Survival Analysis: Comparison of Aspirin and Heparin in IST-3 Dataset

DATA 621: Advanced Statistics Modelling Final Project

Maruthi Kumar Mutnuri

Atlanta Liu

Overview of Stroke

Third-leading cause of death in Canada

Two Main Types:

Ischemic (blockage)

Hemorrhagic (rupture)

Anti-platelets: **Aspirin**

Anticoagulants: **Heparin**

Research Question & Hypothesis

- Is there a significant difference in survival rates between individuals who were randomized to take either Aspirin or a Heparin dosage?
  - How does *age and sex* impact the survival rates between patients who take Aspirin only or Heparin only?
  - What *other variables* significantly contribute to the survival rate between each treatment group?

- Aspirin patients are expected to show an improvement in survival rates in comparison to the control. Heparin is not expected to show any improvements in survivability.

Methods

International Strokes Trial (IST-3) Dataset

- n = 19,435 patients with symptoms of stroke
  - 984 patients in pilot phase
- Outcomes of patients recorded at the 6-month period
- Patient eligibility into the randomized clinical trial was based on physician criterion:
  - Should Aspirin or Heparin be administered?

Variable Definition

Main Variables of Interest

- **Groups**:
  - Control
  - Aspirin (300mg)
  - Low Heparin (5000 IU)
  - High Heparin (12,5000 IU)
- **TD** = Time of Death or Censoring (days)
  - If death occurs beyond 180 days, they are censored
- **DIED** = Status of death (0 = did not die, 1 = died)

Predictive Variables (patient characteristics)

- **AGE** (years)
- **SEX** (Male, Female)
- **RATRIAL** = Atrial fibrillation (Yes or No)
  - 984 rows were removed since RATRIAL was not recorded in pilot study
- **RSBP** = Systolic Blood Pressure (mmHg)
- **RVISINF** = Visible Infarcts (Yes or No)

Survival Analysis Procedures

- Data Exploration
- Kaplan Meier Plot
- Log Rank Test
- Cox Proportional Hazards Modelling:
  - Interaction Terms & Model Reduction
  - Likelihood Ratio Test
- Assumptions:
  - Proportional Hazard (Schoenfeld Plot of Residuals)
  - Linearity in Covariates (Martingale Plot of Residuals)
- Stratified Cox Modelling
- Hazard Ratios & Interpretations

Results

Data Exploration

- Across 19,435 patients, sample is largely representative of the older population between ages 70 (n = 12177, 63%)
- Randomization into four treatment groups:
  - Control (n = 4860, 25%)
  - Aspirin (n = 4858, 25%)
  - Low Heparin (n = 2429, 12%)
  - High Heparin (n = 2426, 12%)

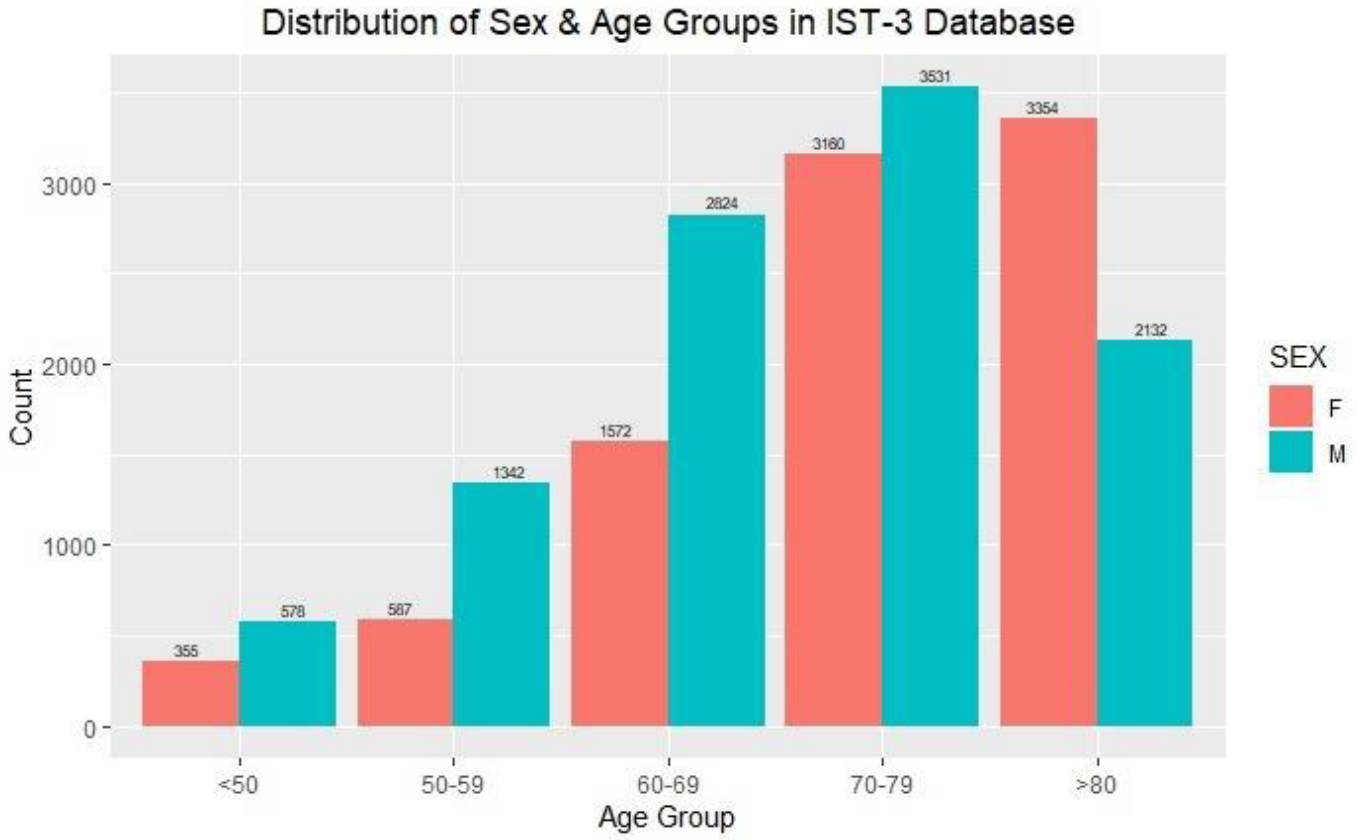
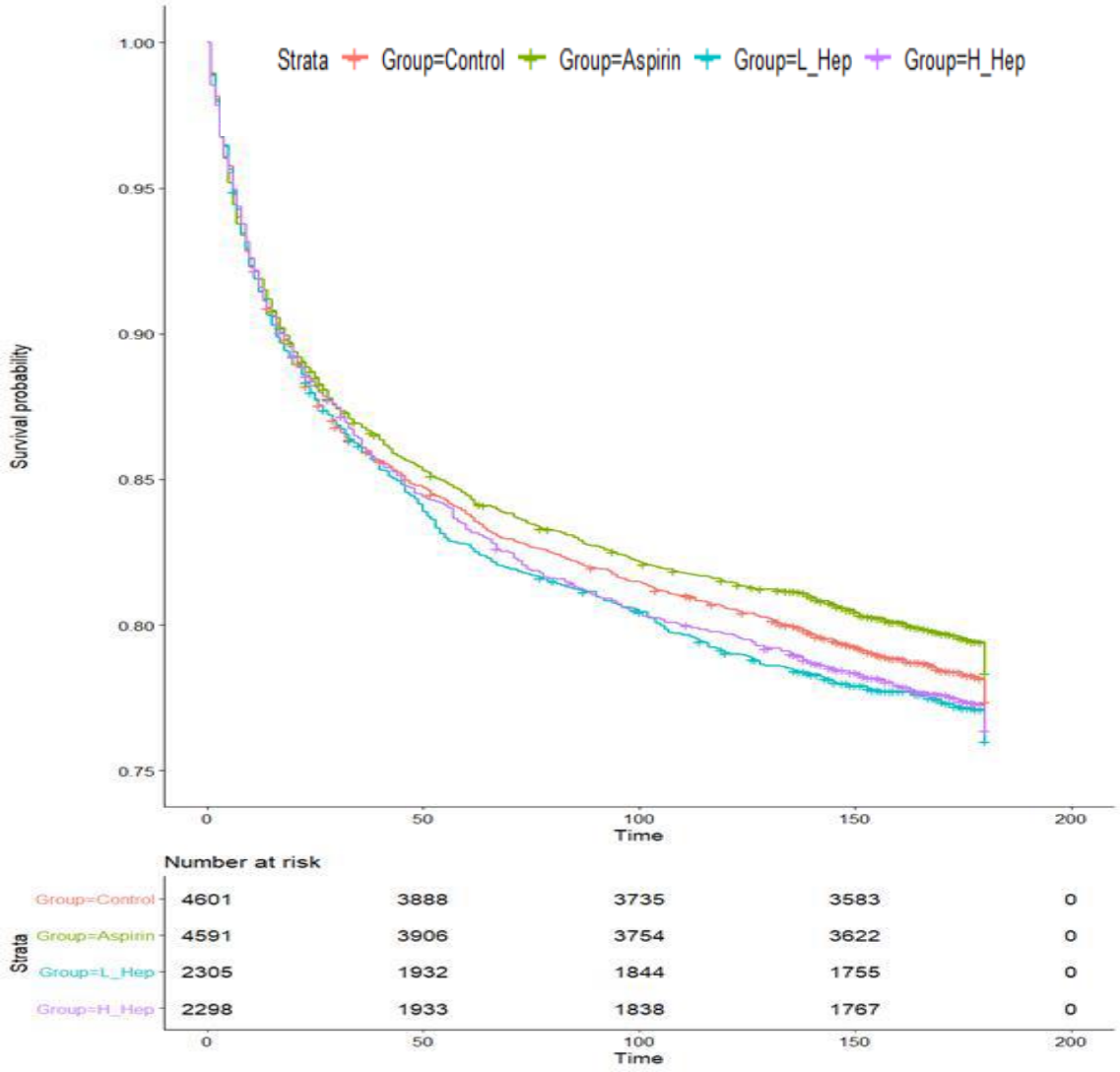


