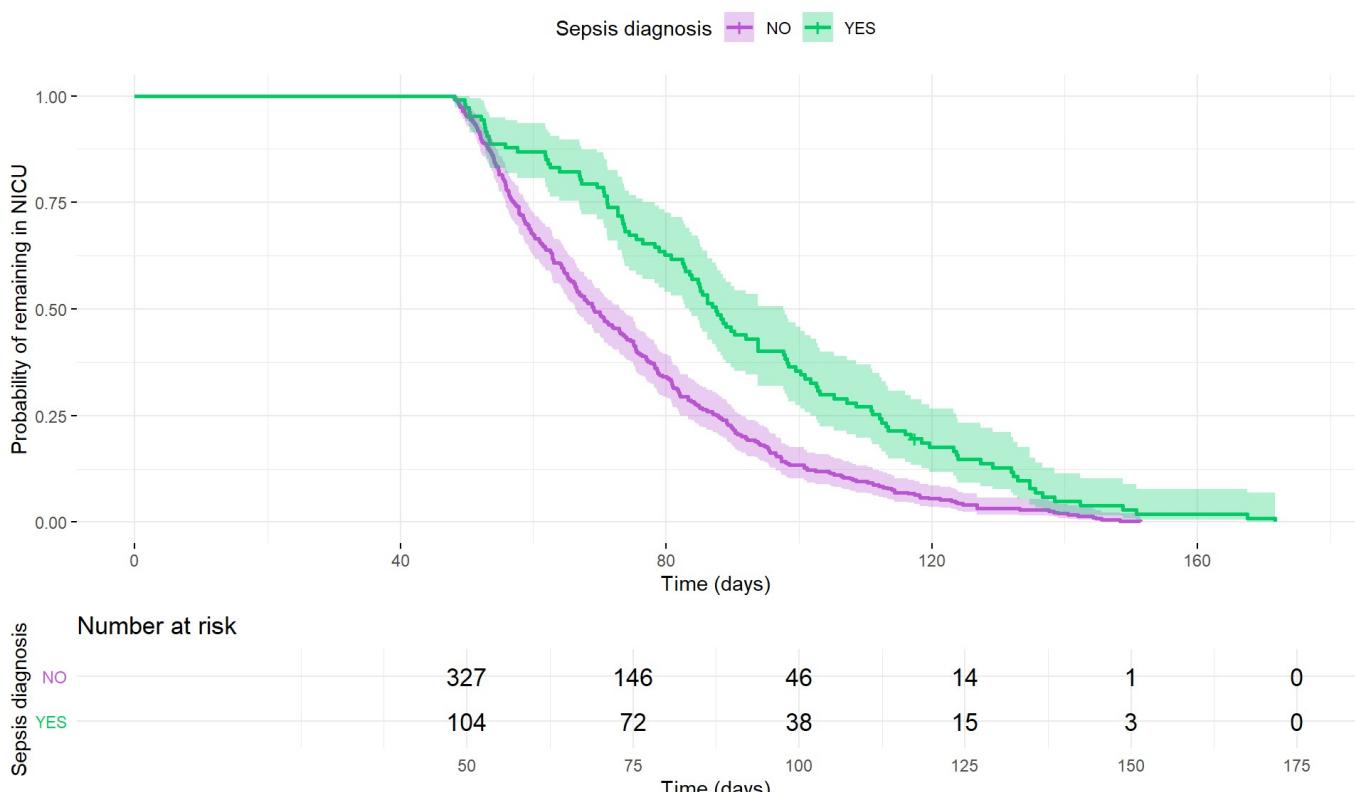


- This is a plot of all 3 levels of prematurity with and without a circulatory system malformation (full-term and preterm have been grayed out to emphasize extreme preterm). We can see that extreme preterm with circulatory malformation appears to have the lowest likelihood of being discharged from NICU overall. However, as was seen above, the curves for extreme preterm with CM, and extreme preterm without CM cross, thus we can't say that either extreme preterm group has a higher or lower likelihood of being discharged from NICU.

## Plot for infants with and without sepsis:

```
sepSurv <- survfit(Surv(LOS, delta) ~ sepsis, data=data)
```

```
ggsurvplot(
  sepSurv,
  data = data,
  risk.table = TRUE,
  conf.int = TRUE,
  ggtheme = theme_minimal(),
  risk.table.y.text.col = T,
  palette=c("mediumorchid", "springgreen3"),
  legend.labs=c('NO', 'YES'),
  legend.title="Sepsis diagnosis",
  ylab='Probability of remaining in NICU',
  xlab='Time (days)'
)
```



- Can see that a diagnosis of sepsis decreased the likelihood of an infant being discharged from NICU.

