

# Inflammatory bowel disease

Nathan Constantine-Cooke

2025-08-13

## Table of contents

Introduction . . . . .	3
IBD Control-8 . . . . .	4
Crohn's disease . . . . .	4
Patient-reported flare . . . . .	4
Proportional hazards assumption test . . . . .	6
DF betas . . . . .	6
Hard flare . . . . .	8
Proportional hazards assumption test . . . . .	11
DF betas . . . . .	11
Ulcerative colitis . . . . .	13
Patient-reported flare . . . . .	13
Proportional hazards assumption test . . . . .	16
DF betas . . . . .	16
Hard flare . . . . .	18
Proportional hazards assumption test . . . . .	21
DF betas . . . . .	21
IBD Control visual analogue scale . . . . .	23
Crohn's disease . . . . .	23
Patient-reported flare . . . . .	23
Proportional hazards assumption test . . . . .	26
DF betas . . . . .	26
Hard flare . . . . .	28
Proportional hazards assumption test . . . . .	31
DF betas . . . . .	31
Ulcerative colitis . . . . .	33
Patient-reported flare . . . . .	33
Proportional hazards assumption test . . . . .	36
DF betas . . . . .	36

Hard flare . . . . .	38
Proportional hazards assumption test . . . . .	41
DF betas . . . . .	41
Biologic use . . . . .	43
Crohn's disease . . . . .	43
Patient-reported flare . . . . .	43
Proportional hazards assumption test . . . . .	46
DF betas . . . . .	46
Hard flare . . . . .	48
Proportional hazards assumption test . . . . .	51
DF betas . . . . .	51
Ulcerative colitis . . . . .	53
Patient-reported flare . . . . .	53
Proportional hazards assumption test . . . . .	56
DF betas . . . . .	56
Hard flare . . . . .	58
Proportional hazards assumption test . . . . .	61
DF betas . . . . .	61
Variables only relevant to Crohn's disease . . . . .	63
Surgery . . . . .	64
Patient-reported flare . . . . .	64
Proportional hazards assumption test . . . . .	66
DF betas . . . . .	66
Hard flare . . . . .	68
Proportional hazards assumption test . . . . .	71
DF betas . . . . .	71
Montreal location . . . . .	73
Patient-reported flare . . . . .	73
Proportional hazards assumption test . . . . .	76
DF betas . . . . .	76
Hard flare . . . . .	78
Proportional hazards assumption test . . . . .	81
DF betas . . . . .	81
Montreal L4 . . . . .	83
Patient-reported flare . . . . .	83
Proportional hazards assumption test . . . . .	86
DF betas . . . . .	86
Hard flare . . . . .	88
Proportional hazards assumption test . . . . .	91
DF betas . . . . .	91
Montreal Behaviour . . . . .	93
Patient-reported flare . . . . .	93
Proportional hazards assumption test . . . . .	96

DF betas . . . . .	96
Hard flare . . . . .	98
Proportional hazards assumption test . . . . .	101
DF betas . . . . .	101
Perianal disease . . . . .	103
Patient-reported flare . . . . .	103
Proportional hazards assumption test . . . . .	106
DF betas . . . . .	106
Hard flare . . . . .	108
Proportional hazards assumption test . . . . .	111
DF betas . . . . .	111
Variables only relevant to UC/IBDU . . . . .	113
Montreal extent . . . . .	114
Patient-reported flare . . . . .	114
Proportional hazards assumption test . . . . .	116
DF betas . . . . .	116
Hard flare . . . . .	118
Proportional hazards assumption test . . . . .	121
DF betas . . . . .	121
Mayo score . . . . .	123
Patient-reported flare . . . . .	124
Proportional hazards assumption test . . . . .	126
DF betas . . . . .	126
Hard flare . . . . .	128
Proportional hazards assumption test . . . . .	131
DF betas . . . . .	131
Reproduction and reproducibility . . . . .	134

## Introduction

```

source("Survival/utils.R")

# Setup analysis environment
analysis_setup <- setup_analysis()
paths <- analysis_setup$paths
demo <- analysis_setup$demo

demo$FC <- log(demo$FC)

flare.df <- readRDS(paste0(paths$outdir, "flares-demographics.RDS"))
flare.cd.df <- readRDS(paste0(paths$outdir, "flares-demographics-cd.RDS"))

```

```

flare.uc.df <- readRDS(paste0(paths$outdir, "flares-demographics-uc.RDS"))

cd.clin.forest <- readRDS(paste0(paths$outdir, "cd-clin-demographics.RDS"))
cd.hard.forest <- readRDS(paste0(paths$outdir, "cd-hard-demographics.RDS"))
uc.clin.forest <- readRDS(paste0(paths$outdir, "uc-clin-demographics.RDS"))
uc.hard.forest <- readRDS(paste0(paths$outdir, "uc-hard-demographics.RDS"))

```

## IBD Control-8

### Crohn's disease

#### Patient-reported flare

```

# Categorize IBD Control-8
flare.cd.df <- categorize_variable(flare.cd.df, "control_8",
                                      breaks = c(0, 13, 16),
                                      labels = c("0-12", "13-16"))

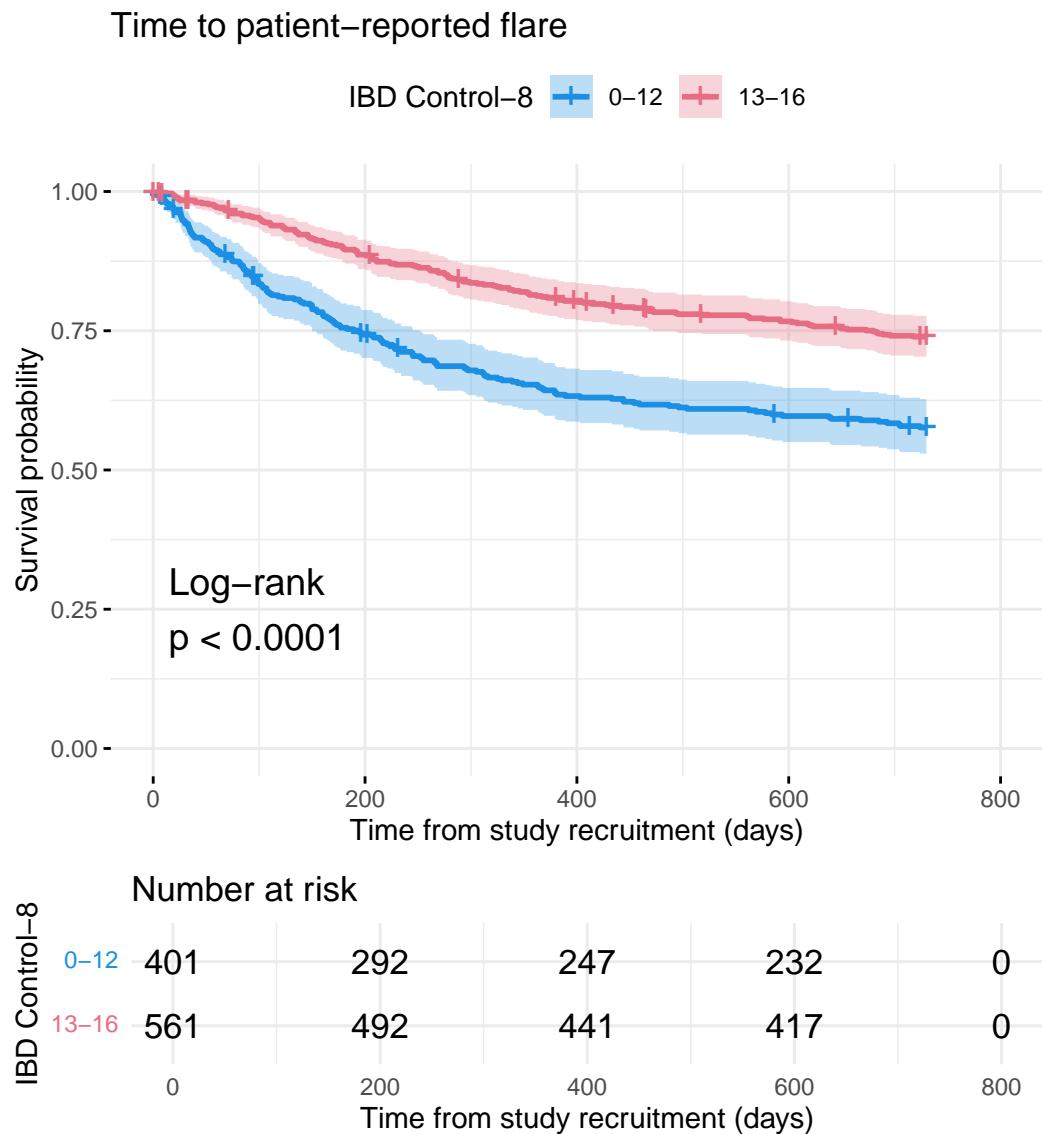
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "control_8",
  outcome_time = "softflare_time",
  outcome_event = "softflare",
  legend_title = "IBD Control-8",
  plot_base_path = "plots/cd/soft-flare/ibd/control-8",
  break_time_by = 200,
  palette = c("#1A8FE3", "#E76D83")
)

# Extract hazard ratio for continuous control_8 variable
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + IMD + cat + control_8 + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

cd.clin.forest <- rbind(cd.clin.forest, get_HR(fit.me, "control_8"))

```

```
# Display plot and model summary  
analysis_result$plot
```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.9227	1.4759	2.5049	0.0000
IMD2	0.8074	0.4922	1.3246	0.3970
IMD3	0.8700	0.5315	1.4241	0.5796
IMD4	0.8528	0.5286	1.3758	0.5141
IMD5	0.9833	0.6177	1.5655	0.9435
catFC 50-250	1.4397	1.0987	1.8866	0.0082
catFC > 250	2.4014	1.7740	3.2506	0.0000
control_8	0.8902	0.8606	0.9207	0.0000

Diagnostics:

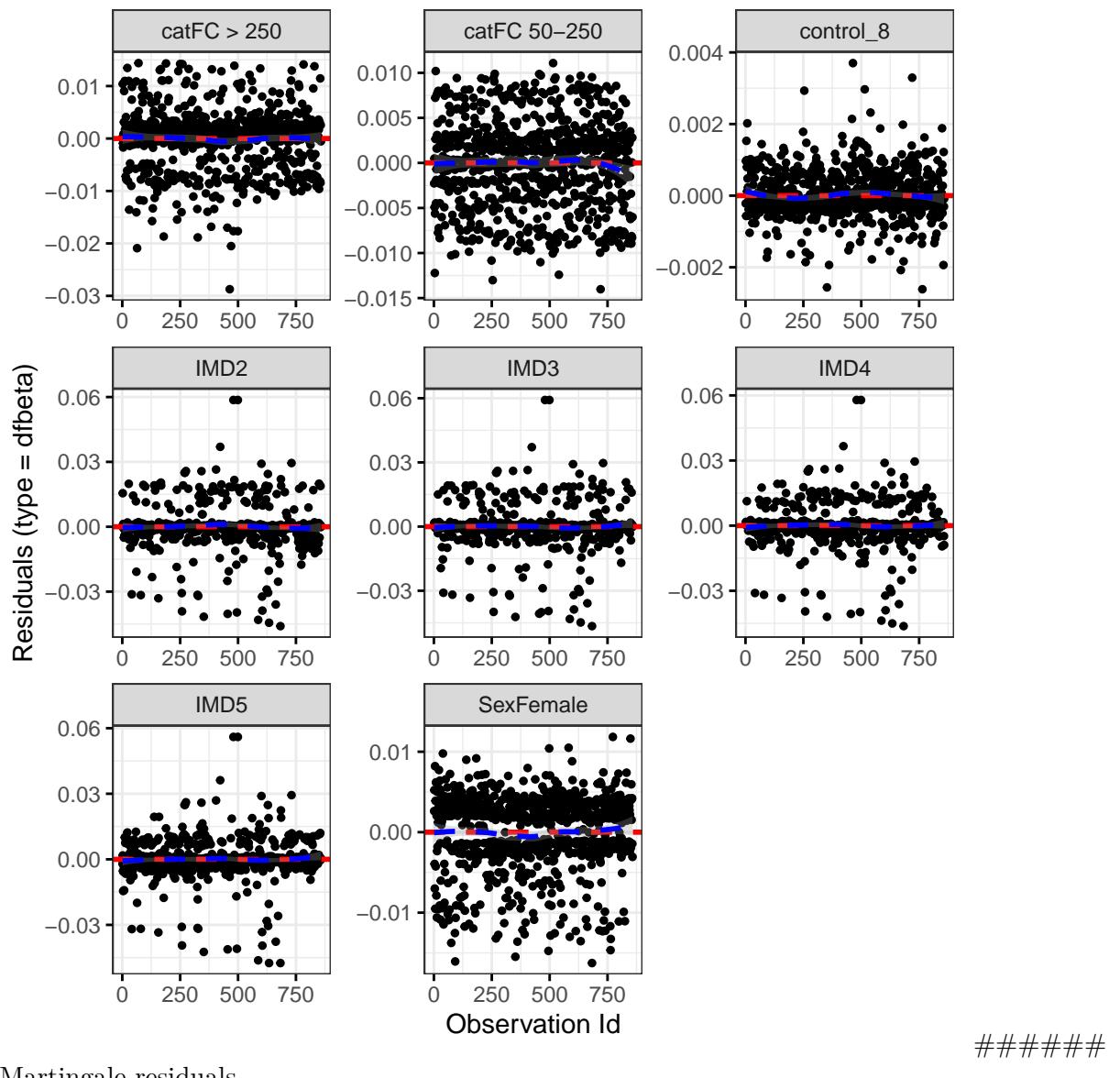
### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.0012	0.9903	0.9717
IMD	6.0925	3.9280	0.1855
cat	1.3577	1.9739	0.5009
control_8	8.4939	0.9670	0.0034
GLOBAL	15.1377	14.9466	0.4376

### DF betas

```
Warning: `gather_()` was deprecated in tidyverse 1.2.0.
  i Please use `gather()` instead.
  i The deprecated feature was likely used in the survminer package.
    Please report the issue at <https://github.com/kassambara/survminer/issues>.
```

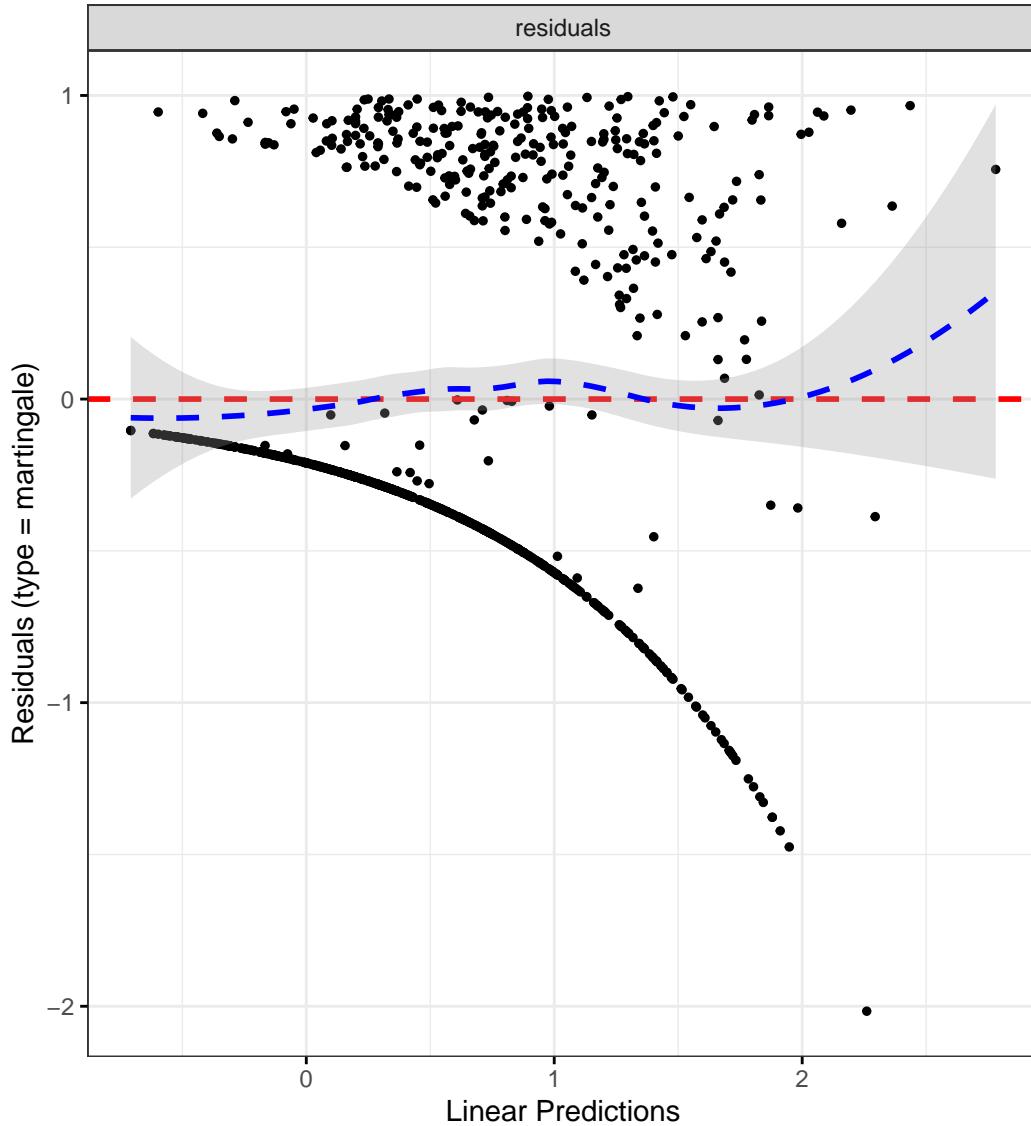
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