Figure 1: Distribution of age and sex in patients suffering from re-occurring stroke from IST-3 dataset

Kaplan Meier Curves



Survival curves for each treatment are plotted over the course of 180 days, along with a risk table below

Kaplan Meier Curve:

- Large initial overlap in survival curves indicate that the proportionality assumption may not be met
- Survival curve for Aspirin seems to be slightly better than control group, particularly near 180 days

Log-Rank Test

```r
survdiff(Surv(TD, DIED)~Group, data = data)
```

```
Call:
survdiff(formula = Surv(TD, DIED) ~ Group, data = data)

                N Observed Expected (O-E)^2/E (O-E)^2/V
Group=Control 4601     1030     1031   0.00108   0.00162
Group=Aspirin 4591      979     1033   2.77983   4.19296
Group=L_Hep   2305      545      513   1.97286   2.37600
Group=H_Hep   2298      536      513   1.01388   1.22105

 Chisq= 5.8  on 3 degrees of freedom, p= 0.1
```

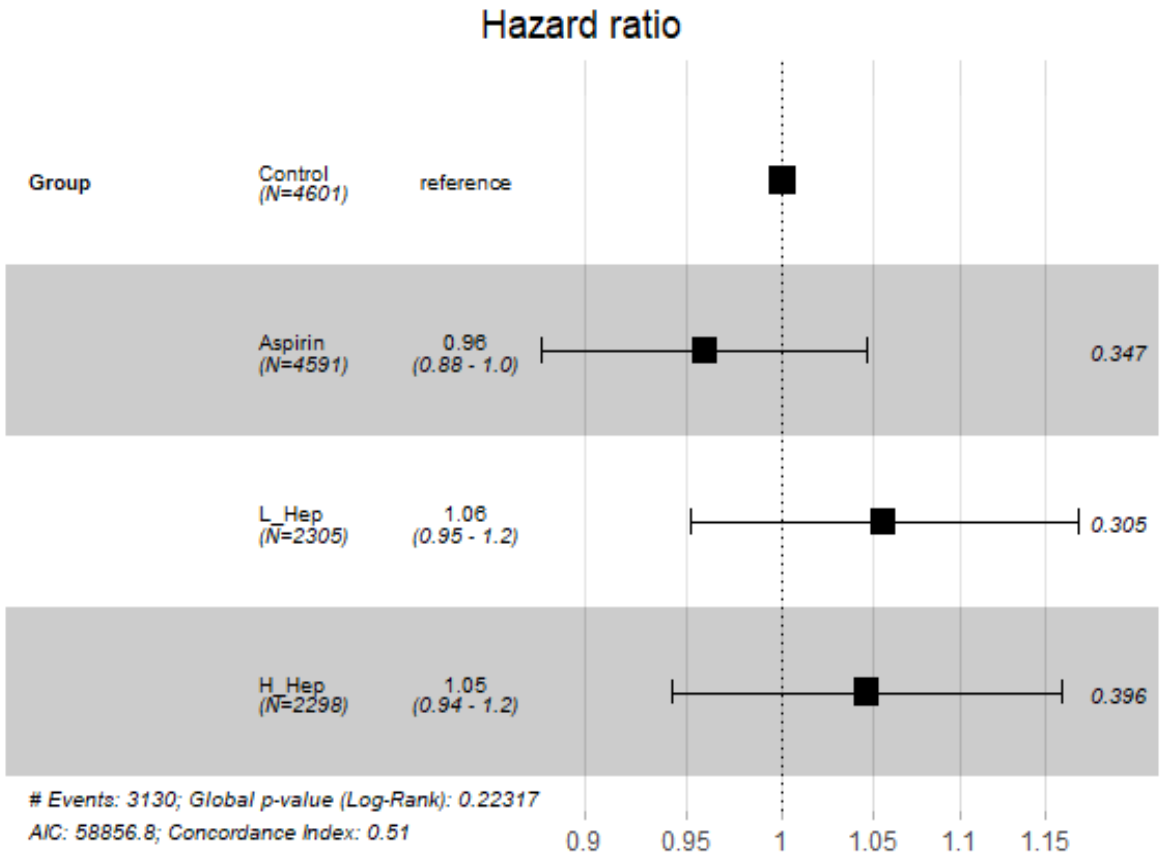
Log-rank test output compares the survival rates for all four treatment groups. Results show that there is not a significant difference between two or more groups.