## Aim 2): Tests for proportional differences between preterm levels

- There was a significant difference between preterm and extreme preterm in

prevalence of circulatory system malformations ( $p = 0.00003$ ), with extreme preterm having higher prevalence.

- There were significant differences between full-term and preterm ( $p = 0.016$ ), and full-term and extreme preterm ( $p = 0.020$ ) for prevalence of heart rate anomalies, where full-term had a lower prevalence of heart rate anomalies than the other two groups.
- There was a significant difference between preterm and extreme preterm in prevalence of sepsis (0.00000047), with extreme preterm having a higher prevalence of sepsis.
- There were significant differences between full-term and preterm ( $p = 0.0023$ ), and full-term and extreme preterm ( $p = 0.0002$ ), for prevalence of jaundice, with full-term having a lower prevalence of jaundice than the other two groups.

## Aim 3): Multinomial Model

```
summary(multi.wvgam)
```

```

## 
## Call:
## vglm(formula = DISCHARGE_LOCATION ~ preterm + respiratory + seps
is,
##       family = multinomial, data = dat)
##
## Pearson residuals:
##                               Min      1Q Median      3Q     Max
## log(mu[,1]/mu[,3]) -2.746 -0.3485 -0.2368 -0.1380 2.84
## log(mu[,2]/mu[,3]) -3.892 -0.7894  0.6066  0.6584 1.17
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept):1 16.0918   644.9393     NA      NA
## (Intercept):2 15.4580   644.9393     NA      NA
## preterm1:1     0.5857    0.7470     0.784    0.4330
## preterm1:2     0.2125    0.6668     0.319    0.7499
## preterm2:1    -1.0933    0.7367    -1.484    0.1378
## preterm2:2    -1.3807    0.6442    -2.143    0.0321 *
## respiratory1:1 -15.3190   644.9391    -0.024   0.9810
## respiratory1:2 -13.4944   644.9392    -0.021   0.9833
## sepsis1:1      -0.2733    0.4592    -0.595   0.5517
## sepsis1:2       0.8059    0.3593     2.243   0.0249 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 746.4526 on 884 degrees of freedom
##
## Log-likelihood: -373.2263 on 884 degrees of freedom
##
## Number of Fisher scoring iterations: 15
##
## Warning: Hauck-Donner effect detected in the following estimate
## (s):
## '(Intercept):1', '(Intercept):2'
##
## 
```

## Reference group is level 3 of the response

Considering that the reference level is level 3 of Discharge location ("Short term hospital stay), after removing the three levels without enough data, we see focus on home and home health care discharges.

$$\ln\left(\frac{P(\text{Discharge} = \text{Home})}{P(\text{Discharge} = \text{Short Term Hospital})}\right) = \hat{\beta}_{1,0} + \hat{\beta}_{1,\text{Preterm1}} \times X_{1,\text{Preterm1}} + \hat{\beta}_{1,\text{Respiratory}} \times X_{1,\text{Respirat}}$$

$$\ln\left(\frac{P(\text{Discharge} = \text{HomeHealthCare})}{P(\text{Discharge} = \text{Short Term Hospital})}\right) = \hat{\beta}_{2,0} + \hat{\beta}_{2,\text{Preterm}} \times X_{2,\text{Preterm}} + \hat{\beta}_{2,\text{Respiratory}} \times X_{2,\text{Respirat}}$$

If we input the numbers from the above modelling into these equations, they become:

$$\ln\left(\frac{P(\text{Discharge} = \text{Home})}{P(\text{Discharge} = \text{Short Term Hospital})}\right) = 16.0918 + 0.5857 \times X_{1,\text{Preterm1}} - 15.3190 \times X_{1,\text{Respirator}}$$

$$\ln\left(\frac{P(\text{Discharge} = \text{HomeHealthCare})}{P(\text{Discharge} = \text{Short Term Hospital})}\right) = 15.4580 + 0.2125 \times X_{2,\text{Preterm}} - 13.4944 \times X_{2,\text{Respirator}}$$