## Hard flare

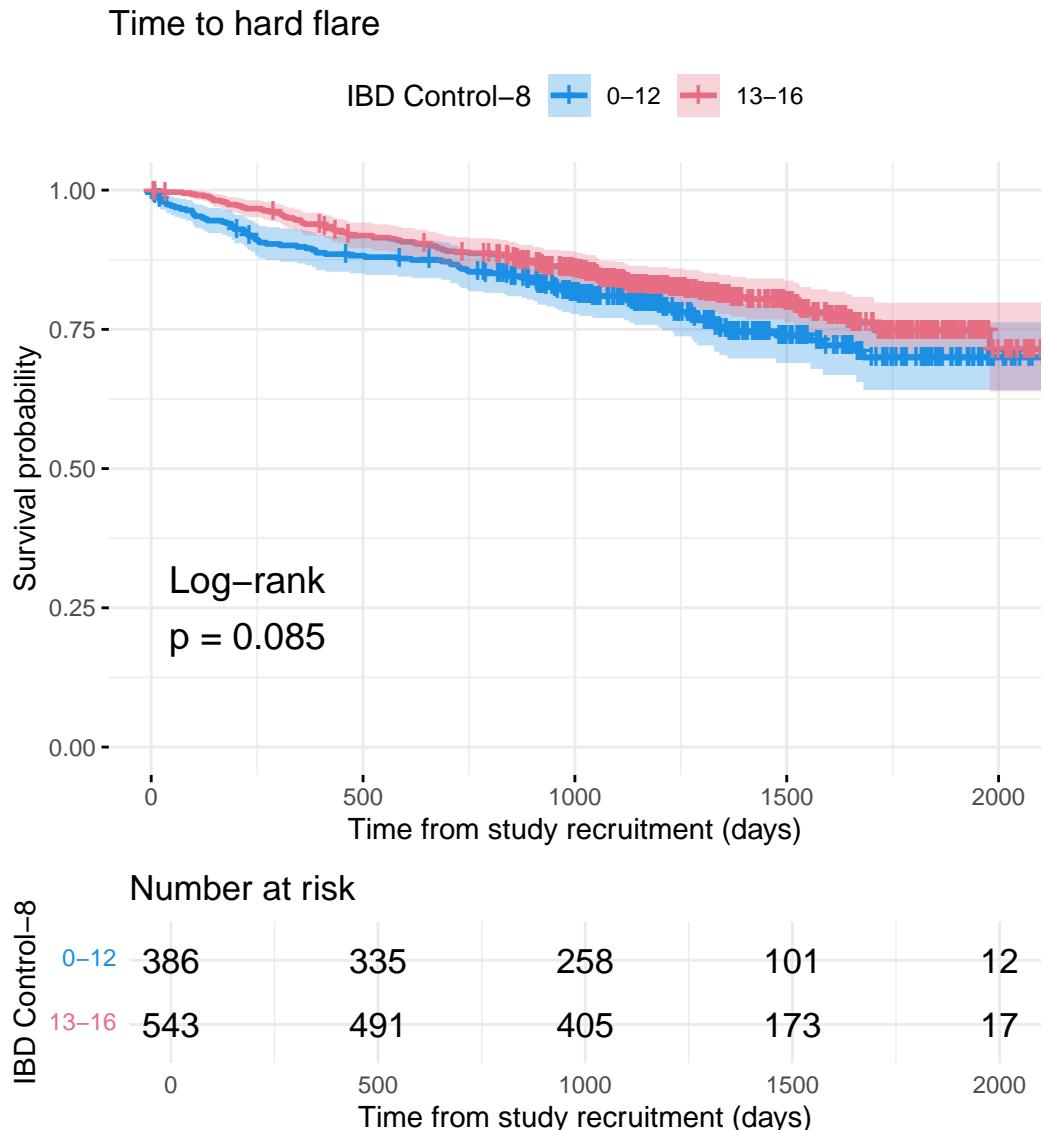
```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "control_8",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "IBD Control-8",
```

```
plot_base_path = "plots/cd/hard-flare/ibd/control-8",
break_time_by = 500,
palette = c("#1A8FE3", "#E76D83")
)

# Extract hazard ratio for continuous control_8 variable
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + IMD + cat + control_8 + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

cd.hard.forest <- rbind(cd.hard.forest, get_HR(fit.me, "control_8"))

# Display plot and model summary
analysis_result$plot
```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.3756	1.0112	1.8714	0.0423
IMD2	0.6980	0.3790	1.2855	0.2485
IMD3	0.8591	0.4697	1.5715	0.6221

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	0.7891	0.4392	1.4178	0.4282
IMD5	0.7981	0.4509	1.4124	0.4387
catFC 50-250	1.9417	1.3669	2.7584	0.0002
catFC > 250	3.5422	2.4220	5.1804	0.0000
control_8	0.9792	0.9359	1.0245	0.3621

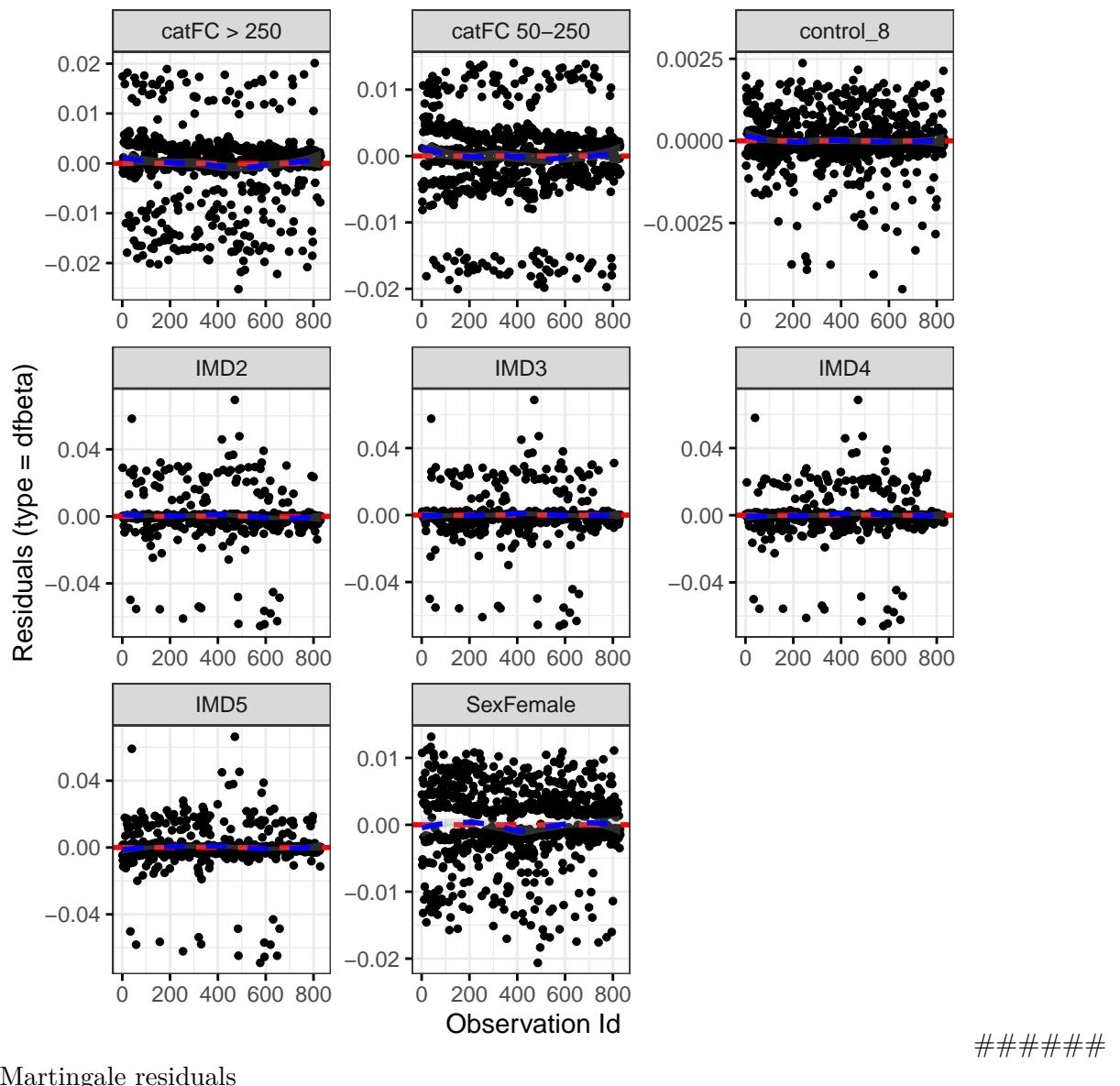
Diagnostics:

### Proportional hazards assumption test

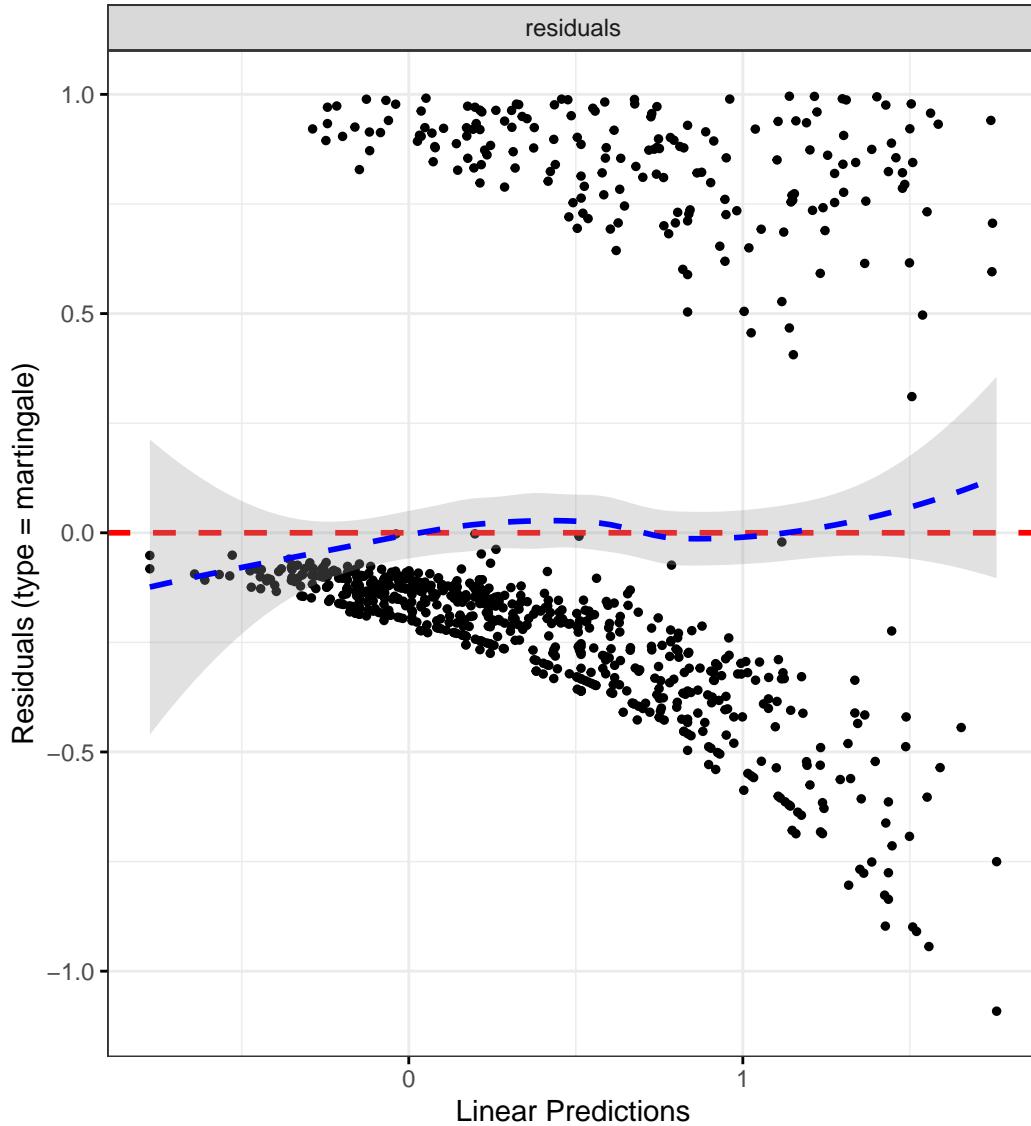
	Chi-squared statistic	DF	P-value
Sex	1.5133	0.9850	0.2149
IMD	8.1252	3.9413	0.0840
cat	7.2016	1.9860	0.0269
control_8	2.8005	0.9806	0.0918
GLOBAL	20.3170	15.1753	0.1677

### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



```
`geom_smooth()` using formula = 'y ~ x'
```



### Ulcerative colitis

#### Patient-reported flare

```
# Categorize IBD Control-8
flare.uc.df <- categorize_variable(flare.uc.df, "control_8",
                                      breaks = c(0, 13, 16),
                                      labels = c("0-12", "13-16"))
```

```

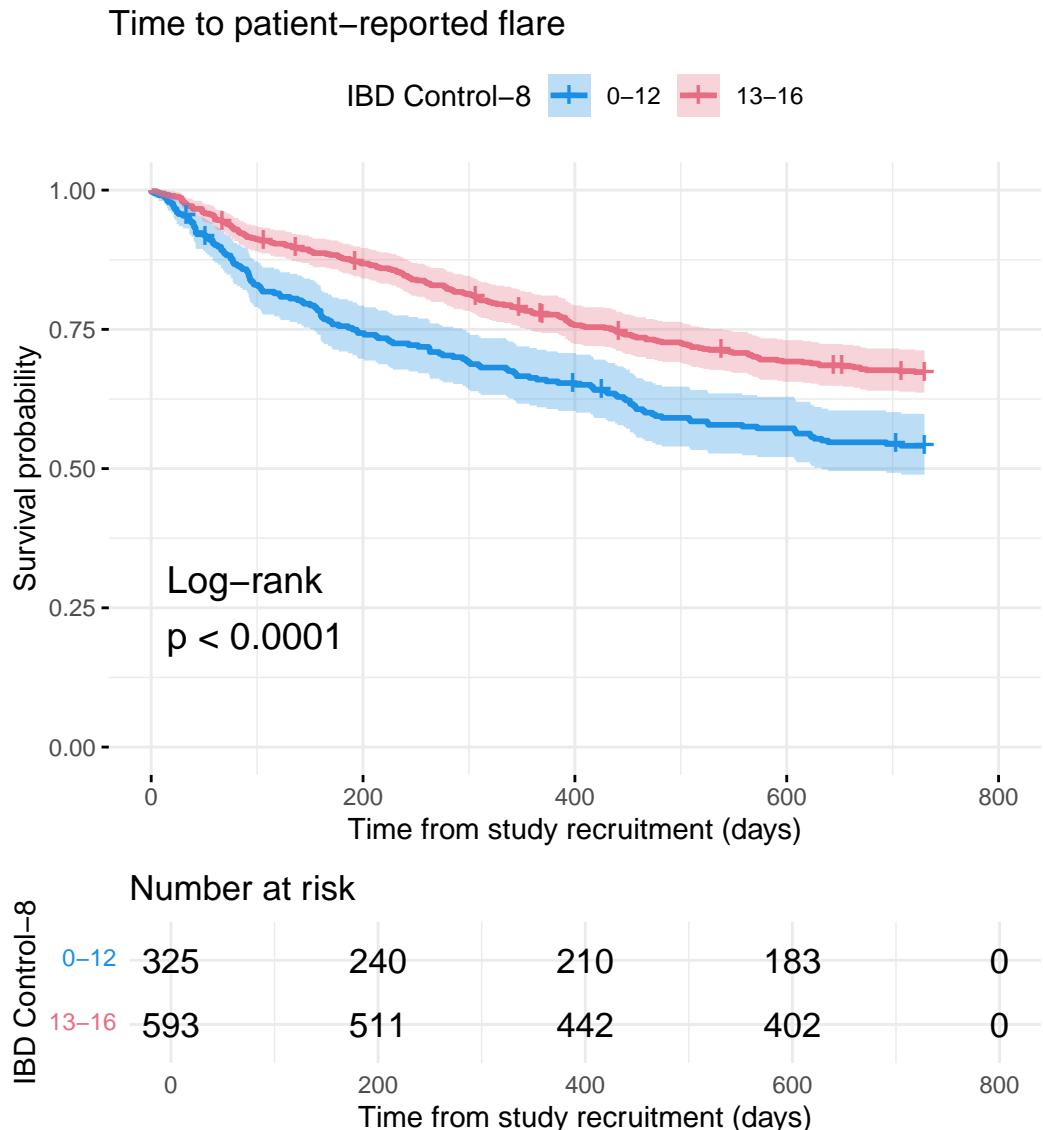
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.uc.df,
  var_name = "control_8",
  outcome_time = "softflare_time",
  outcome_event = "softflare",
  legend_title = "IBD Control-8",
  plot_base_path = "plots/uc/soft-flare/ibd/control-8",
  break_time_by = 200,
  palette = c("#1A8FE3", "#E76D83")
)

# Extract hazard ratio for continuous control_8 variable
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + IMD + cat + control_8 + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

uc.clin.forest <- rbind(uc.clin.forest, get_HR(fit.me, "control_8"))

# Display plot and model summary
analysis_result$plot