Cox Proportional Hazards Base Model

Variables	coef	exp(coef)	se(coef)	z	Pr(>|z|)
Group (Aspirin)	-0.04164	0.95922	0.04428	-0.94	0.347
Group (Low Heparin)	0.05416	1.05565	0.05281	1.026	0.305
Group (High Heparin)	0.04497	1.046	0.05297	0.849	0.396

Model Equation:

$$log(h_i) = log(h_0)\ (\beta_1 X_{Group}\ )$$



Cox Proportional Hazards Full Model

Model Equation:

$$log(h_i) = log(h_0) \beta_1 X_{Group} + \beta_2 X_{Age} + \beta_3 X_{Sex\ =\ M} + \beta_4 X_{RSBP} + \beta_5 X_{RATRIAL\ =\ Y} + \beta_6 X_{RVISFINF=Y} + \beta_7 X_{Group \times Age} + \beta_8 X_{Group \times Sex} + \beta_9 X_{Group \times RSBP} + \beta_{10} X_{Group \times RATRIAL\ =Y} + \beta_{11} X_{Group \times RVISINF=Y}$$

Variables	coef	exp(coef)	se(coef)	z	Pr(>|z|)	Significance
GroupAspirin	1.061712	2.8913179	0.478145	2.22	0.0264	*
GroupL_Hep	-0.61647	0.5398462	0.58227	-1.059	0.2897	
GroupH_Hep	0.754326	2.1261785	0.582032	1.296	0.195	
AGE	0.057812	1.0595162	0.003595	16.081	2.00E-16	***
SEXM	0.027474	1.0278546	0.064033	0.429	0.6679	
RSBP	-0.0027	0.9973014	0.001142	-2.366	0.018	*
RATRIALY	0.56179	1.7538091	0.0695	8.083	6.31E-16	***
RVISINFY	0.36817	1.4450883	0.063391	5.808	6.33E-09	***
GroupAspirin:AGE	-0.0102	0.9898499	0.00504	-2.024	0.043	*
GroupL_Hep:AGE	0.00422	1.0042286	0.006087	0.693	0.4882	
GroupH_Hep:AGE	-0.00639	0.9936286	0.006116	-1.045	0.2959	
GroupAspirin:SEXM	-0.08124	0.92197	0.091791	-0.885	0.3761	
GroupL_Hep:SEXM	0.060437	1.0623004	0.108162	0.559	0.5763	
GroupH_Hep:SEXM	-0.00883	0.991209	0.109489	-0.081	0.9357	
GroupAspirin:RSBP	-0.0019	0.9980989	0.001646	-1.156	0.2476	
GroupL_Hep:RSBP	0.002167	1.0021693	0.001961	1.105	0.2692	
GroupH_Hep:RSBP	-0.00094	0.9990631	0.001997	-0.469	0.6388	
GroupAspirin:RATRIALY	0.045002	1.0460303	0.098795	0.456	0.6487	
GroupL_Hep:RATRIALY	-0.15578	0.8557442	0.11996	-1.299	0.1941	
GroupH_Hep:RATRIALY	0.010476	1.0105312	0.119522	0.088	0.9302	
GroupAspirin:RVISINFY	-0.02445	0.9758504	0.090899	-0.269	0.788	
GroupL_Hep:RVISINFY	-0.00805	0.9919832	0.1079	-0.075	0.9405	
GroupH_Hep:RVISINFY	-0.14373	0.8661232	0.110076	-1.306	0.1917	

Cox Proportional Hazards Reduced Model (1)

Model Equation:

$$log(h_i) = log(h_0)\,(\beta_1 X_{Group} + \beta_2 X_{Age} + \beta_3 X_{Sex\,=\,M} + \beta_4 X_{RSBP} + \beta_5 X_{RATRIAL\,=\,Y} + \beta_6 X_{RVISFINF=Y} + \beta_7 X_{Group\times Age})$$

Variables	coef	exp(coef)	se(coef)	z	Pr(>|z|)	Significance
GroupAspirin	0.608978	1.8385521	0.370209	1.645	0.1	.
GroupL_Hep	-0.1427	0.8670139	0.458286	-0.311	0.7555	
GroupH_Hep	0.498204	1.6457621	0.450767	1.105	0.2691	
AGE	0.057614	1.0593063	0.003457	16.666	2.00E-16	***
SEXM	0.010551	1.0106066	0.036888	0.286	0.7749	
RSBP	-0.00311	0.9968945	0.000667	-4.662	3.13E-06	***
RATRIALY	0.553481	1.7392964	0.040093	13.805	2.00E-16	***
RVISINFY	0.33507	1.398038	0.036741	9.12	2.00E-16	***
GroupAspirin:AGE	-0.00866	0.9913809	0.004747	-1.823	0.0682	.
GroupL_Hep:AGE	0.002388	1.0023912	0.005845	0.409	0.6828	
GroupH_Hep:AGE	-0.00571	0.9943063	0.005804	-0.984	0.3252	

Model	Log Likelihood	Chisq	Df	P(>|Chi|)
Full Model	-28739			
Reduced Model (1)	-28744	11.27	12	0.506

Likelihood Ratio Test

Cox Proportional Hazards Reduced Model (2)

Model Equation:

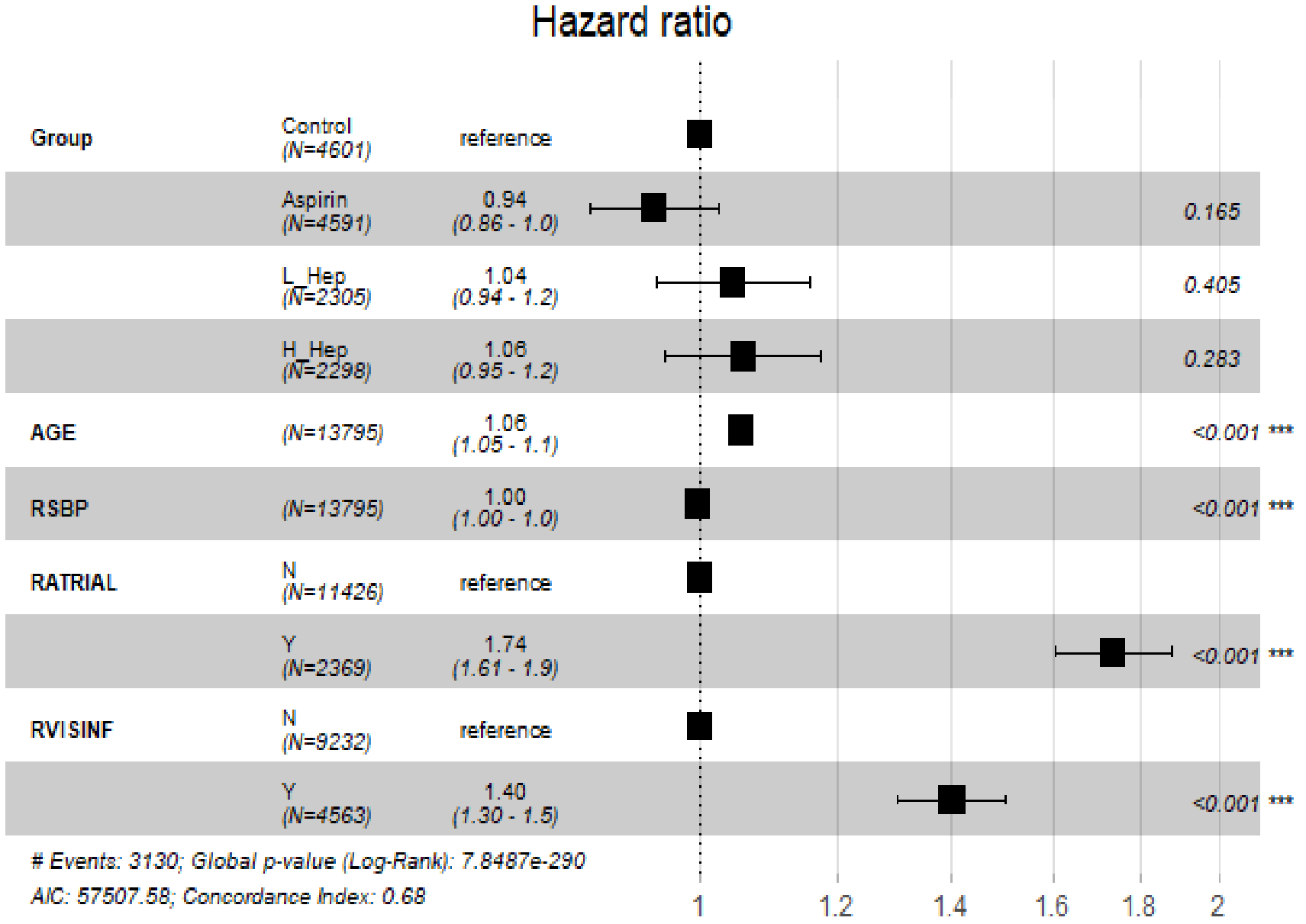
$$log(h_i) = log(h_0)\ (\beta_1 X_{Group} + \beta_2 X_{Age} + \beta_3 X_{RSBP} + \beta_4 X_{RATRIAL} = Y + \beta_5 X_{RVISFINF=Y})\ )$$