Therefore, we would conclude that for the first question, the log odds of a newborn going home compared to staying in the hospital, with an intercept of 16.0918289, slightly preterm infants (preterm1) show a log odds increase of 0.5856823 ( $Z=0.7840627$ ,  $p = 0.4330033$ ), which is to say, an odds ratio 1.7962161 times greater, to go home than have a short term hospital stay. In extremely preterm children, the log odds that they will go home decreases by -1.0932917 ( $Z=-1.484068$ ,  $p = 0.1377909$ ). Therefore, we would expect the odds of going home to decrease by 0.3351116 times in favour of a short term hospital stay. In this model, the effect of being slightly preterm and extremely preterm are not additive, a child will fall into one category or the other at birth, and according to this data set, will not change into the other category after ex-utero maturation as these clinical presentations are considered different in complexity and risk. Furthermore, respiratory issues decrease the log odds of a newborn going straight home by -15.3189769 ( $Z=-0.0237526$ ,  $p = 0.98105$ ). If we convert that main effect to an

odds ratio, it is 0.0000002 times less likely that the patient will go directly home. Similarly, with sepsis, we see a log odds ratio estimate of -0.2733377 ( $Z=-0.5951852$ ,  $p = 0.5517197$ ), which translates to 0.7608358 times as likely to be discharged directly home if sepsis or infections are the cause for NICU admission.

The equation to determine if a child will be discharge to home with health care aids follows a very similar pattern, with slightly different values. With an intercept of 15.4580127, slightly preterm babies (preterm1) show a log odds increase of 0.2125326 ( $Z=0.3187341$ ,  $p = 0.7499281$ ), which indicates a change in the odds ratio of 1.2368065 times. Extremely preterm children show a log odds of going home with health care of -1.3806838 ( $Z=-2.1431628$ ,  $p = 0.0321$ ), or an odds ratio of 0.2514066 times as likely to go home for health care as to be discharged from NICU to short term hospital stay if they were born extremely premature. If respiratory issues were present, we would expect a decrease log odds of a newborn going straight home by -13.4943861 ( $Z=-0.0209235$ ,  $p = 0.9833067$ ), or that the odds of going to home health care decrease by 0.0000014 times. Finally, the presence of sepsis indicates that the log odds increase by 0.8059061 ( $Z=2.2432339$ ,  $p = 0.0248817$ ), which translates to 2.2387241 times.

## Conclusions

Extremely preterm infants were less likely than full-term infants to be discharged from NICU, while there was no difference between full-term and preterm. This is in agreement with much of the current literature, in which increasing levels of prematurity require lengthy hospital stays to complete development (Manktelow et al., 2010). We also found that having a circulatory system malformation decreased infants' likelihood of being discharged from NICU, but less so for extremely preterm infants. This was counter to what we had expected, as we had expected any health insult to have a more deleterious effect on the more vulnerable extremely preterm infants. However, this may make sense when considering that extremely preterm infants are already expected to have a lengthier stay than full-term infants, and so any extension caused by a malformation is less likely to cause a large relative extension to their stay. Not unexpectedly, sepsis caused decreased likelihood to be discharged from NICU for all infants, meaning it caused lengthier stays. Again not unexpectedly, we found that prematurity level and the prevalence of certain health conditions were not independent, with circulatory system malformations, heart rate anomalies, jaundice and sepsis tending to be associated with greater levels of prematurity.

We found that respiratory issues, sepsis, and preterm status were significant predictors of the discharge location. We had hypothesized that sepsis and prematurity would be

included in the model, but we had believed that interactions between these variables would be just as important as the variables themselves. However, the interactions were not important to the model, either testing initially or at the end. While not surprising that respiratory issues would lead to decreased odds of going home or going home with health care in place, that they did not interact with prematurity was surprising. Extremely premature babies have inadequately developed lungs, and respiration is one of the latest developments to be made in utero, so this finding was unexpected.

Thus, overall, we see that an infant's health outcome after admission to NICU is affected by multiple factors. Interestingly, the interactions between maturity level and health conditions did not pan out as we had expected. We had expected that increasing levels of prematurity would make any health upset even more dangerous, causing a larger relative increase in hospital stay, or a greater likelihood of requiring further care after NICU. This was not the case, however; Interactions between prematurity and health conditions either were non-significant, or counter to our initial predictions. As explained above, the most intuitive reason that this may be the case is that NICU stay length tends to increase with increasingly levels of prematurity by default, as these infants must advance in their development while at the NICU. As such, any influence on length of stay from other conditions is likely to simply be absorbed into these already lengthy stays.