```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.4904	1.1840	1.8761	0.0007
IMD2	1.3212	0.7897	2.2106	0.2888
IMD3	1.2137	0.7394	1.9922	0.4437

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	1.4949	0.9289	2.4059	0.0977
IMD5	1.3596	0.8507	2.1729	0.1992
catFC 50-250	1.5424	1.1893	2.0002	0.0011
catFC > 250	2.0139	1.5095	2.6868	0.0000
control_8	0.9225	0.8909	0.9553	0.0000

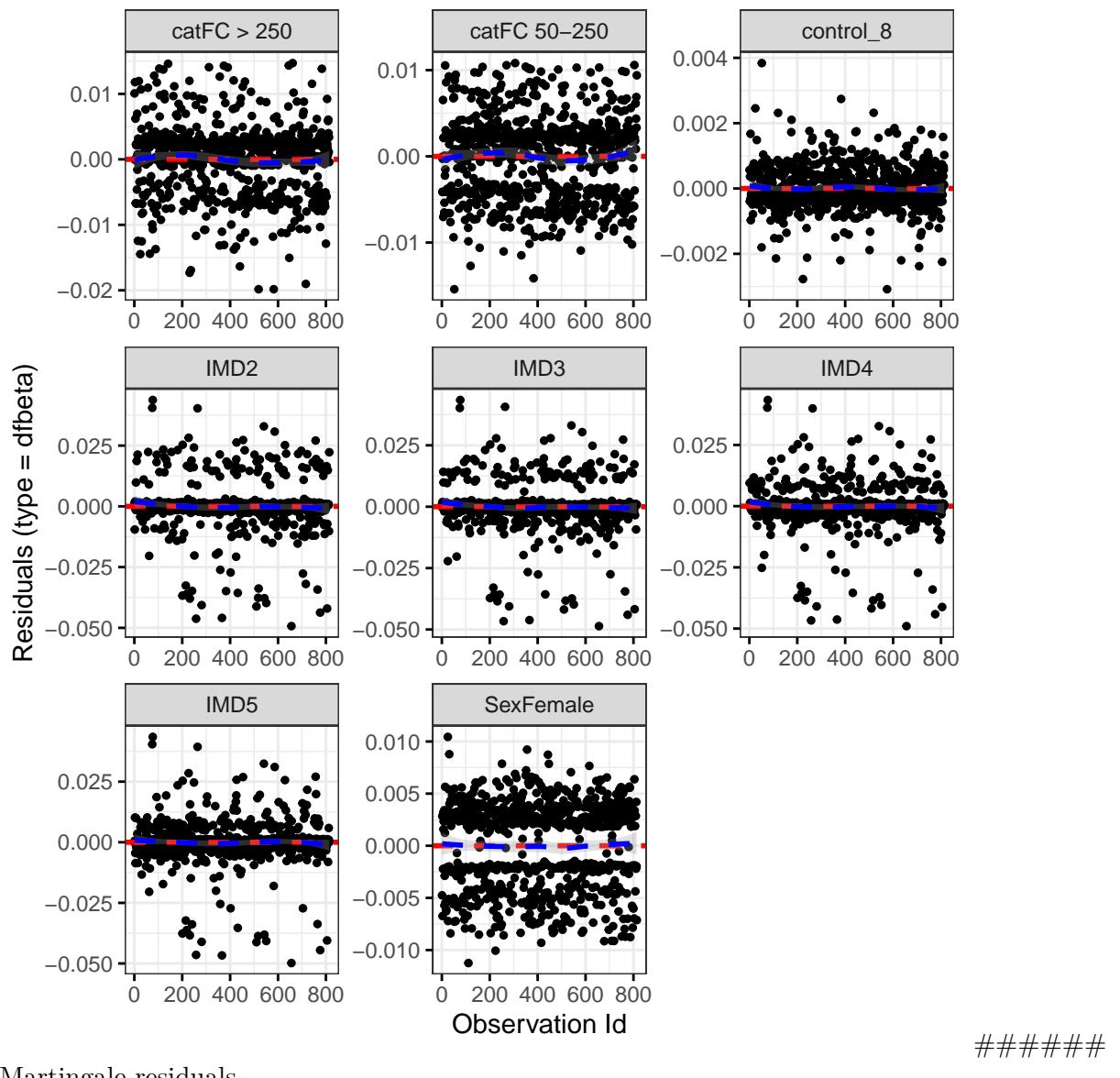
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	2.9678	0.9953	0.0844
IMD	4.2823	3.9667	0.3644
cat	4.5769	1.9835	0.1000
control_8	5.2699	0.9948	0.0215
GLOBAL	15.8660	12.7158	0.2390

### DF betas

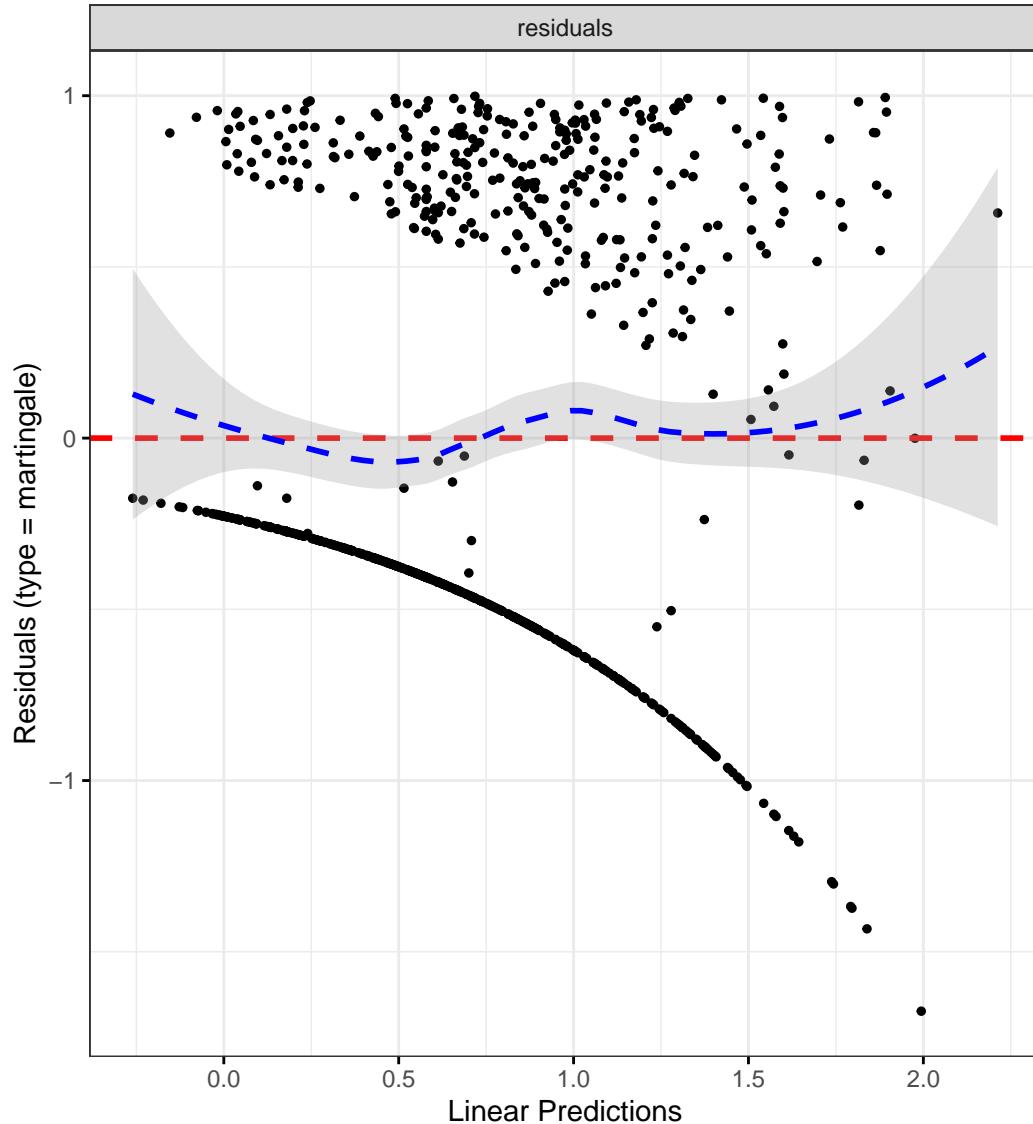
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



## Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.uc.df,
  var_name = "control_8",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "IBD Control-8",
```

```

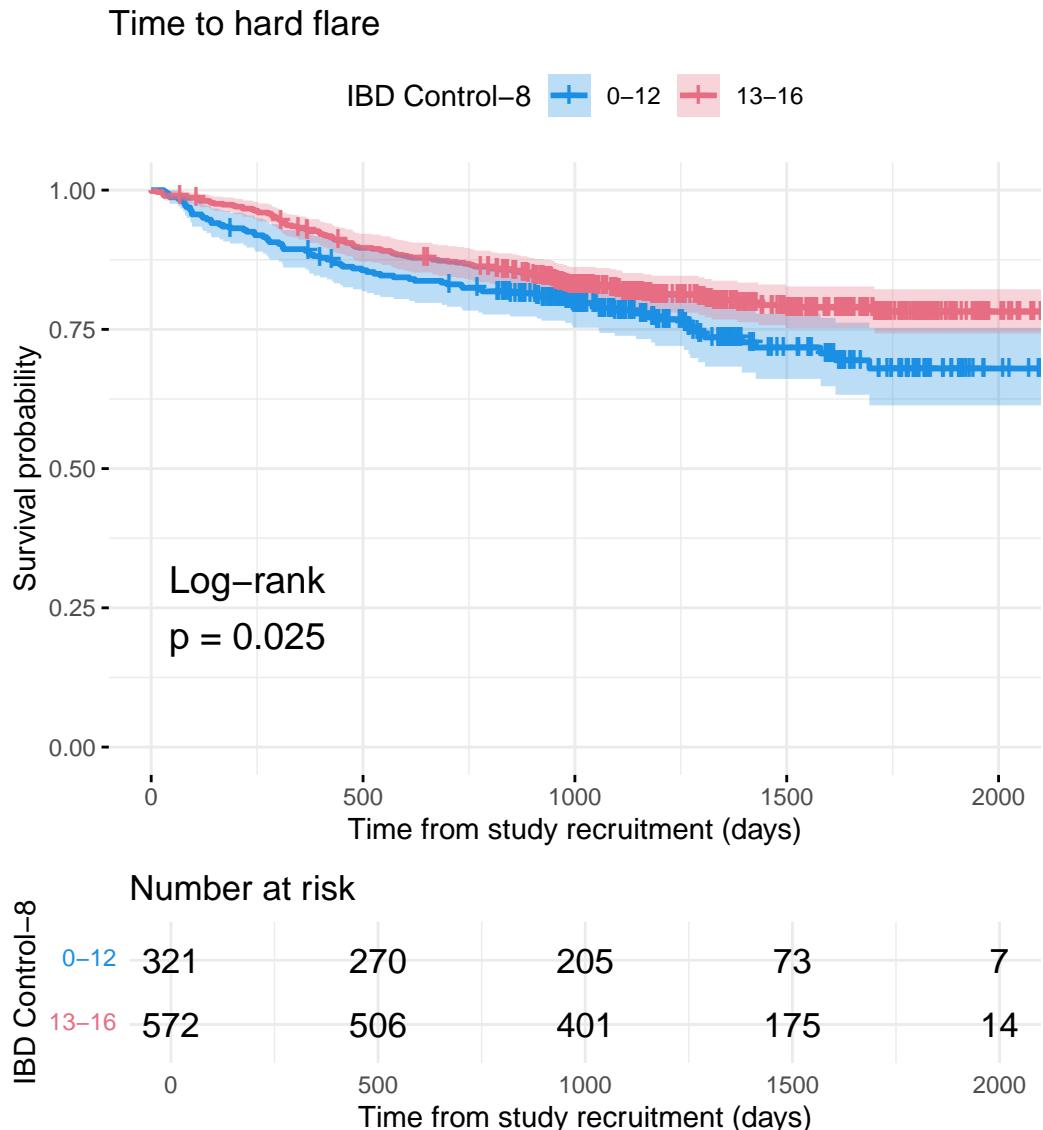
plot_base_path = "plots/uc/hard-flare/ibd/control-8",
break_time_by = 500,
palette = c("#1A8FE3", "#E76D83")
)

# Extract hazard ratio for continuous control_8 variable
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + IMD + cat + control_8 + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

uc.hard.forest <- rbind(uc.hard.forest, get_HR(fit.me, "control_8"))