Variables	coef	exp(coef)	se(coef)	z	Pr(>|z|)	Significance
GroupAspirin	-0.06145	0.9403968	0.044288	-1.388	0.165	
GroupL_Hep	0.044005	1.044988	0.052816	0.833	0.405	
GroupH_Hep	0.056924	1.058575	0.052976	1.075	0.283	
AGE	0.054077	1.0555655	0.001996	27.093	2.00E-16	***
RSBP	-0.00314	0.9968629	0.000666	-4.72	2.36E-06	***
RATRIALY	0.552014	1.7367474	0.040064	13.778	2.00E-16	***
RVISINFY	0.336335	1.3998074	0.036717	9.16	2.00E-16	***

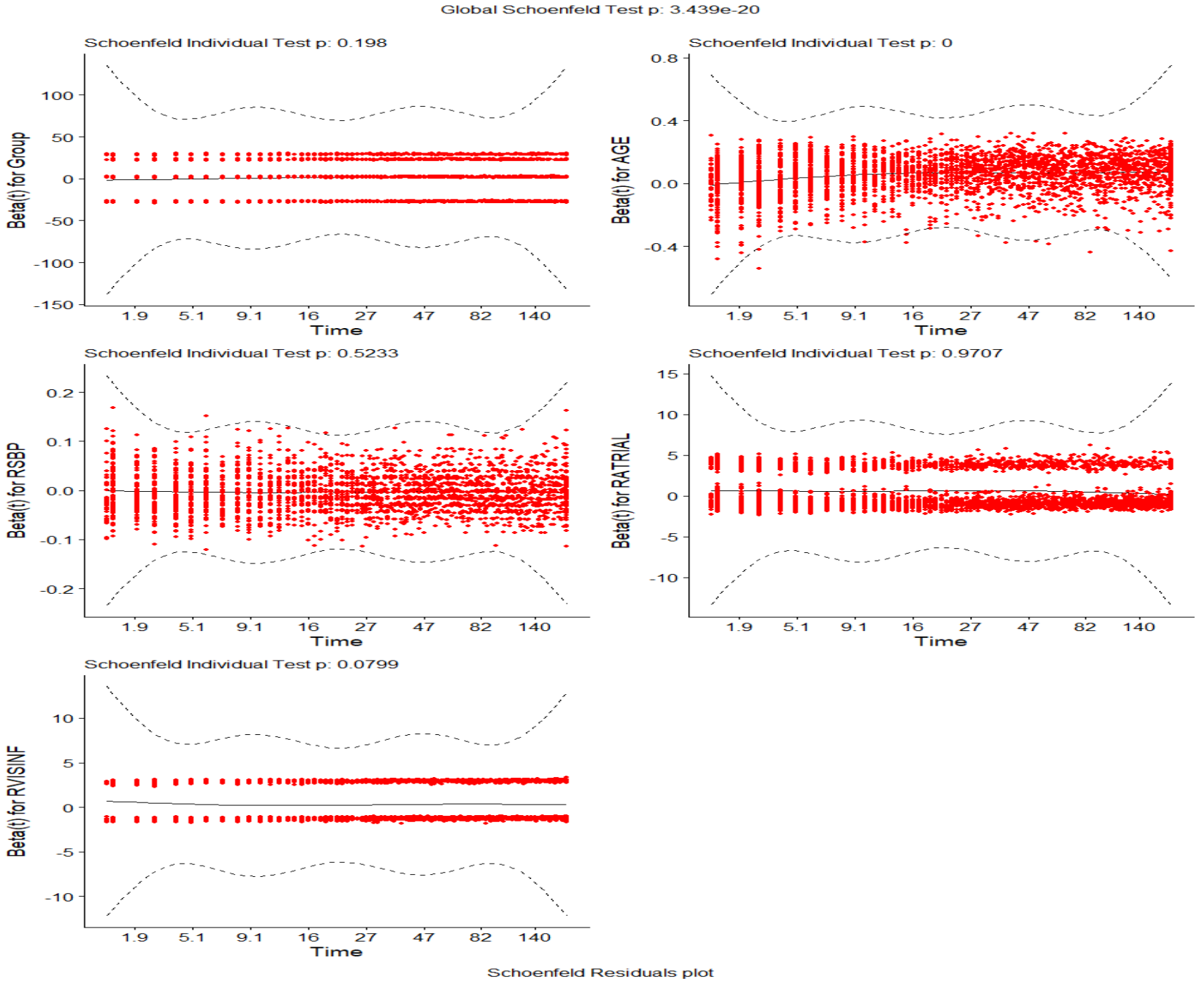
Model	Log Likelihood	Chisq	Df	P(>|Chi|)
Reduced Model (1)	-28744			
Reduced Model (2)	-28747	5.3222	4	0.2558