This study suffered from a few limitations which future studies could address. For one, we collapsed many different diagnoses into somewhat broad categories, so this could have hidden a lot of variability in stay length based on diagnosis. In fact, some diagnoses were ignored entirely, and so could be influencing stay length without our accounting for them. Additionally, as we used the diagnoses that infants were admitted to NICU for, and did not factor in whether these diagnoses were refined or added to at subsequent time points, we may have some incorrect diagnoses in our data set. Major diagnostic errors have been found to occur in NICU settings (e.g. errors in 19.2% of diagnoses of NICU deaths (Custer et al., 2015)), and so this could have caused deviations from reality in our data. Our study only assessed health outcomes directly relating to NICU. This is likely not the end of the health journey for many infants, and so future studies could also look at longer-term outcomes for infants discharged from NICU.

## References:

- Aita, M., Stremler, R., Feeley, N., Lavallée, A., & De Clifford-Faugère, G. (2017). Effectiveness of interventions during NICU hospitalization on the neurodevelopment of

- preterm infants: A systematic review protocol. *Systematic Reviews*, 6(1), 1–5. <https://doi.org/10.1186/s13643-017-0613-5> (<https://doi.org/10.1186/s13643-017-0613-5>)
- Couturier, D.-L., Ma, J., Heritier, S., & Manuguerra, M. (2017). survivalMPL: Penalised Maximum Likelihood for Survival Analysis Models. Retrieved from <https://cran.r-project.org/package=survivalMPL> (<https://cran.r-project.org/package=survivalMPL>)
- Cummings, J. J., & Marlow, N. (2018). Reducing variations in neonatal outcomes: Look at practices, systems, and the patient. *Pediatrics*, 141(5), 2018–2021. <https://doi.org/10.1542/peds.2018-0402> (<https://doi.org/10.1542/peds.2018-0402>)
- Custer, J. W., Winters, B. D., Goode, V., Robinson, K. A., Yang, T., Pronovost, P. J., & Newman-Toker, D. E. (2015). Diagnostic errors in the pediatric and neonatal ICU: A systematic review. *Pediatric Critical Care Medicine*, 16(1), 29–36. <https://doi.org/10.1097/PCC.0000000000000274> (<https://doi.org/10.1097/PCC.0000000000000274>)
- Discenza, D. (2018). Real Awareness in the NICU and Through the Lifespan. *Neonatal Network*, 37(4), 248–249.
- Foglia, E. E., Langeveld, R., Heimall, L., Deveney, A., Ades, A., Jensen, E. A., & Nadkarni, V. M. (2017). Incidence, characteristics, and survival following cardiopulmonary resuscitation in the quaternary neonatal intensive care unit. *Resuscitation*, 110, 32–36. <https://doi.org/10.1016/j.resuscitation.2016.10.012> (<https://doi.org/10.1016/j.resuscitation.2016.10.012>)
- Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., ... Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3, 1–9. <https://doi.org/10.1038/sdata.2016.35> (<https://doi.org/10.1038/sdata.2016.35>)
- Karlsson, A., & Clements, M. (2019). biostat3: Utility Functions, Datasets and Extended Examples for Survival Analysis. Retrieved from <https://cran.r-project.org/package=biostat3> (<https://cran.r-project.org/package=biostat3>)
- Kassambara, A., Kosinski, M., & Biecek, P. (2019). survminer: Drawing Survival Curves using “ggplot2.” Retrieved from <https://cran.r-project.org/package=survminer> (<https://cran.r-project.org/package=survminer>)
- Kono, Y., Mishina, J., Yonemoto, N., Kusuda, S., Fujimura, M., & NICU Network, Japan. (2011). Neonatal correlates of adverse outcomes in very low-birthweight infants in the NICU Network. *Pediatrics International*, 53(6), 930–935.
- Kurek Eken, M., Tüten, A., Özkaya, E., Karatekin, G., & Karateke, A. (2017). Major

determinants of survival and length of stay in the neonatal intensive care unit of newborns from women with premature preterm rupture of membranes. *Journal of Maternal-Fetal and Neonatal Medicine*, 30(16), 1972–1975. <https://doi.org/10.1080/14767058.2016.1235696> (<https://doi.org/10.1080/14767058.2016.1235696>)