# Display plot and model summary
analysis_result$plot

```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.3913	1.0274	1.8840	0.0328
IMD2	1.3477	0.6476	2.8046	0.4248
IMD3	1.5031	0.7484	3.0187	0.2520

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	2.1278	1.0942	4.1375	0.0261
IMD5	1.6178	0.8327	3.1432	0.1557
catFC 50-250	2.0148	1.4131	2.8728	0.0001
catFC > 250	3.0295	2.0721	4.4293	0.0000
control_8	0.9453	0.9026	0.9899	0.0169

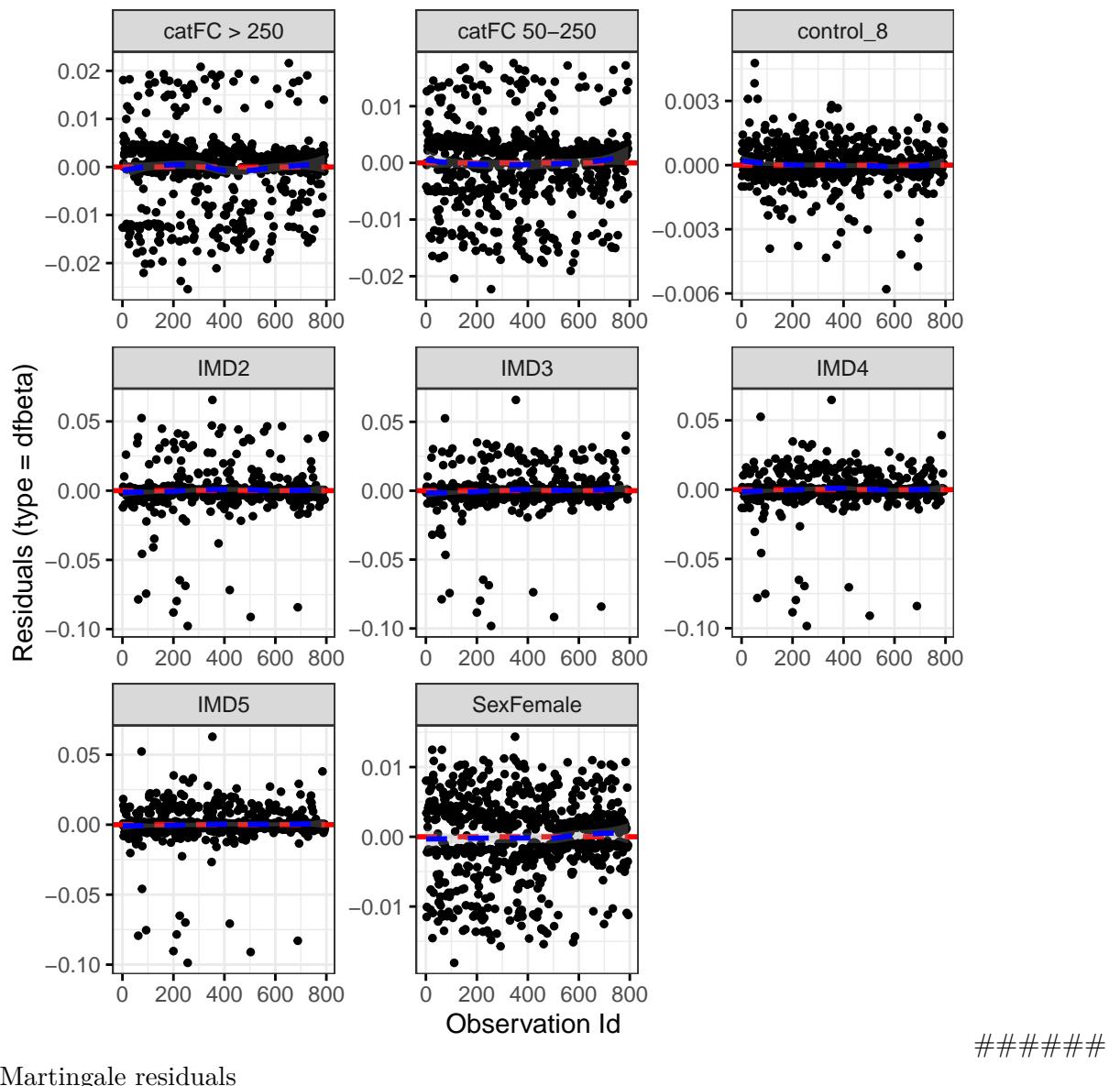
Diagnostics:

### Proportional hazards assumption test

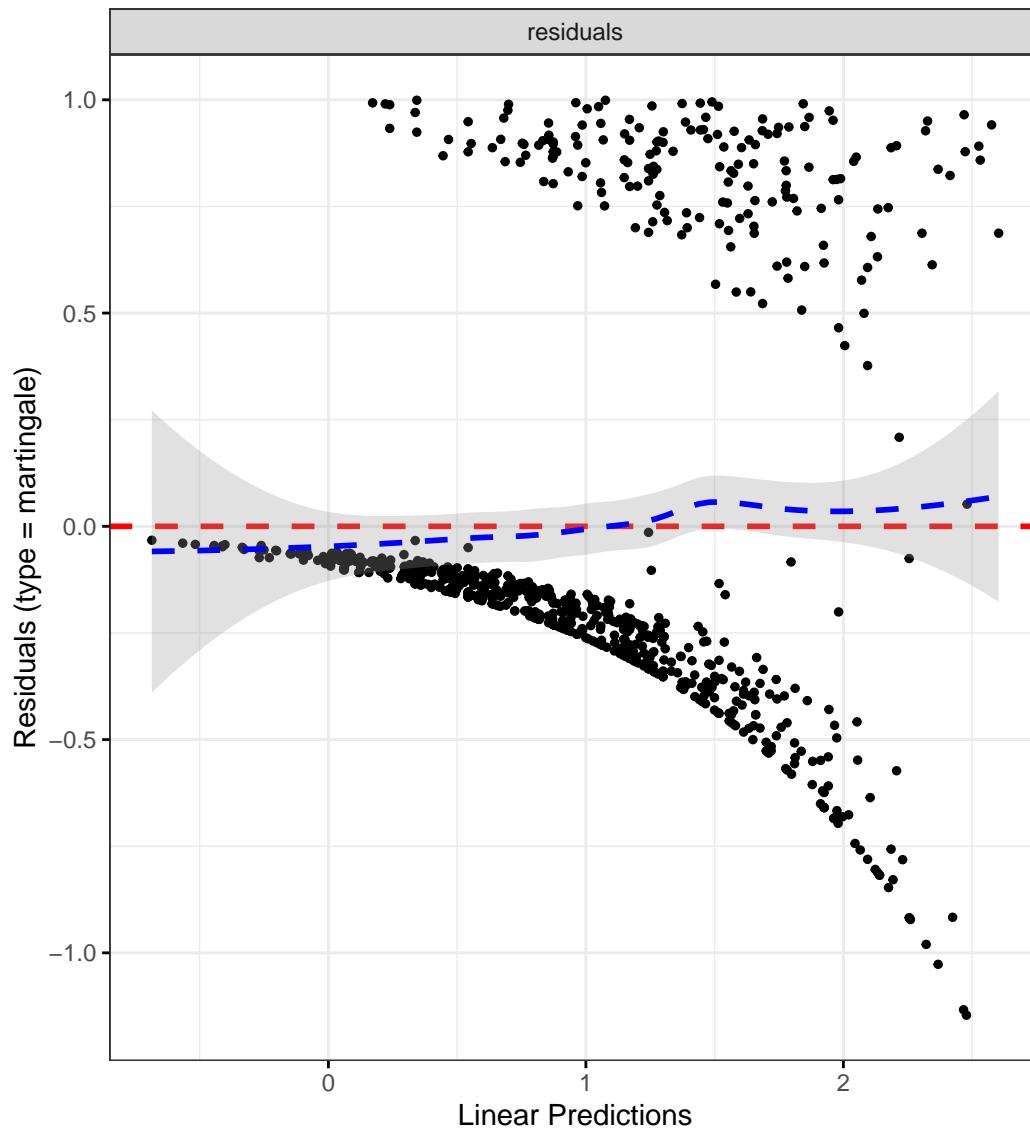
	Chi-squared statistic	DF	P-value
Sex	0.1400	0.9860	0.7027
IMD	2.7617	3.9396	0.5890
cat	3.5264	1.9665	0.1671
control_8	1.7488	0.9845	0.1826
GLOBAL	8.2255	23.1839	0.9982

### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



```
`geom_smooth()` using formula = 'y ~ x'
```



### IBD Control visual analogue scale

Crohn's disease

Patient-reported flare

```
# Handle VAS control (already categorized as character)
flare.cd.df$vas_control_cat <- factor(flare.cd.df$vas_control)
```

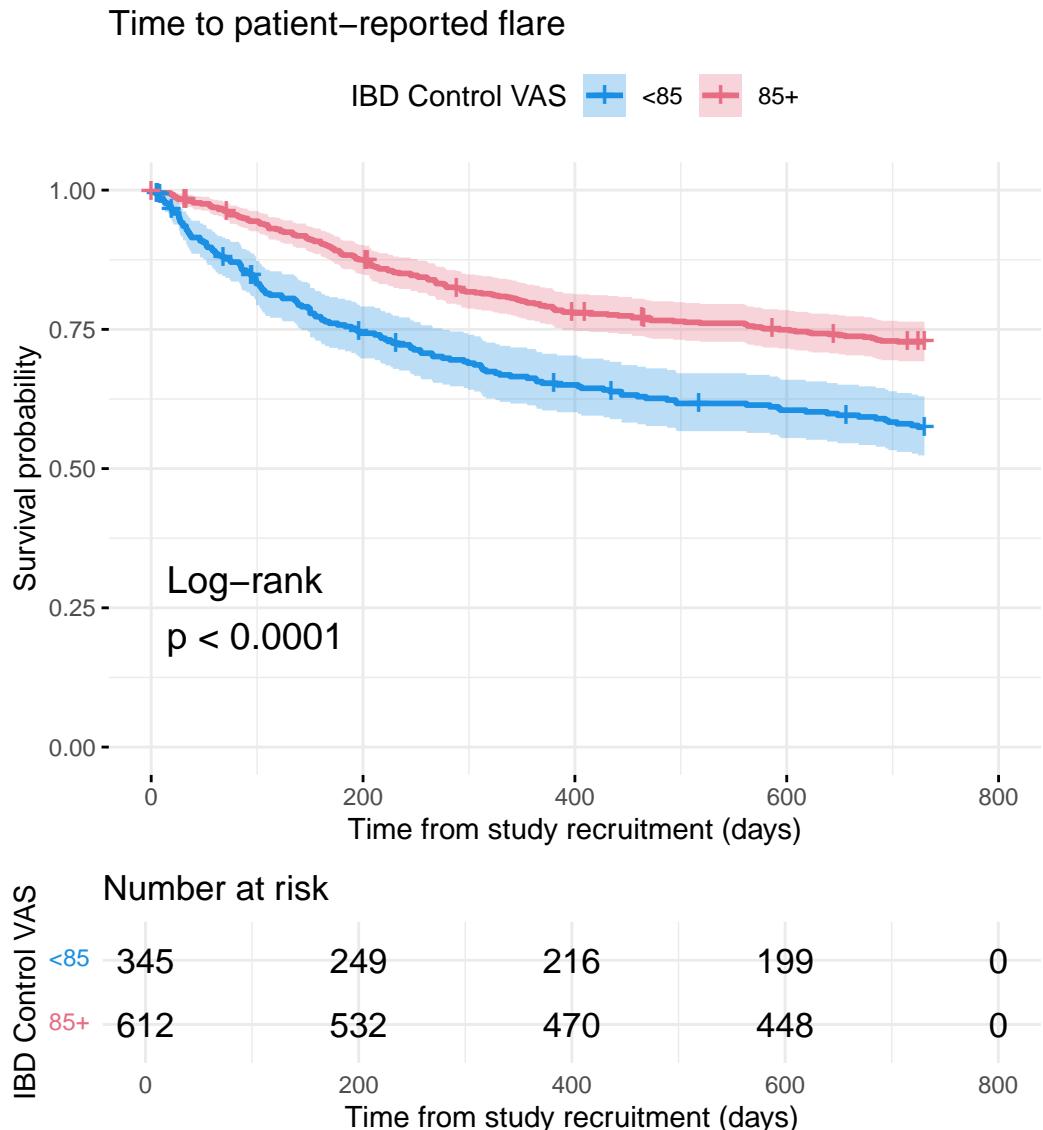
```

# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "vas_control",
  outcome_time = "softflare_time",
  outcome_event = "softflare",
  legend_title = "IBD Control VAS",
  plot_base_path = "plots/cd/soft-flare/ibd/control-vas",
  break_time_by = 200,
  palette = c("#1A8FE3", "#E76D83")
)

# Extract hazard ratio for vas_control variable
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + IMD + cat + vas_control + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

# Display plot and model summary
analysis_result$plot

```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	2.0996	1.6160	2.7280	0.0000
IMD2	0.7695	0.4706	1.2582	0.2962
IMD3	0.9004	0.5525	1.4672	0.6736

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	0.8423	0.5237	1.3546	0.4789
IMD5	0.9688	0.6121	1.5333	0.8924
catFC 50-250	1.3678	1.0415	1.7963	0.0243
catFC > 250	2.3897	1.7645	3.2362	0.0000
vas_control85+	0.5710	0.4500	0.7246	0.0000

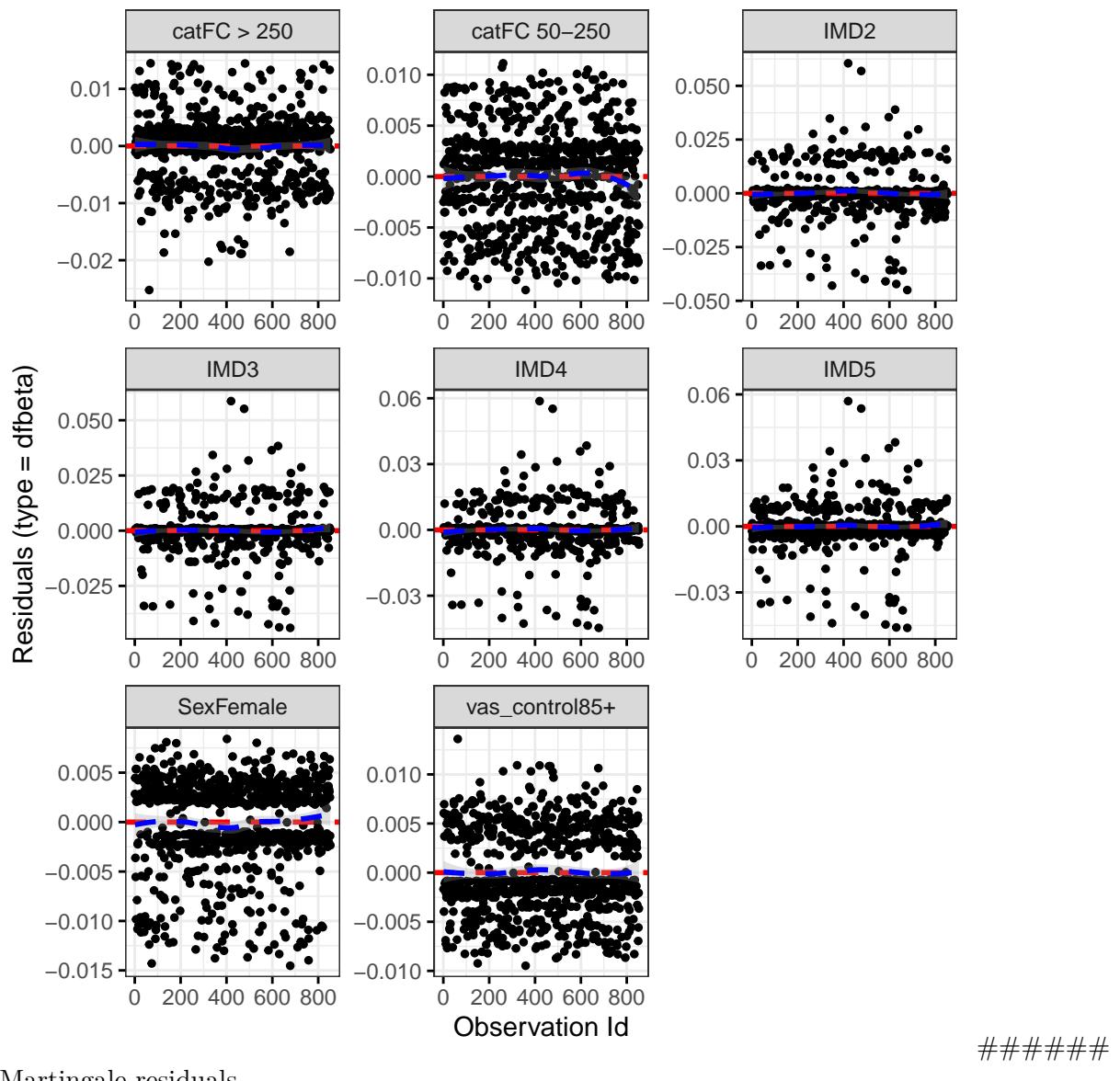
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.0298	0.9934	0.8610
IMD	5.8028	3.9533	0.2095
cat	1.3945	1.9837	0.4940
vas_control	6.4668	0.9893	0.0108
GLOBAL	13.2734	12.1071	0.3576

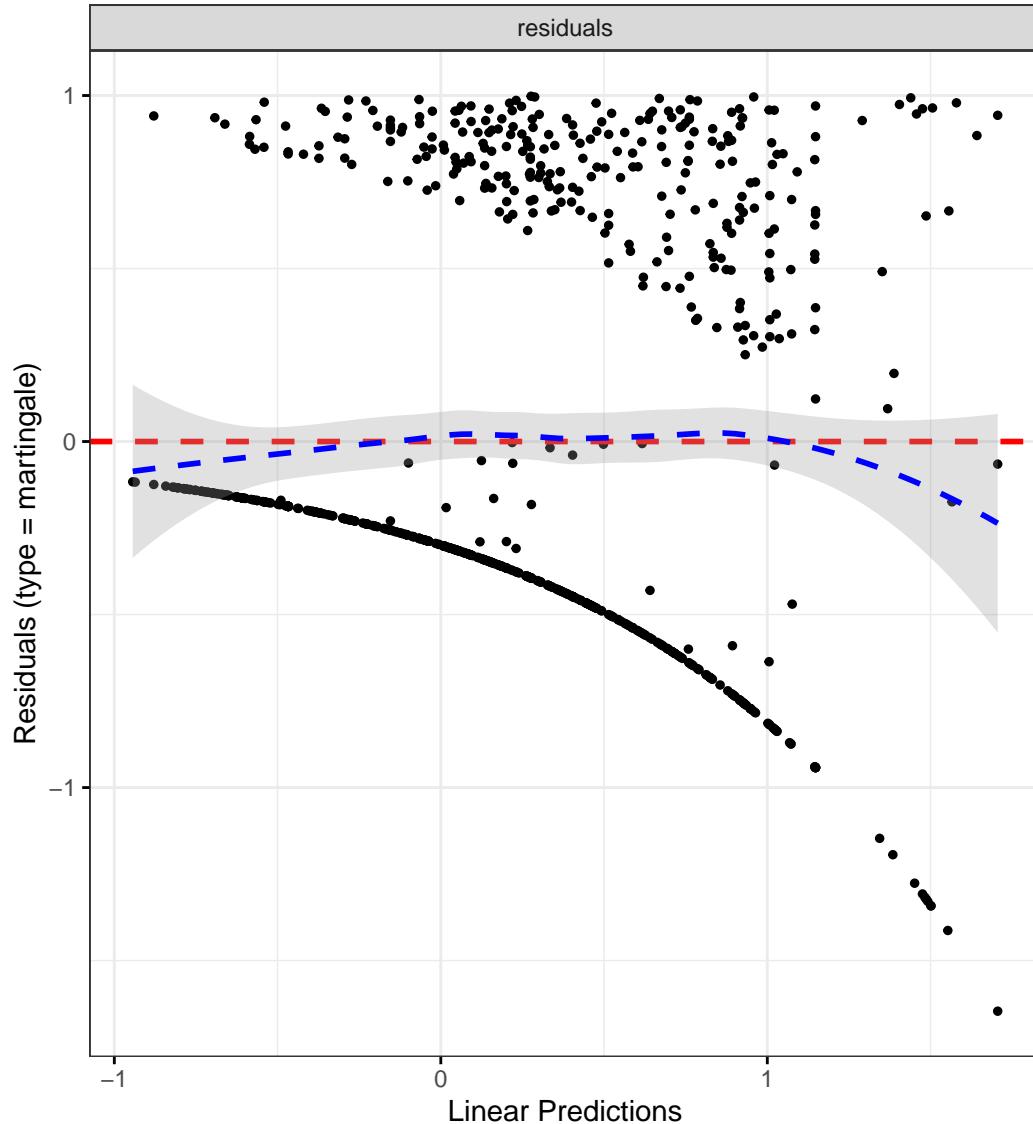
### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



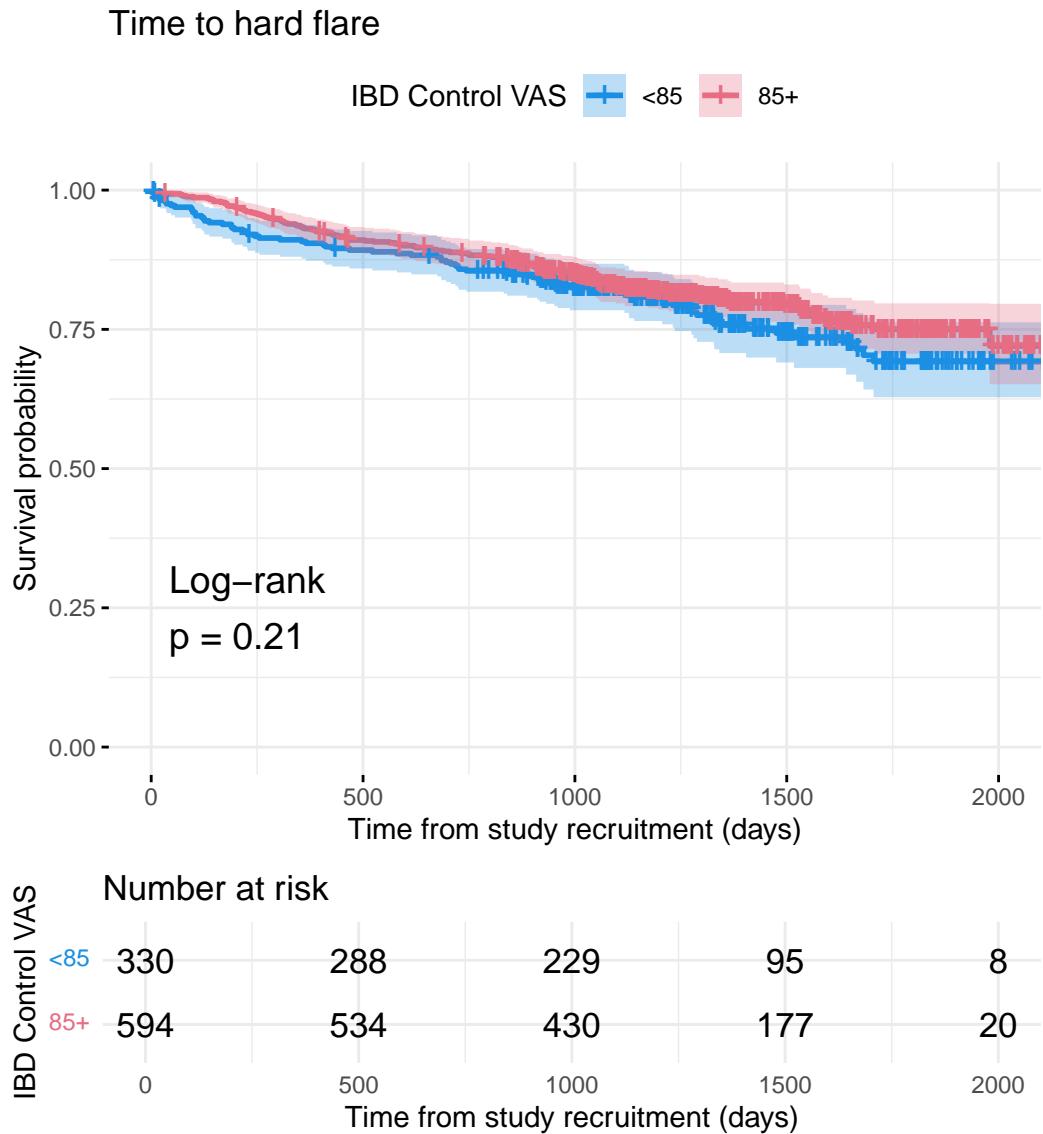
## Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "vas_control",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "IBD Control VAS",
```

```
plot_base_path = "plots/cd/hard-flare/ibd/control-vas",
break_time_by = 500,
palette = c("#1A8FE3", "#E76D83")
)

# Extract hazard ratio for continuous vas_control variable
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + IMD + cat + vas_control + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

# Display plot and model summary
analysis_result$plot
```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.3751	1.0101	1.8721	0.0430
IMD2	0.6859	0.3715	1.2665	0.2282
IMD3	0.8599	0.4699	1.5734	0.6244