Likelihood Ratio Test

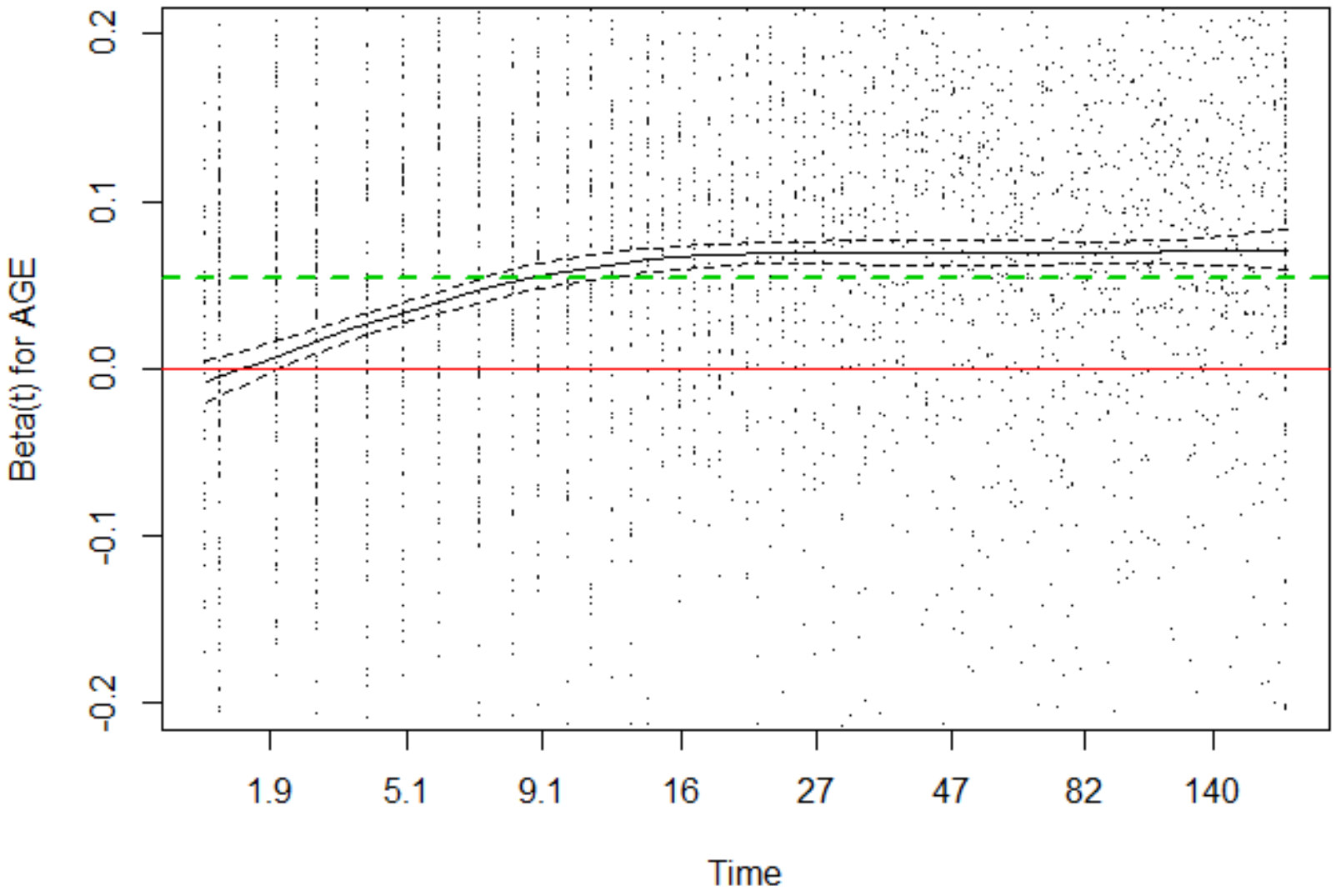
Final Reduced Model



Assumptions:
1. Testing Proportional Hazards Assumption



PH Assumption AGE



There is a noticeable change in slope before 16 days for the effect of age so stratifying age might help meeting the assumption

Using SurvSplit to Stratify Data

```
data_tf <- survSplit(Surv(TD, DIED) ~ ., data= data, cut=c(4, 16), start="tstart", zero=0, episode= "tgroup", id="
id")
head(data_tf)
```

Group	SEX	AGE	RSBP	RATRIAL	RVISINF	TD	DIED
<fctr>	<fctr>	<int>	<int>	<fctr>	<fctr>	<dbl>	<dbl>
Control	F	73	120	N	N	8	1
Control	M	74	160	N	Y	180	0
Control	M	80	200	Y	Y	180	0
Control	M	61	180	N	Y	180	0
Control	M	70	135	N	N	180	0
Control	M	62	170	N	N	180	0
Control	F	87	170	N	N	17	1
Control	M	73	175	N	N	180	0
Control	M	71	150	N	N	180	0
Control	M	63	180	Y	N	180	0

Group	SEX	AGE	RSBP	RATRIAL	RVISINF	id	tstart	TD	DIED	tgroup
<fctr>	<fctr>	<int>	<int>	<fctr>	<fctr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
Control	F	73	120	N	N	1	0	4	0	1
Control	F	73	120	N	N	1	4	8	1	2
Control	M	74	160	N	Y	2	0	4	0	1
Control	M	74	160	N	Y	2	4	16	0	2
Control	M	74	160	N	Y	2	16	180	0	3
Control	M	80	200	Y	Y	3	0	4	0	1
Control	M	80	200	Y	Y	3	4	16	0	2
Control	M	80	200	Y	Y	3	16	180	0	3
Control	M	61	180	N	Y	4	0	4	0	1
Control	M	61	180	N	Y	4	4	16	0	2

```
coxph.fit_FINAL <- coxph(Surv(tstart, TD, DIED) ~ Group+RSBP+RATRIAL+RVISINF + AGE:strata(tgroup), data=data_tf, t
ies ="breslow")
summary(coxph.fit_FINAL)
```

Final SurvSplit Model

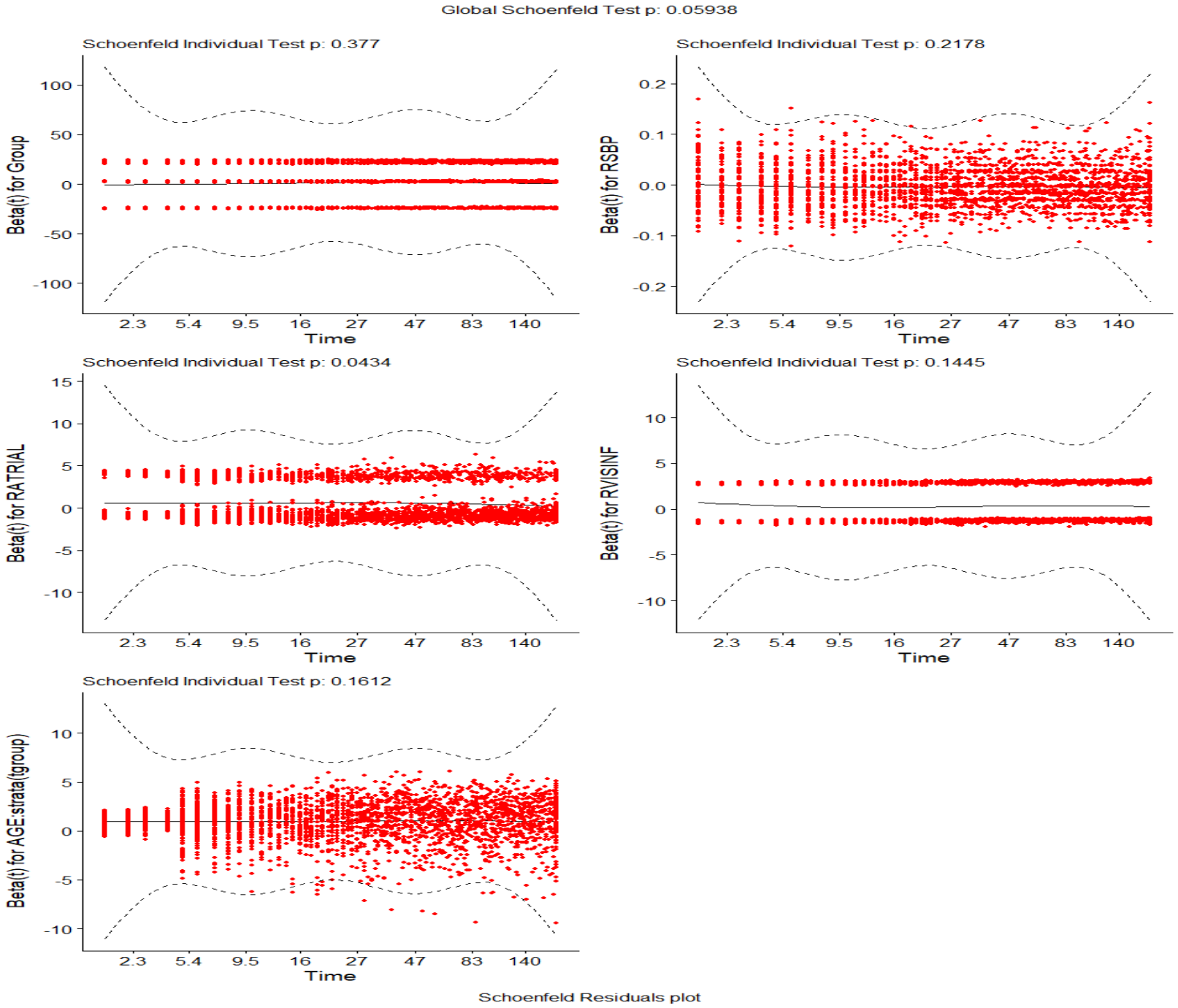
Model Equation:

$$log(h_i) = \beta_1 X_{Control} + \beta_2 X_{RSBP} + \beta_3 X_{RATRIAL=Y} + \beta_4 X_{RVISINF=Y} + \beta_5 X_{AGE:tgroup=1} + \beta_6 X_{AGE:tgroup=2} + \beta_7 X_{AGE:tgroup=3}$$

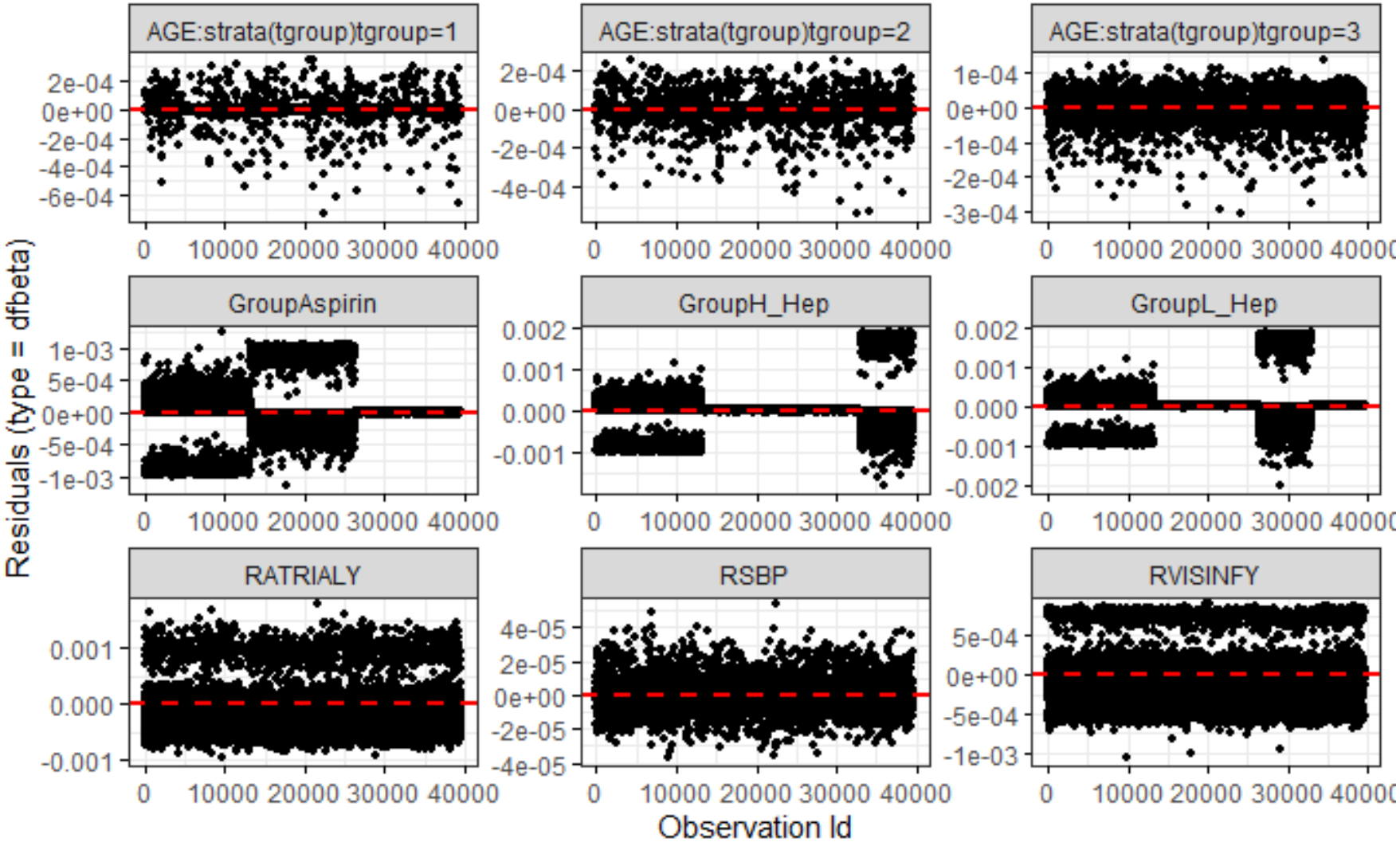
Aspirin group is borderline significant at 95% of CI

Variables	coef	exp(coef)	se(coef)	z	Pr(>|z|)
GroupAspirin	-0.073757	0.928898	0.044645	-1.652	0.09852
GroupL_Hep	0.05114	1.052471	0.052977	0.965	0.33438
GroupH_Hep	0.058118	1.05984	0.053268	1.091	0.27525
RSBP	-0.002975	0.99703	0.000669	-4.446	8.74E-06
RATRIALY	0.550428	1.733996	0.040332	13.647	2.00E-16
RVISINFY	0.339934	1.404854	0.036961	9.197	2.00E-16
AGE:strata(tgroup)tgroup=1	0.013189	1.013276	0.00406	3.249	0.00116
AGE:strata(tgroup)tgroup=2	0.055798	1.057384	0.00387	14.419	2.00E-16
AGE:strata(tgroup)tgroup=3	0.069314	1.071773	0.002766	25.062	2.00E-16