Manktelow, B., Draper, E. S., Field, C., & Field, D. (2010). Estimates of length of neonatal stay for very premature babies in the UK. *Archives of Disease in Childhood: Fetal and Neonatal Edition*, 95(4). <https://doi.org/10.1136/adc.2009.168633> (<https://doi.org/10.1136/adc.2009.168633>)

Meadow, W., Lagatta, J., Andrews, B., & Lantos, J. (2012). The mathematics of morality for neonatal resuscitation. *Clinics in Perinatology*, 39(4), 941–956. <https://doi.org/10.1016/j.clp.2012.09.013> (<https://doi.org/10.1016/j.clp.2012.09.013>)

Platt, M. J. (2014). Outcomes in preterm infants. *Public health*, 128(5), 399-403.

Peebles-Kleiger, M. J. (2000). Pediatric and neonatal intensive care hospitalization as traumatic stressor: implications for intervention. *Bulletin of the Menninger Clinic*, 64(2), 257.

R Core Team. (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://doi.org/https://doi.org/ISBN> (<https://doi.org/https://doi.org/ISBN>) 3-900051-07-0

RStudio Team. (2015). RStudio: Integrated Development for R. RStudio, Inc. Retrieved from <http://www.rstudio.com/> (<http://www.rstudio.com/>).

Rysavy, M. A. (2018). Prognosis as an Intervention. *Clinics in Perinatology*, 45(2), 231–240. <https://doi.org/10.1016/j.clp.2018.01.009> (<https://doi.org/10.1016/j.clp.2018.01.009>)

Seaton, S. E., Barker, L., Jenkins, D., Draper, E. S., Abrams, K. R., & Manktelow, B. N. (2016). What factors predict length of stay in a neonatal unit: A systematic review. *BMJ Open*, 6(10). <https://doi.org/10.1136/bmjopen-2015-010466> (<https://doi.org/10.1136/bmjopen-2015-010466>)

Sengupta, S., Carrion, V., Shelton, J., Wynn, R. J., Ryan, R. M., Singhal, K., & Lakshminrusimha, S. (2013). Adverse neonatal outcomes associated with early-term birth. *JAMA pediatrics*, 167(11), 1053-1059.

Subedi, D., Deboer, M. D., & Scharf, R. J. (2017). Developmental trajectories in children with prolonged NICU stays. *Archives of Disease in Childhood*, 102(1), 29–34. <https://doi.org/10.1136/archdischild-2016-310777> (<https://doi.org/10.1136/archdischild-2016-310777>)

2016-310777)

Tang, Y., Horikoshi, M., & Li, W. (2016). *ggfortify: Data Visualization Tools for Statistical Analysis Results*. Retrieved from <https://cran.r-project.org/package=ggfortify> (<https://cran.r-project.org/package=ggfortify>)

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Retrieved from <https://cran.r-project.org/package=survival> (<https://cran.r-project.org/package=survival>)

Umberger, E., Canvasser, J., & Hall, S. L. (2018). Enhancing NICU parent engagement and empowerment. *Seminars in Pediatric Surgery*, 27(1), 19–24. <https://doi.org/10.1053/j.sempeudsurg.2017.11.004> (<https://doi.org/10.1053/j.sempeudsurg.2017.11.004>)

Wickam, H. (2009). *ggplot2: Elegant Graphics for Data Analysis* (35th ed.). <https://doi.org/https://doi.org/10.1007/978-0-387-98141-3> (<https://doi.org/https://doi.org/10.1007/978-0-387-98141-3>)

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *dplyr: A Grammar of Data Manipulation*. Retrieved from <https://cran.r-project.org/package=dplyr> (<https://cran.r-project.org/package=dplyr>)

Wickham, H., & Henry, L. (2020). *tidyverse: Tidy Messy Data*. Retrieved from <https://cran.r-project.org/package=tidyr> (<https://cran.r-project.org/package=tidyr>)

Yee, T. W. (2019). *VGAM: Vector Generalized Linear and Additive Models*. Retrieved from <https://cran.r-project.org/package=VGAM> (<https://cran.r-project.org/package=VGAM>)

Yee, T. W., Stoklosa, J., & Huggins, R. M. (2015). The VGAM Package for Capture-Recapture Data Using the Conditional Likelihood. *Journal of Statistical Software*, 65(5), 1–33. Retrieved from <http://www.jstatsoft.org/v65/i05/> (<http://www.jstatsoft.org/v65/i05/>)