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	0.7481	0.4148	1.3491	0.3347
IMD5	0.7802	0.4412	1.3800	0.3937
catFC 50-250	1.8928	1.3283	2.6973	0.0004
catFC > 250	3.4895	2.3810	5.1142	0.0000
vas_control85+	0.9359	0.6894	1.2705	0.6708

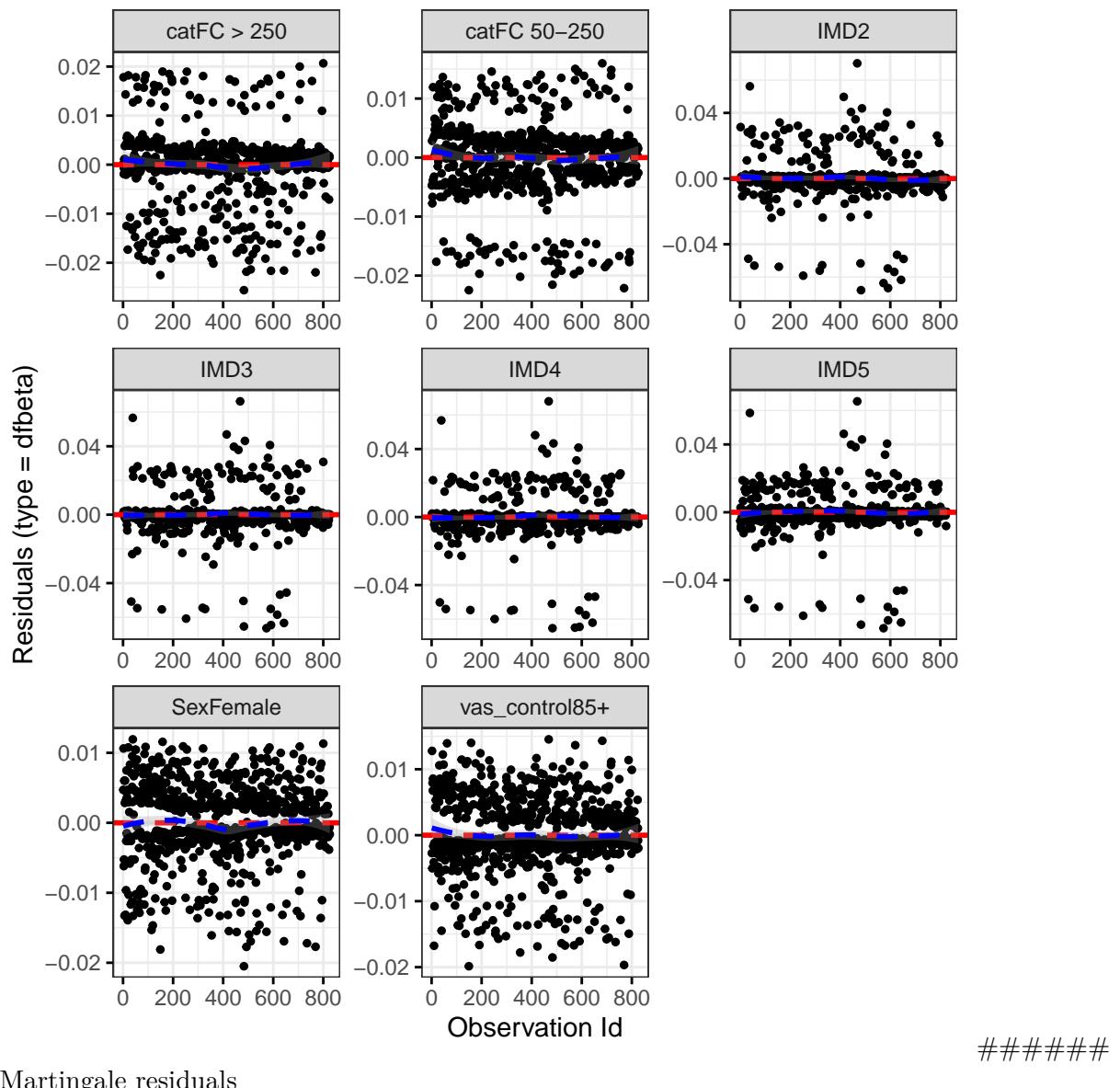
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	1.6756	0.9837	0.1918
IMD	7.9021	3.9395	0.0918
cat	8.1082	1.9864	0.0171
vas_control	0.2524	0.9877	0.6102
GLOBAL	20.3223	15.8147	0.1970

### DF betas

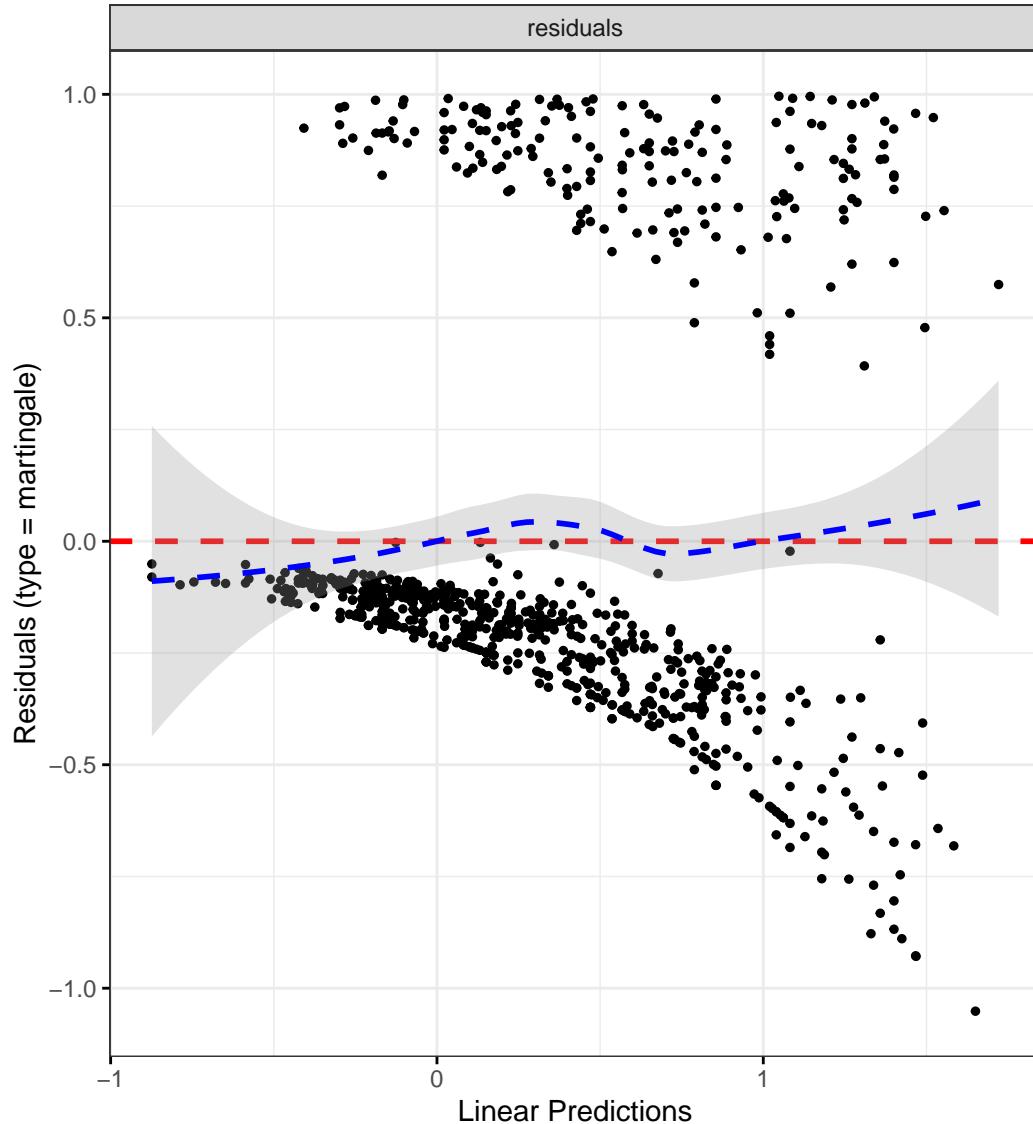
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



### Ulcerative colitis

#### Patient-reported flare

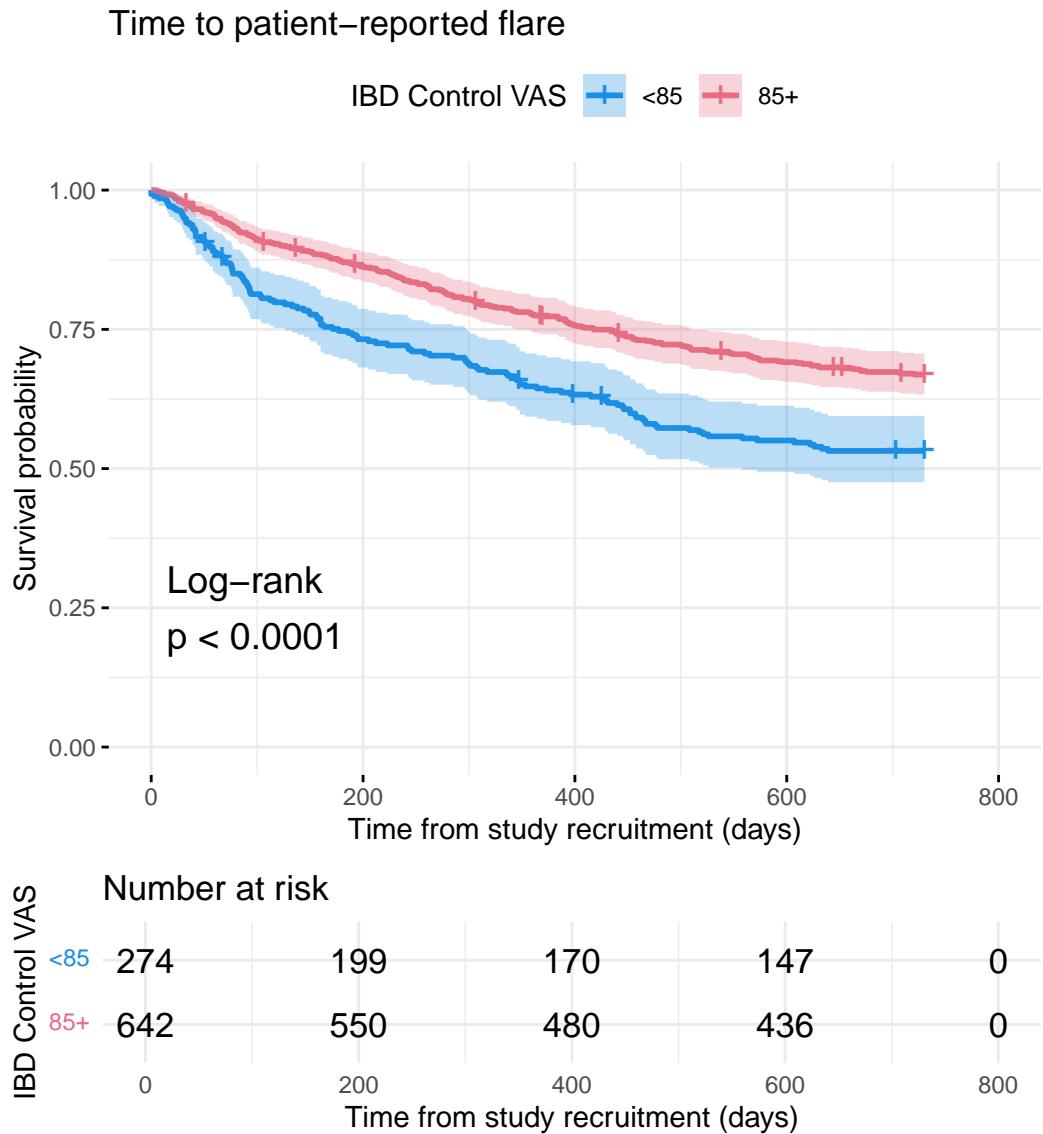
```
# Handle VAS control (already categorized as character)
flare.uc.df$vas_control_cat <- factor(flare.uc.df$vas_control)

# Run survival analysis using utility function
analysis_result <- run_survival_analysis()
```

```
data = flare.uc.df,
var_name = "vas_control",
outcome_time = "softflare_time",
outcome_event = "softflare",
legend_title = "IBD Control VAS",
plot_base_path = "plots/uc/soft-flare/ibd/control-vas",
break_time_by = 200,
palette = c("#1A8FE3", "#E76D83")
)

# Extract hazard ratio for vas_control variable
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + IMD + cat + vas_control + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

# Display plot and model summary
analysis_result$plot
```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.5419	1.2247	1.9413	0.0002
IMD2	1.3325	0.7893	2.2496	0.2826
IMD3	1.2056	0.7280	1.9968	0.4675