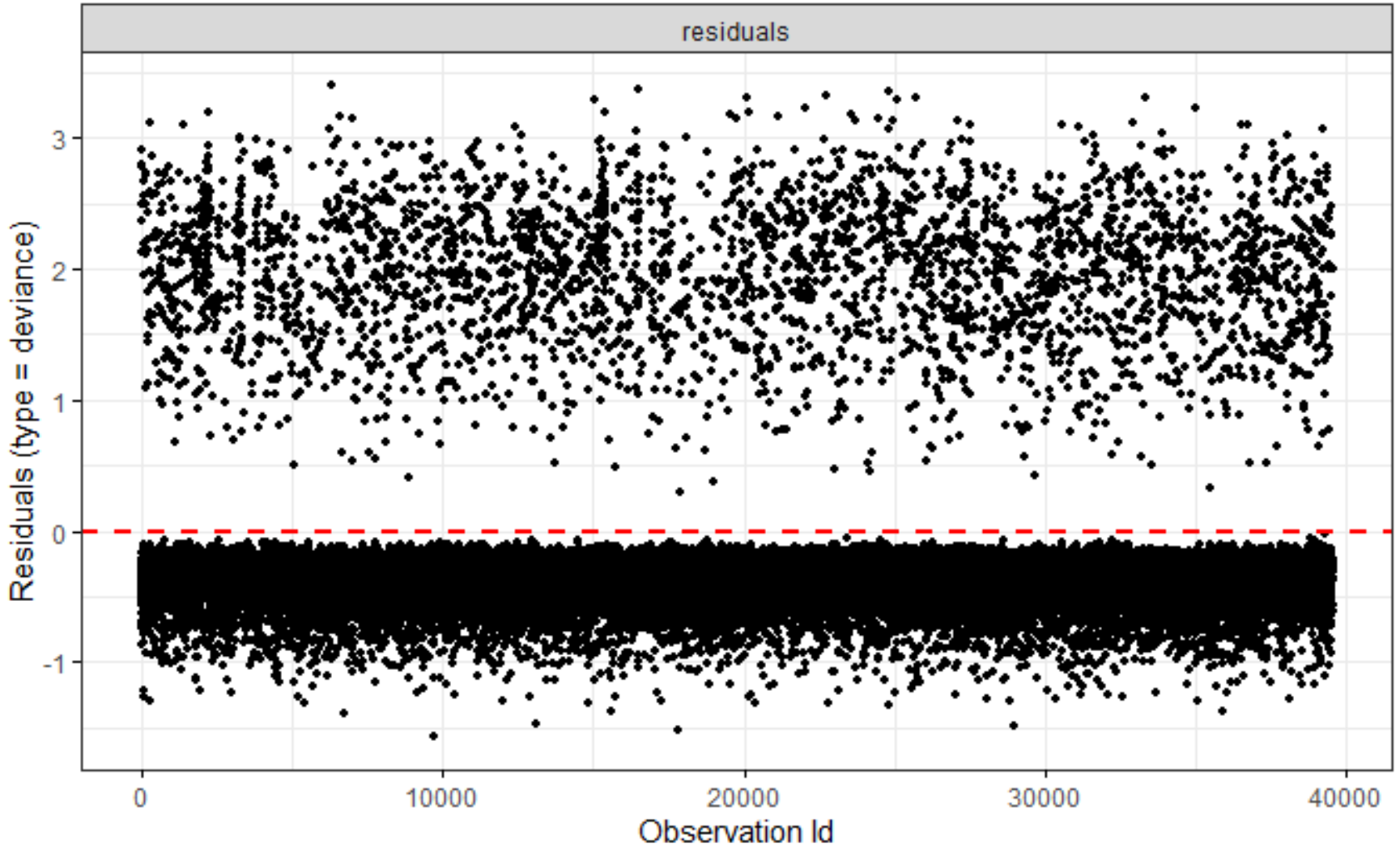
Assumptions:
1. Testing Proportional Hazards



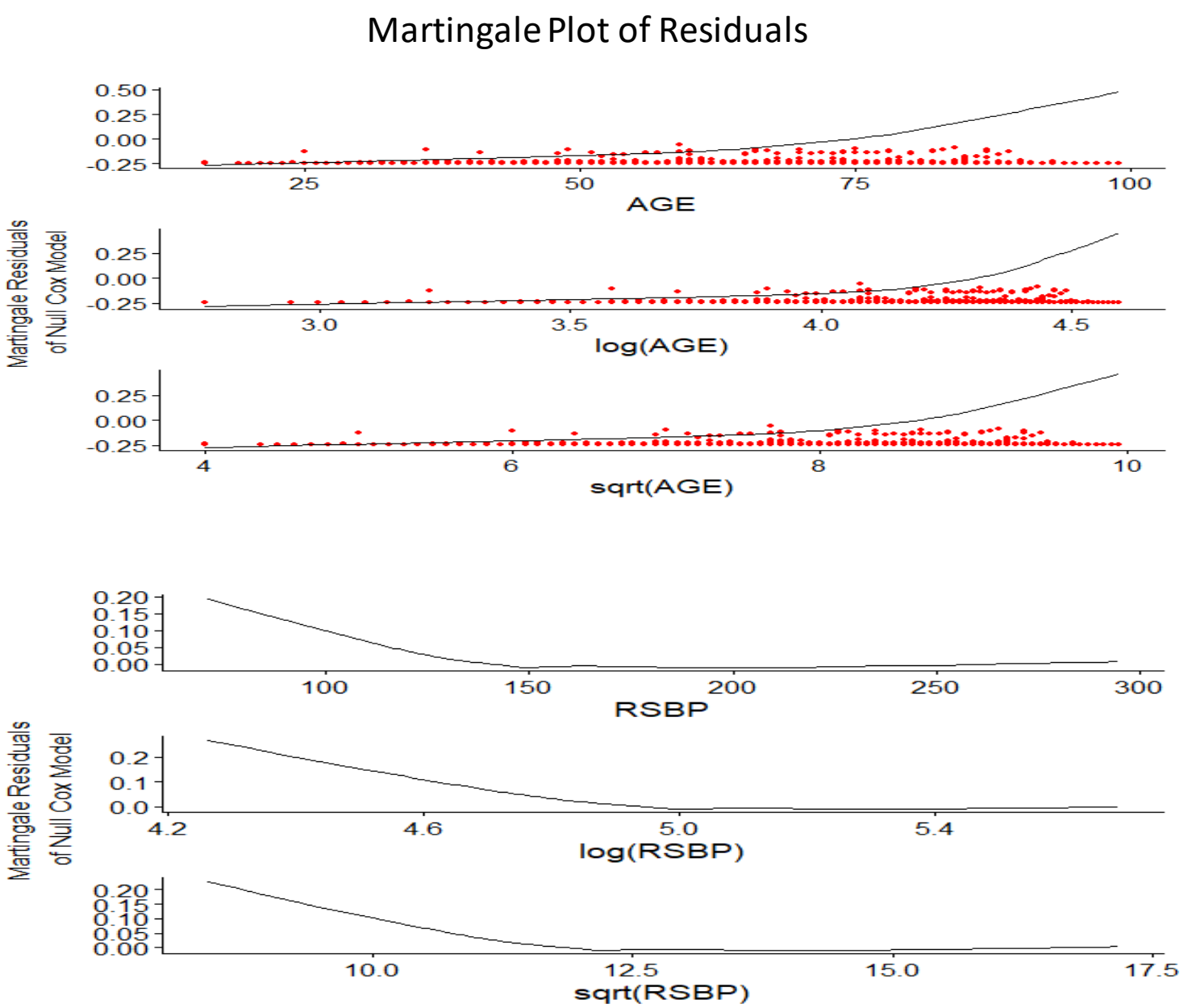
Assumptions:
2. Testing
Outliers



Assumptions:
2. Testing
Outliers



Assumptions:
3. Testing
Non-Linearity



Interpretation

Hazard for Aspirin is slightly lower than the low heparin and high heparin groups.