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	1.4696	0.9046	2.3874	0.1199
IMD5	1.3493	0.8371	2.1749	0.2188
catFC 50-250	1.5798	1.2174	2.0502	0.0006
catFC > 250	1.9587	1.4652	2.6185	0.0000
vas_control85+	0.6861	0.5423	0.8679	0.0017

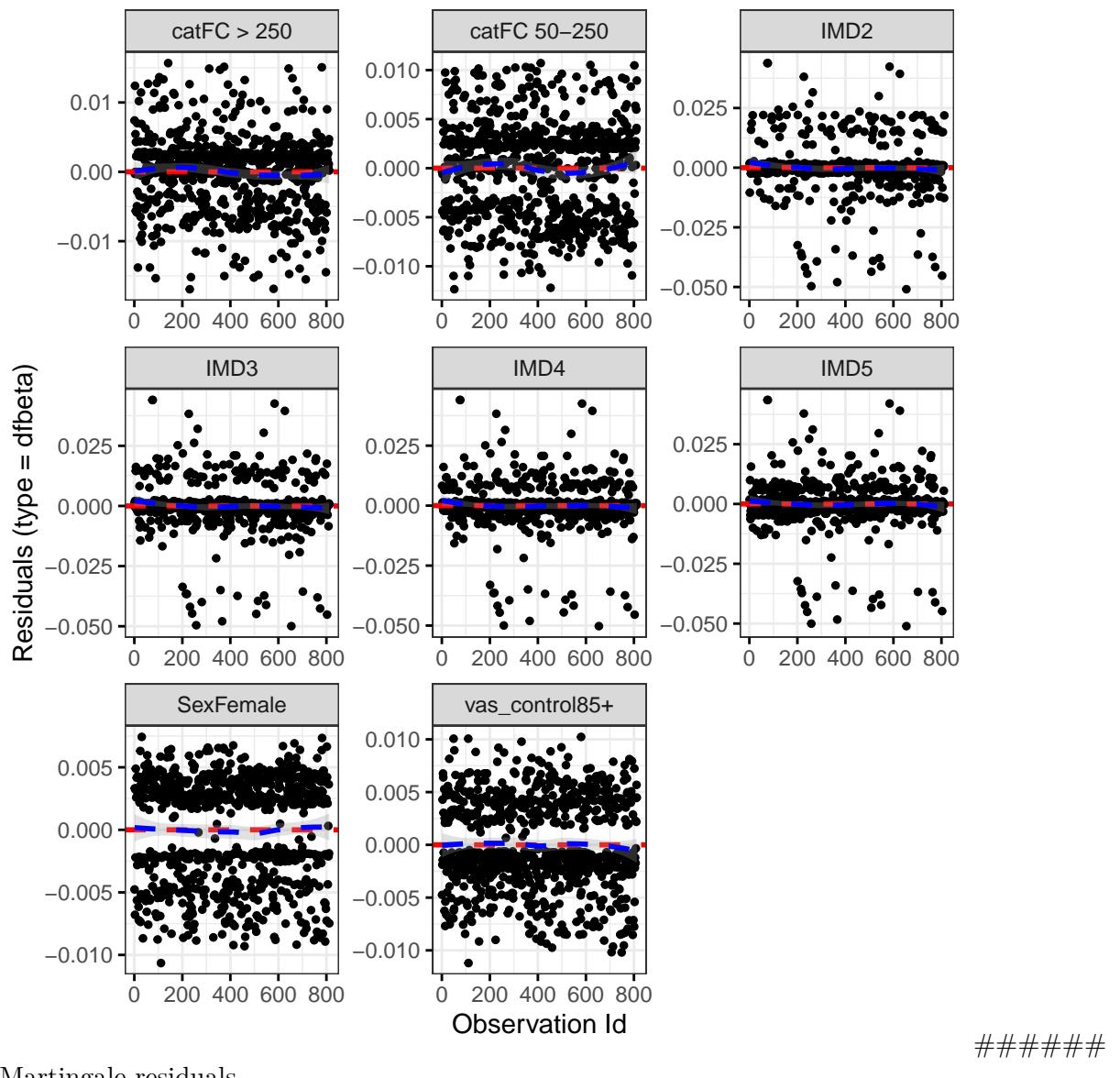
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	2.9736	0.9943	0.0840
IMD	4.3116	3.9618	0.3601
cat	5.7100	1.9817	0.0566
vas_control	3.5219	0.9923	0.0599
GLOBAL	16.5023	13.4488	0.2495

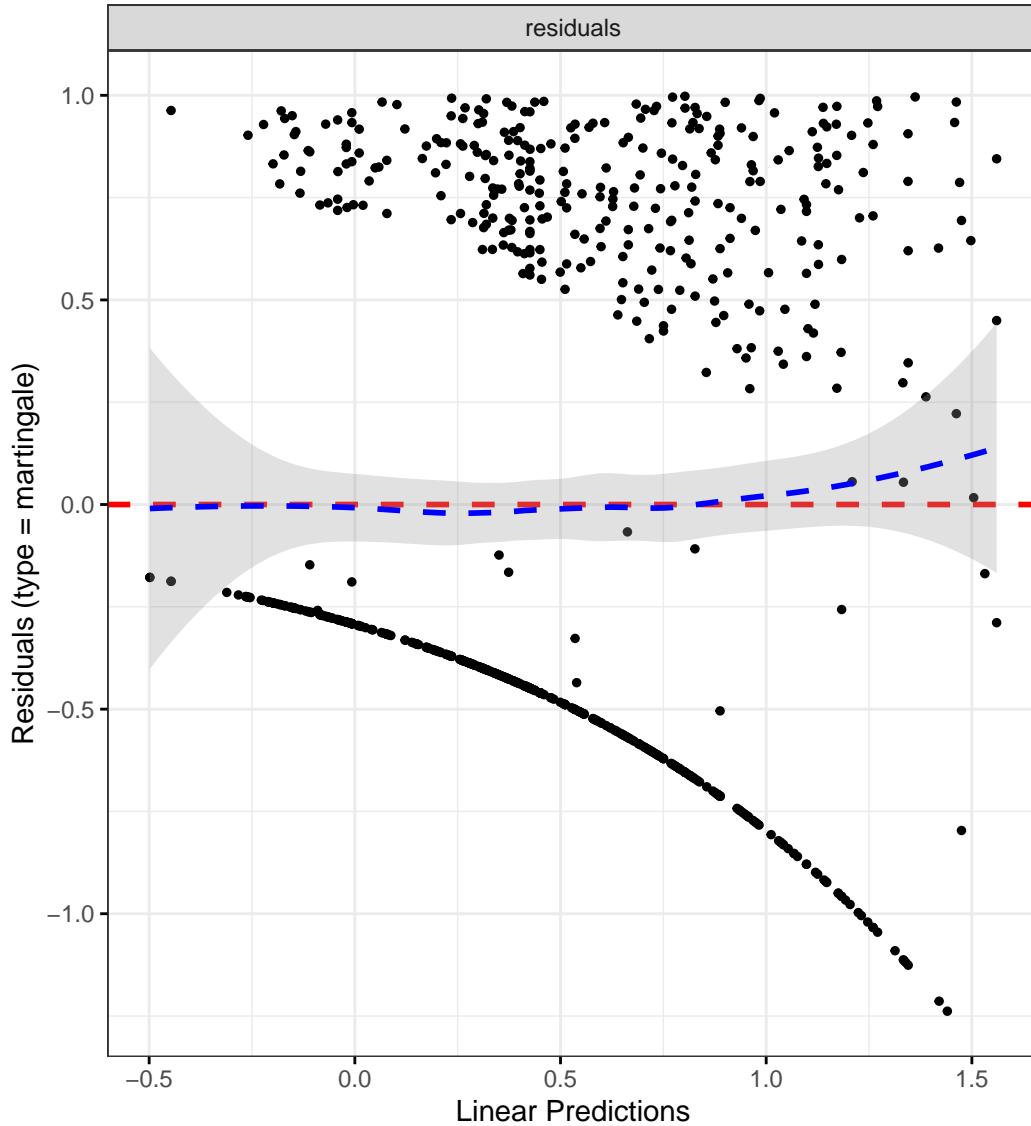
### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



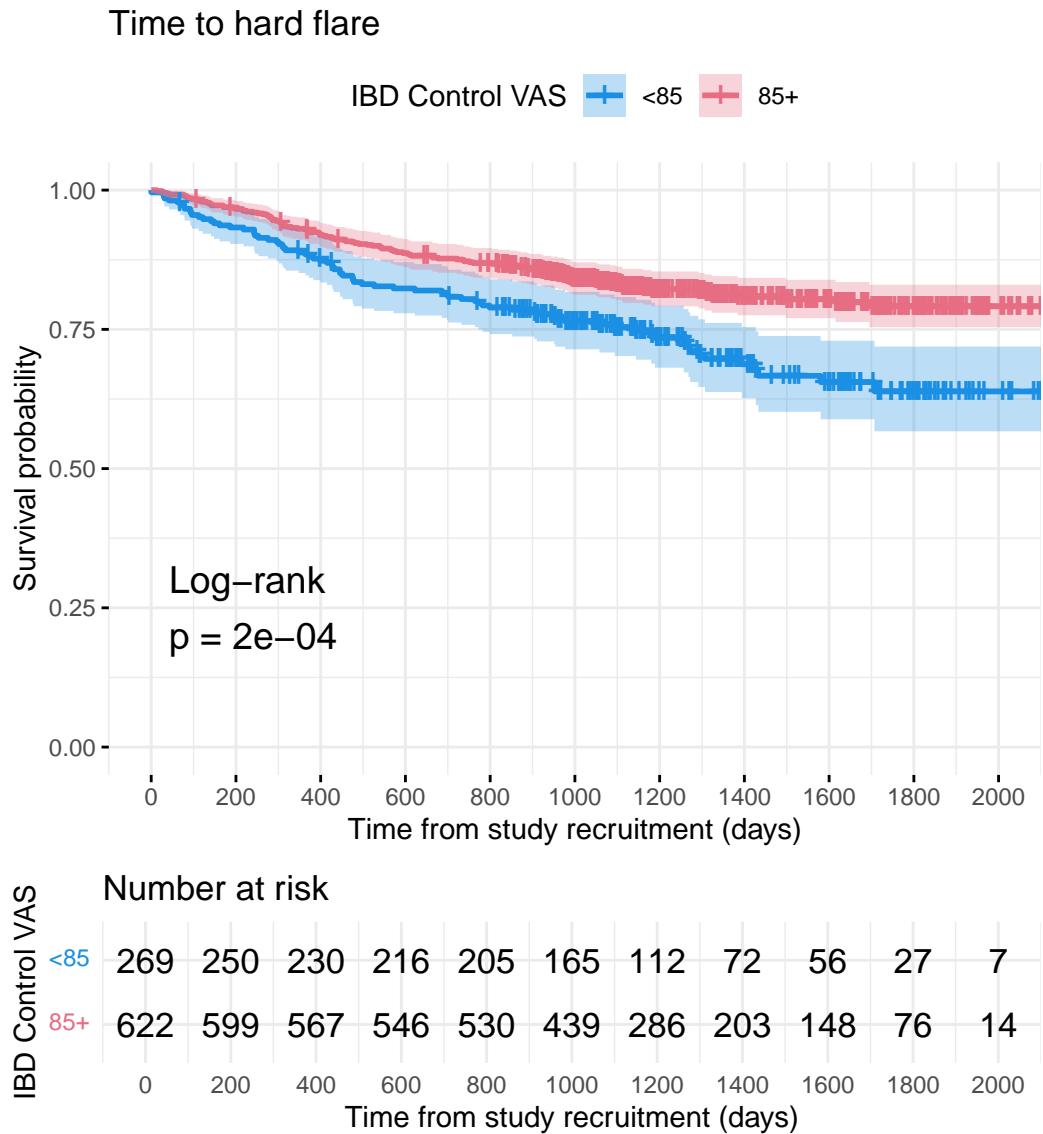
## Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.uc.df,
  var_name = "vas_control",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "IBD Control VAS",
```

```
plot_base_path = "plots/uc/hard-flare/ibd/control-vas",
break_time_by = 200,
palette = c("#1A8FE3", "#E76D83")
)

# Extract hazard ratio for vas_control variable
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + IMD + cat + vas_control + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

# Display plot and model summary
analysis_result$plot
```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.3848	1.0220	1.8763	0.0357
IMD2	1.2918	0.6208	2.6877	0.4935
IMD3	1.4220	0.7087	2.8532	0.3218

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	1.9936	1.0297	3.8599	0.0407
IMD5	1.5530	0.8031	3.0034	0.1908
catFC 50-250	2.0459	1.4357	2.9155	0.0001
catFC > 250	2.9077	1.9851	4.2592	0.0000
vas_control85+	0.6559	0.4815	0.8933	0.0075

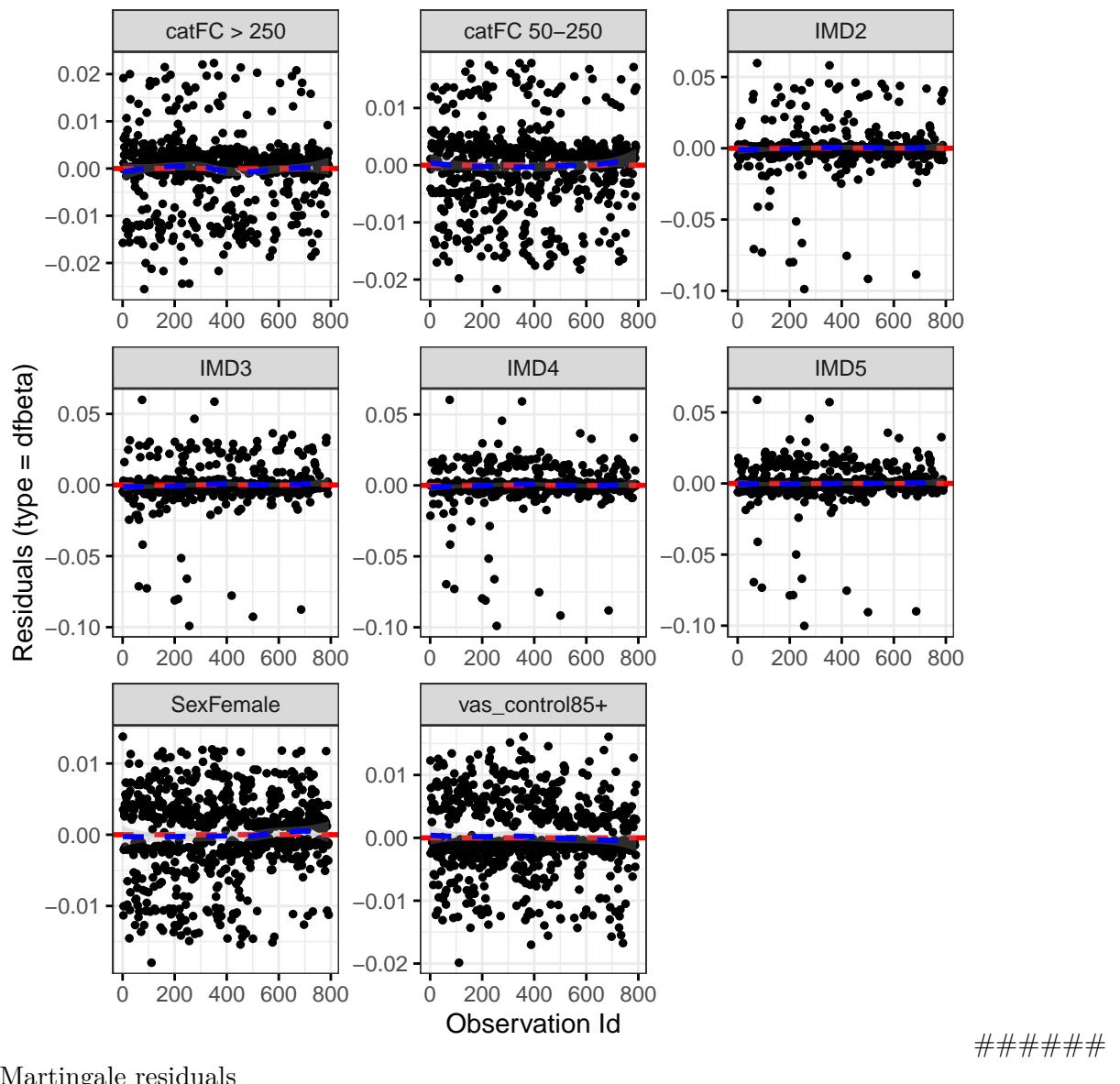
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.1785	0.9852	0.6666
IMD	2.8643	3.9383	0.5711
cat	3.3111	1.9665	0.1862
vas_control	0.2632	0.9852	0.6017
GLOBAL	7.3842	23.0999	0.9992

### DF betas

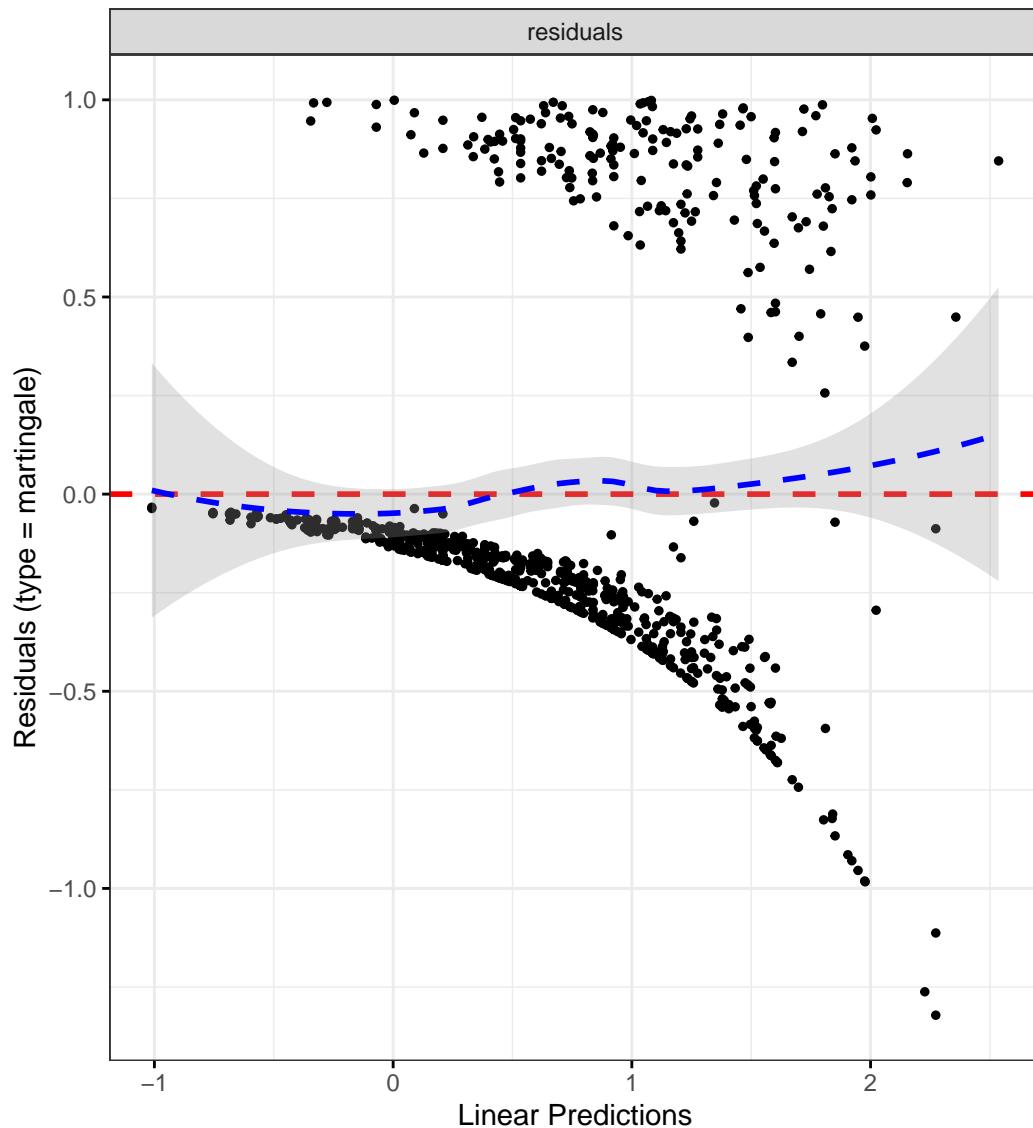
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



## Biologic use

Crohn's disease

Patient-reported flare

```
# Transform biologic variable and create categorized version
flare.cd.df <- flare.cd.df %>%
  mutate(Biologic = plyr::mapvalues(Biologic,
```

```

        from = c("Current",
                 "Previously",
                 "Never prescribed"),
      to = c("Prescribed",
                 "Not prescribed",
                 "Not prescribed")))

# Create categorized version for survival analysis
flare.cd.df$Biologic_cat <- flare.cd.df$Biologic

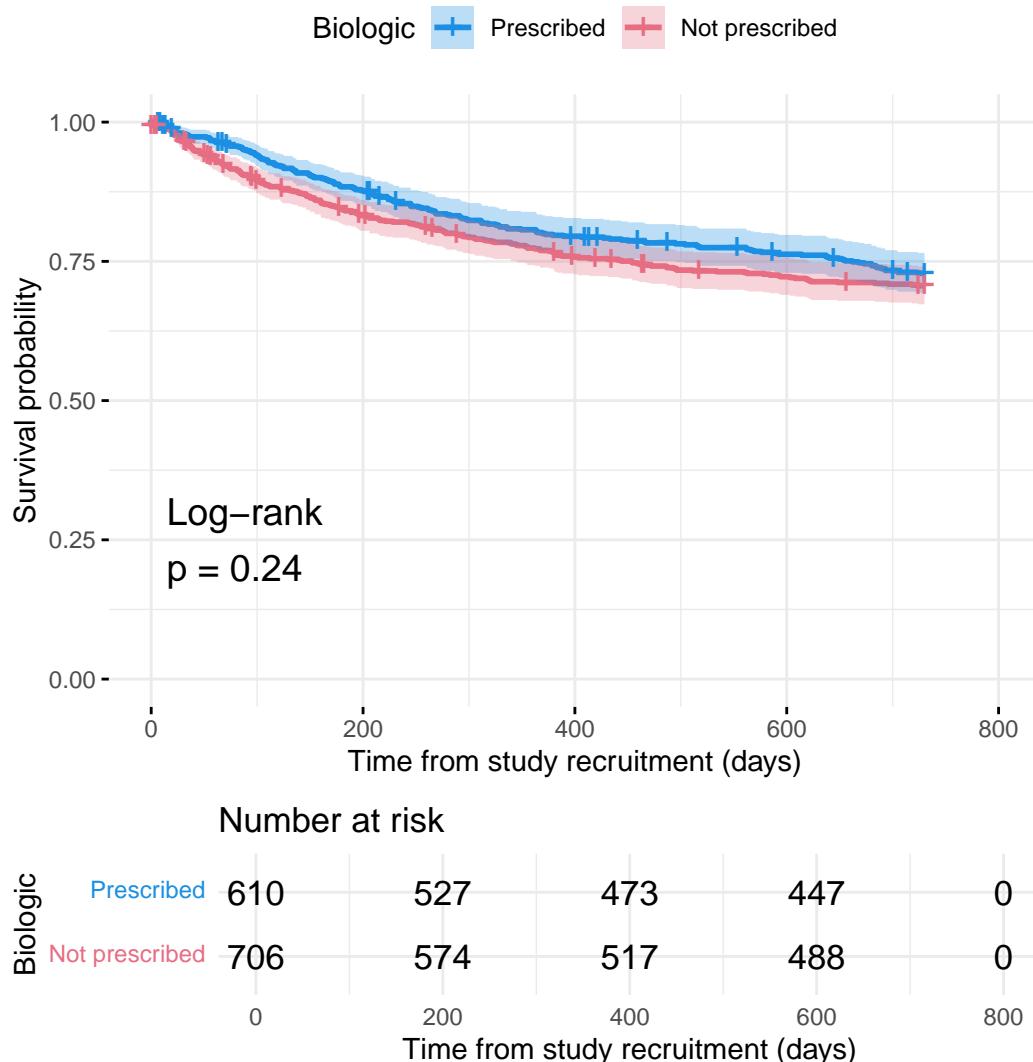
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Biologic",
  outcome_time = "softflare_time",
  outcome_event = "softflare",
  legend_title = "Biologic",
  plot_base_path = "plots/cd/soft-flare/ibd/biologic",
  break_time_by = 200,
  palette = c("#1A8FE3", "#E76D83")
)

# Cox model
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + IMD + cat + Biologic + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

# Display plot and model summary
analysis_result$plot

```

### Time to patient-reported flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.9618	1.5445	2.4920	0.0000
IMD2	0.9342	0.5961	1.4640	0.7665
IMD3	0.8935	0.5657	1.4112	0.6293

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	0.9413	0.6065	1.4608	0.7872
IMD5	0.9803	0.6415	1.4982	0.9269
catFC 50-250	1.5753	1.2209	2.0325	0.0005
catFC > 250	2.4212	1.8260	3.2105	0.0000
BiologicNot prescribed	1.1577	0.9238	1.4509	0.2034

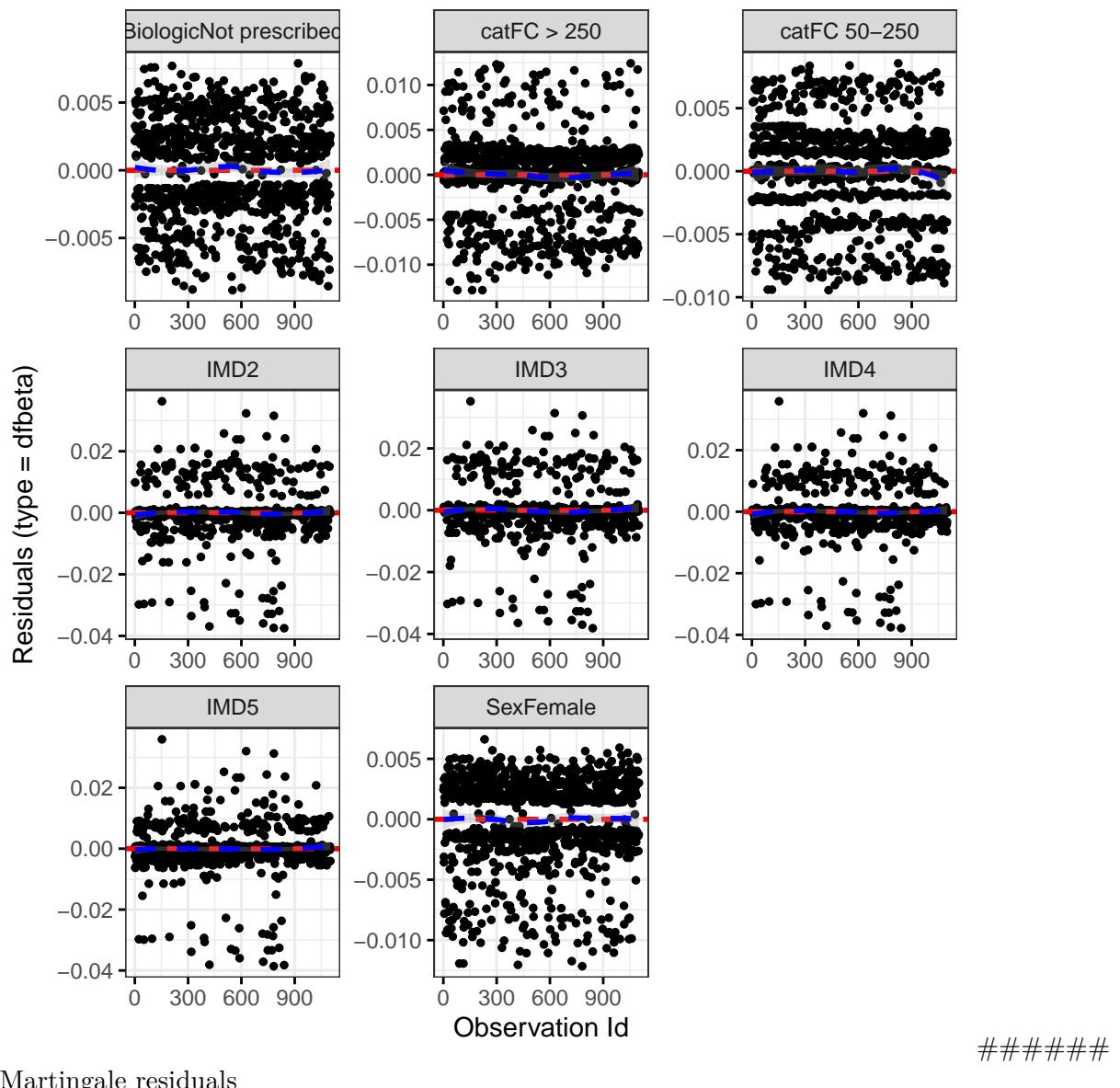
Diagnostics:

#### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.3184	0.9938	0.5700
IMD	5.8410	3.9587	0.2071
cat	2.2735	1.9848	0.3178
Biologic	6.5065	0.9707	0.0102
GLOBAL	15.8608	13.0633	0.2607

#### DF betas

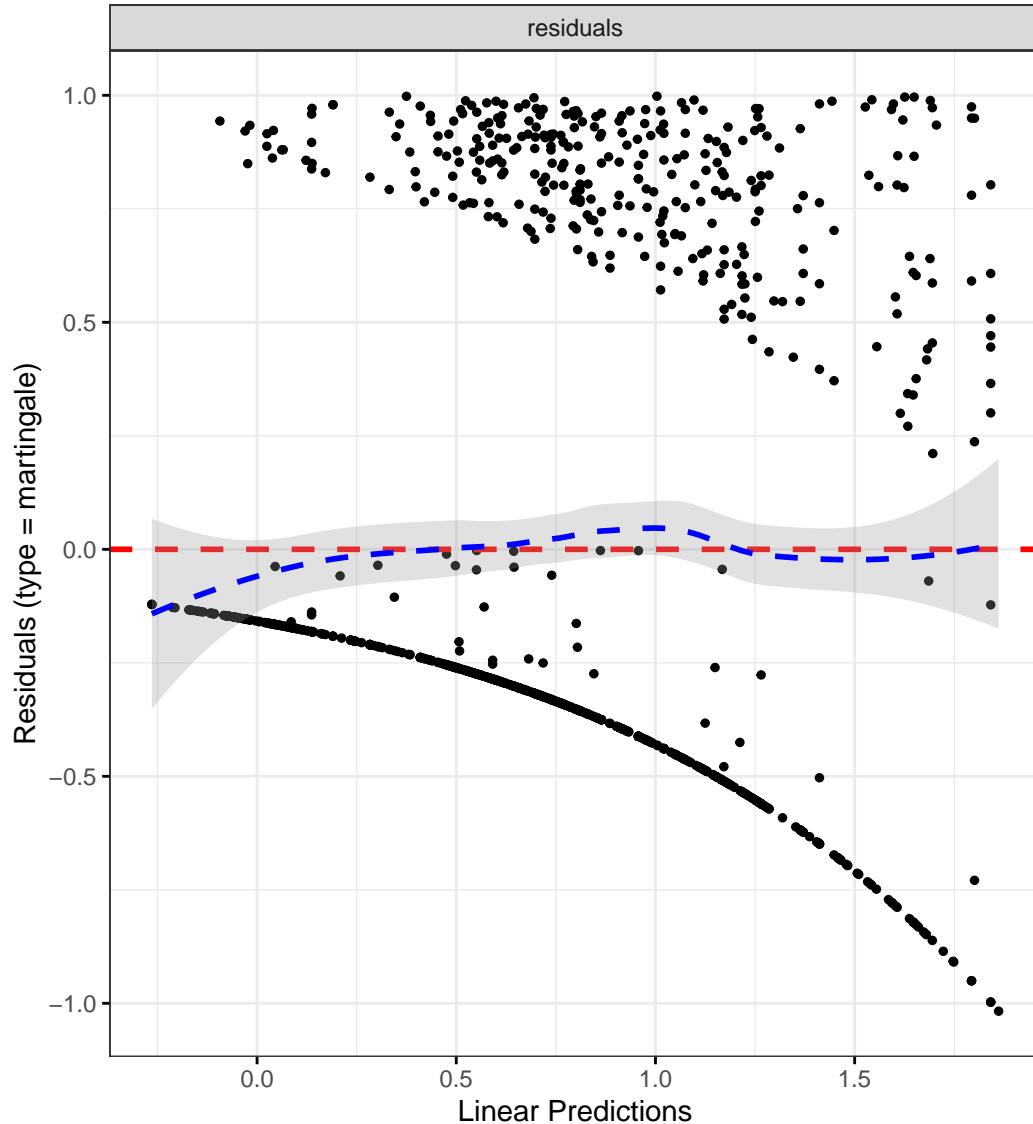
```
`geom_smooth()` using formula = 'y ~ x'
```



Martingale residuals

#####