Variables	exp(coef)	exp(-coef)	lower .95	upper .95
GroupAspirin	0.9289	1.0765	0.8511	1.0138
GroupL_Hep	1.0525	0.9501	0.9487	1.1676
GroupH_Hep	1.0598	0.9435	0.9548	1.1765
RSBP	0.997	1.003	0.9957	0.9983
RATRIALY	1.734	0.5767	1.6022	1.8766
RVISINFY	1.4049	0.7118	1.3067	1.5104
AGE:strata(tgroup)tgroup=1	1.0133	0.9869	1.0052	1.0214
AGE:strata(tgroup)tgroup=2	1.0574	0.9457	1.0494	1.0654
AGE:strata(tgroup)tgroup=3	1.0718	0.933	1.066	1.0776

Higher Order Model

Variables	coef	exp(coef)	se(coef)	z	Pr(>|z|)	Significance
GroupAspirin	-7.32E-02	9.29E-01	4.47E-02	-1.64	0.10094	
GroupL_Hep	5.42E-02	1.06E+00	5.30E-02	1.023	0.30642	
GroupH_Hep	6.12E-02	1.06E+00	5.33E-02	1.148	0.25081	
RSBP	-2.37E-02	9.77E-01	5.01E-03	-4.725	2.30E-06	***
I(RSBP^2)	6.29E-05	1.00E+00	1.50E-05	4.184	2.87E-05	***
RATRIALY	5.53E-01	1.74E+00	4.03E-02	13.698	2.00E-16	***
RVISINFY	3.40E-01	1.41E+00	3.70E-02	9.209	2.00E-16	***
AGE:strata(tgroup)tgroup=1	1.33E-02	1.01E+00	4.05E-03	3.284	0.00102	**
AGE:strata(tgroup)tgroup=2	5.57E-02	1.06E+00	3.86E-03	14.426	2.00E-16	***
AGE:strata(tgroup)tgroup=3	6.91E-02	1.07E+00	2.76E-03	25.046	2.00E-16	***

Variables	exp(coef)	exp(-coef)	lower .95	upper .95
GroupAspirin	0.9294	1.076	0.8515	1.0144
GroupL_Hep	1.0557	0.9473	0.9516	1.1712
GroupH_Hep	1.0631	0.9407	0.9577	1.1801
RSBP	0.9766	1.024	0.9671	0.9862
I(RSBP^2)	1.0001	0.9999	1	1.0001
RATRIALY	1.7378	0.5755	1.6057	1.8807
RVISINFY	1.4055	0.7115	1.3073	1.5111
AGE:strata(tgroup)tgroup=1	1.0134	0.9868	1.0054	1.0214
AGE:strata(tgroup)tgroup=2	1.0572	0.9459	1.0493	1.0653
AGE:strata(tgroup)tgroup=3	1.0715	0.9333	1.0657	1.0773

Discussion

Study Limitations

- Large overlap in survival curve and issues with proportionality assumption
  - Alternative modelling approach involved stratifying age and cutting time intervals
- Higher order was not directly accounted for in our SurvSplit model
  - Schoenfeld residuals can not be plotted
  - RSBP showed significant higher order terms
- Systolic blood pressure was not binned
  - Not certain if effect size is the same across each interval
  - Extreme ends of blood pressure (<140 and >200) reduced risk of death

What Insight Was Gained?

- The main contributors to survival rates in patients with stroke were
  - Age
  - Systolic Blood Pressure
  - Atrial Fibrillation
  - Visible Infarcts
- Sex was not a significant predictor of one's survival rate
- There was not a significant difference between any treatment groups. However, aspirin was borderline significant.

Takeaway:

- Future modelling should avoid relying on proportionality (additive cox model)
- Physicians may want to reconsider providing heparin to patients suffering from re-occurring stroke in the future

References

1. Public Health Agency of Canada. Government of Canada [Internet]. , 2017 - Canada.ca. Government of Canada; 2019. Available from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/stroke-canada-fact-sheet.html

2. International Stroke Trial Collaborative Group. "The International Stroke Trial (IST): a randomized trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke," Lancet, vol. 349, pp. 1569-1581, 1997.

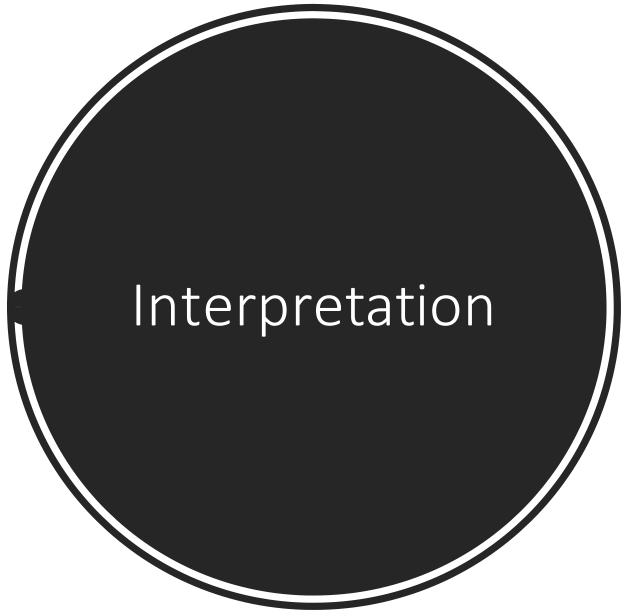
3. Hankey GJ, Berge E, Sandercock P. Anticoagulants Versus Antiplatelet Agents for Acute Ischemic Stroke. Stroke. 2003;34(6):1571–2.

4. Sandercock PA, Niewada M, Członkowska A. The International Stroke Trial database. Trials. 2011;12(1). Available from: https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-12-101#Tab2

5. Schober P, Vetter TR. Survival Analysis and Interpretation of Time-to-Event Data. Anesthesia & Analgesia. 2018;127(3):792–8. Available from: https://journals.lww.com/anesthesia-analgesia/Fulltext/2018/09000/Survival_Analysis_and_Interpretation_of.32.aspx

6. Cox Model Assumptions [Internet]. STHDA. Available from: http://www.sthda.com/english/wiki/cox-model-assumptions

Appendix: Full Interpretation of the Hazard Ratios



Variables	Interpretation
GroupAspirin	The hazard ratio for patients in the Aspirin group to control group is 0.9289 (CI 95% = 0.8511 – 1.0138).
GroupL_Hep	The hazard ratio for patients in the low heparin group to control group is 1.0525 (CI95% = 0.9487 – 1.1676).
GroupH_Hep	The HR for patients in the high heparin group to control group is 1.0598 (CI95% = 0.9548 – 1.1765).
RSBP	The HR for systolic blood pressure (RSBP) is 0.997 (CI95% = 0.9957 – 0.9983). As blood pressure increases by 1 mm Hg then the odds of dying decreases by 0.3% (1-0.997 = 0.003 = 0.3%).
RATRIALY	The HR for RATRIAL is 1.734 (CI95% = 1.6022 – 1.8766). It indicates that the relative risk of death for patients with atrial fibrillation is 1.734 times higher than patients without it.
RVISINFY	The HR for RVISINF is 1.4049 (CI95% = 1.3067 – 1.5104). So, the relative risk is 1.4049 times more for those who do have a visible infarct in comparison to those who don't
AGE:strata(tgroup)tgroup=1	THE HR for age in the first time group (0 – 4 days) is 1.0133 (CI95% = 1.0052 – 1.0214). As age increases by 1 year the odds of dying increases by 1.33% (1-1.0133 = 0.0133 = 1.33%).
AGE:strata(tgroup)tgroup=2	The HR for age in the second time group (4 – 16 days) is 1.0574 (CI95% = 1.0494 – 1.0654). As age increases by 1 year the odds of dying increases by 5.74% (1-1.0574 = 0.0574 = 5.74%).
AGE:strata(tgroup)tgroup=3	The HR for age in the third time group (16 – 180 days) is 1.0718 (CI95% = 1.066 – 1.0776). As age increases by 1 year the odds of dying increases by 7.18% (1-1.0718 = 0.0718 = 7.18%).