```
`geom_smooth()` using formula = 'y ~ x'
```



## Hard flare

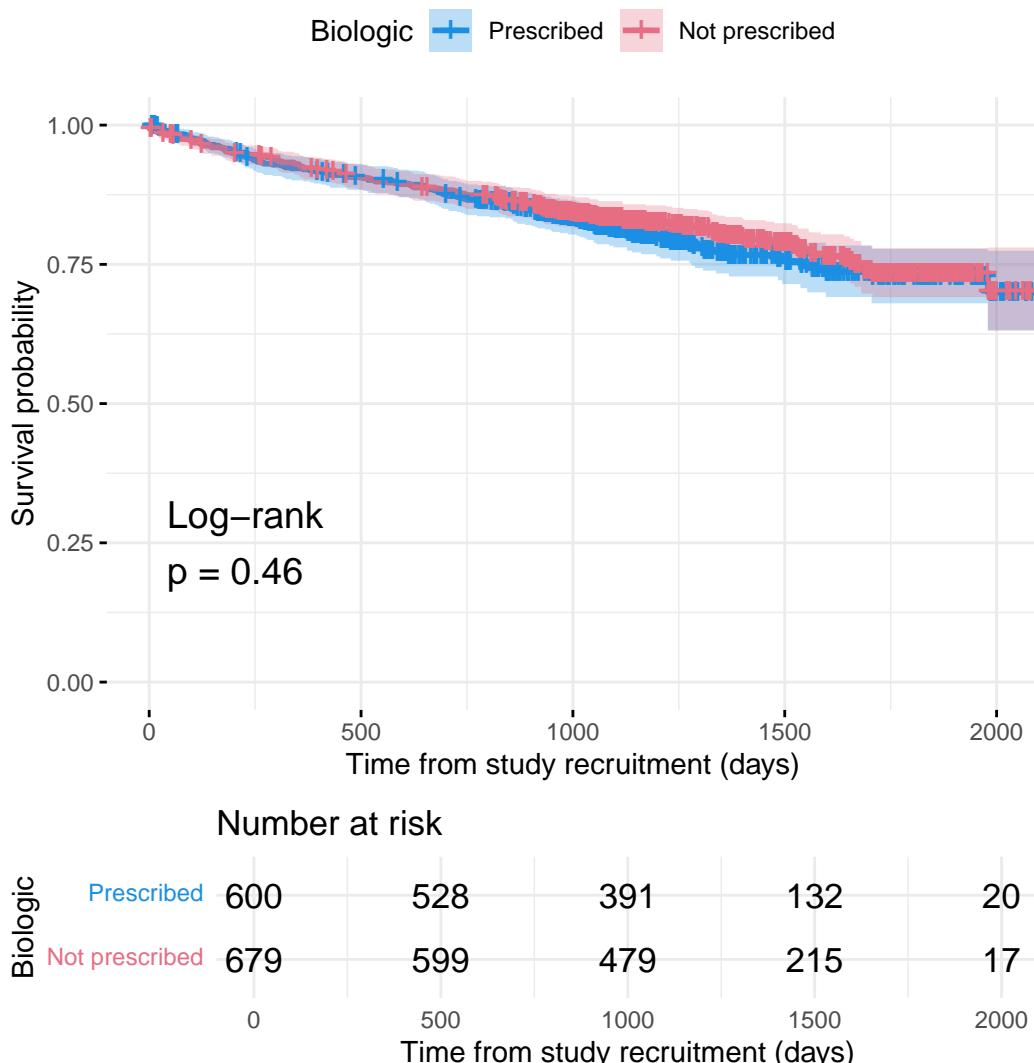
```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Biologic",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "Biologic",
```

```
plot_base_path = "plots/cd/hard-flare/ibd/biologic",
break_time_by = 500,
palette = c("#1A8FE3", "#E76D83")
)

# Cox model
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + IMD + cat + Biologic + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

# Display plot and model summary
analysis_result$plot
```

### Time to hard flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.3883	1.0557	1.8257	0.0189
IMD2	0.9219	0.5363	1.5847	0.7685
IMD3	0.9676	0.5566	1.6820	0.9070

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	0.8948	0.5220	1.5340	0.6862
IMD5	0.9034	0.5369	1.5200	0.7019
catFC 50-250	2.0215	1.4724	2.7754	0.0000
catFC > 250	3.3369	2.3692	4.6999	0.0000
BiologicNot prescribed	1.0022	0.7633	1.3158	0.9876

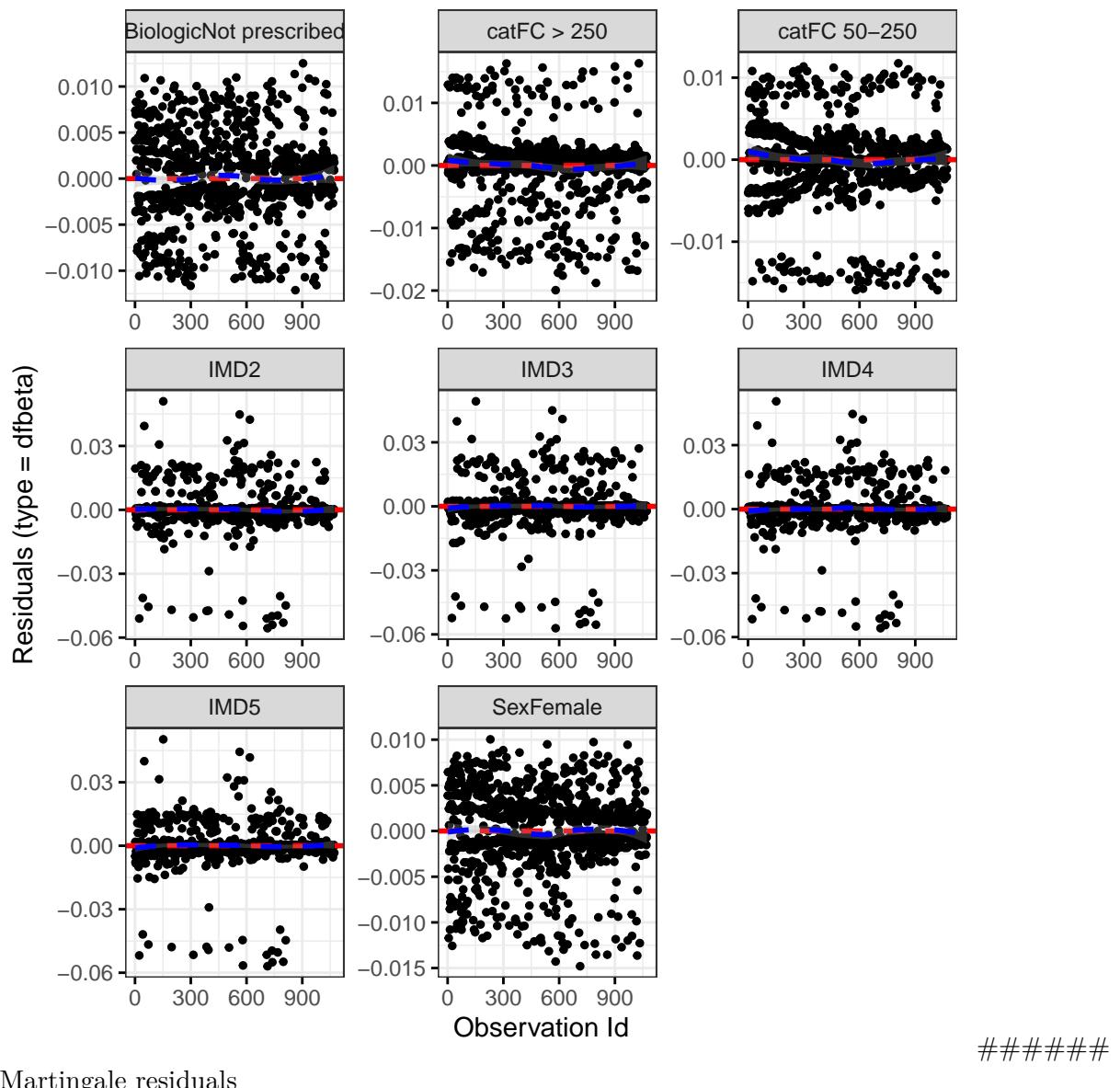
Diagnostics:

#### Proportional hazards assumption test

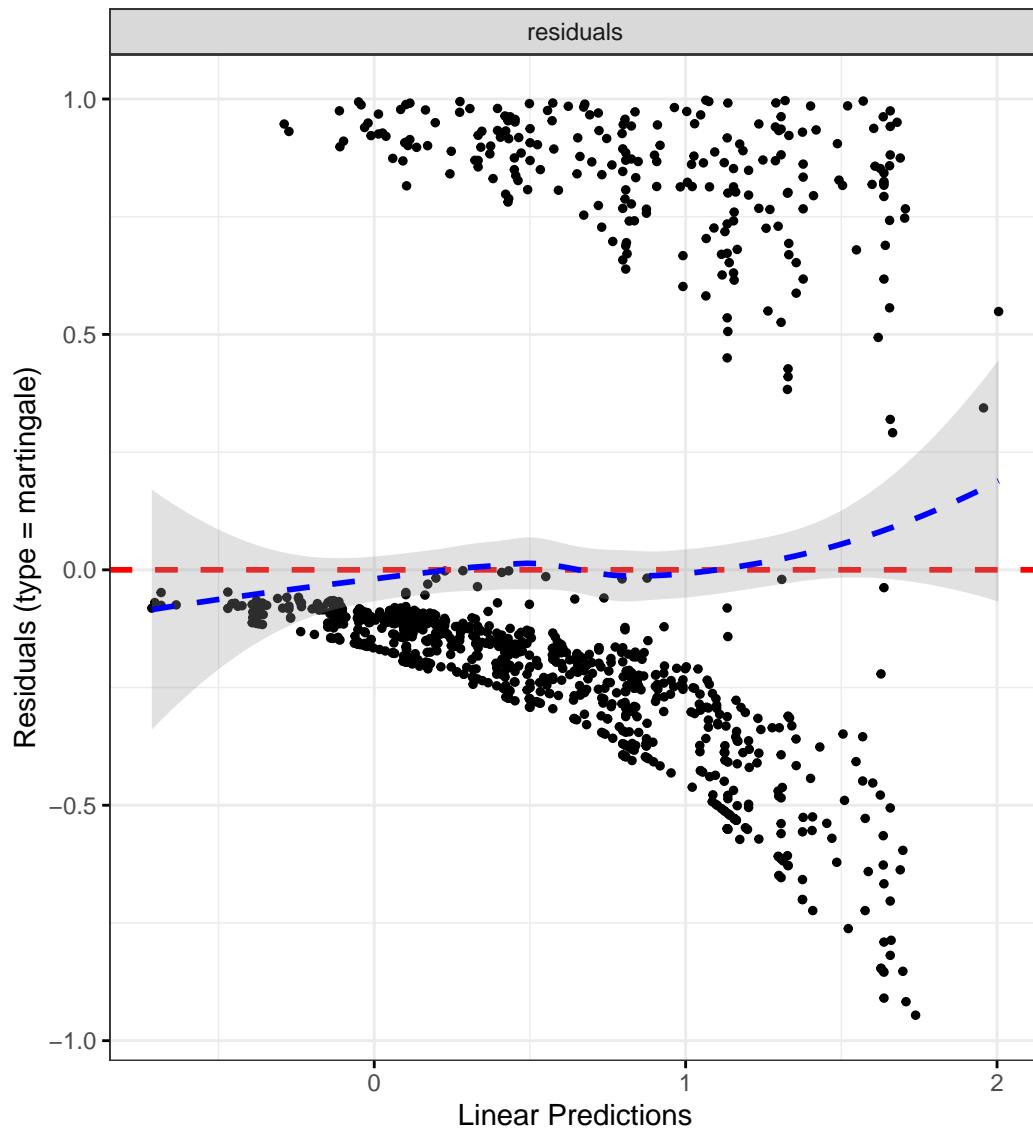
	Chi-squared statistic	DF	P-value
Sex	0.2572	0.9867	0.6065
IMD	4.2175	3.9404	0.3689
cat	8.8719	1.9846	0.0116
Biologic	0.0442	0.9583	0.8196
GLOBAL	14.1651	20.5809	0.8464

#### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



```
`geom_smooth()` using formula = 'y ~ x'
```



### Ulcerative colitis

#### Patient-reported flare

```
# Transform biologic variable and create categorized version
flare.uc.df <- flare.uc.df %>%
  mutate(Biologic = plyr::mapvalues(Biologic,
    from = c("Current",
    "Previously",
```

```

        "Never prescribed"),
      to = c("Prescribed",
            "Not prescribed",
            "Not prescribed")))

# Create categorized version for survival analysis
flare.uc.df$Biologic_cat <- flare.uc.df$Biologic

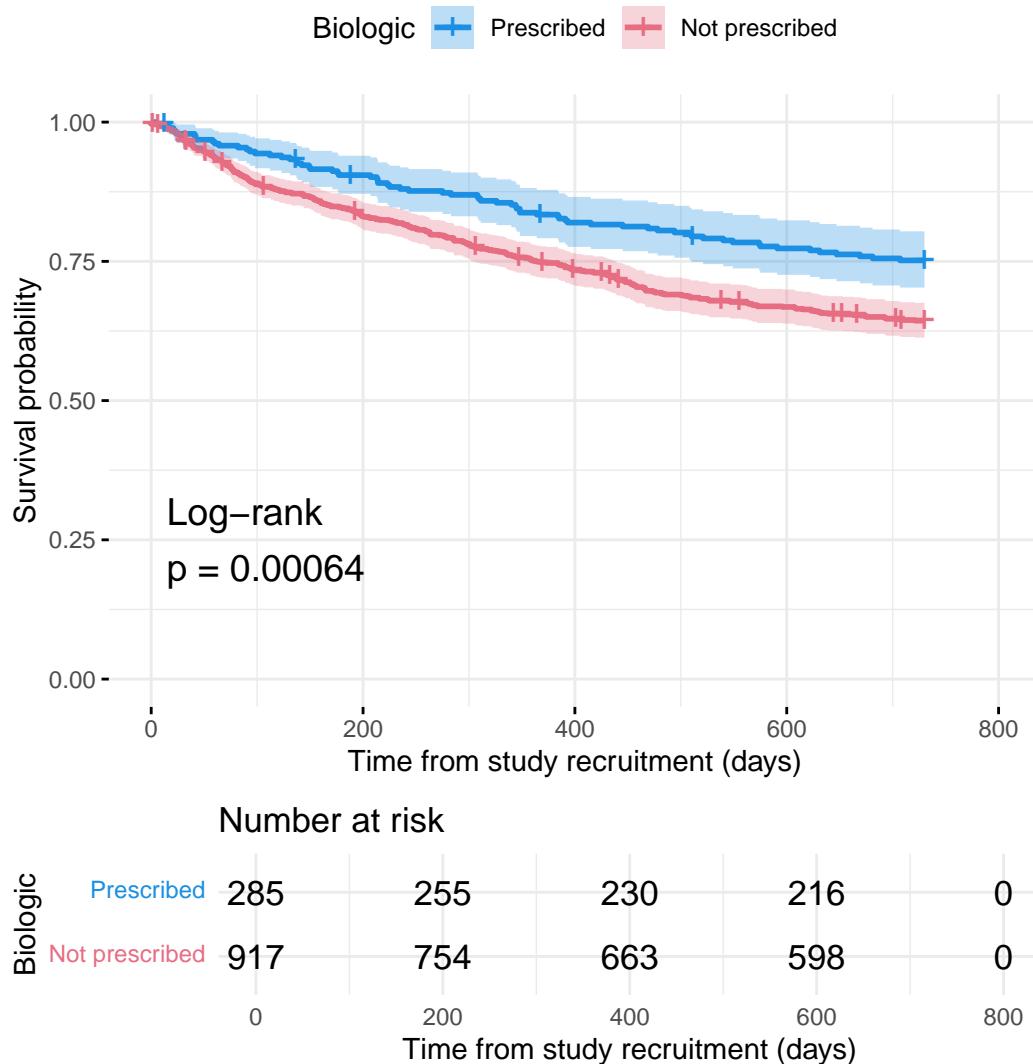
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.uc.df,
  var_name = "Biologic",
  outcome_time = "softflare_time",
  outcome_event = "softflare",
  legend_title = "Biologic",
  plot_base_path = "plots/uc/soft-flare/ibd/biologic",
  break_time_by = 200,
  palette = c("#1A8FE3", "#E76D83")
)

# Cox model
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + IMD + cat + Biologic + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

# Display plot and model summary
analysis_result$plot

```

### Time to patient-reported flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.5457	1.2492	1.9126	0.0001
IMD2	1.2605	0.7970	1.9934	0.3223
IMD3	1.0597	0.6766	1.6596	0.8001

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	1.4046	0.9150	2.1560	0.1202
IMD5	1.1438	0.7503	1.7438	0.5322
catFC 50-250	1.5415	1.2061	1.9702	0.0005
catFC > 250	2.1682	1.6622	2.8283	0.0000
BiologicNot prescribed	1.6853	1.2700	2.2363	0.0003

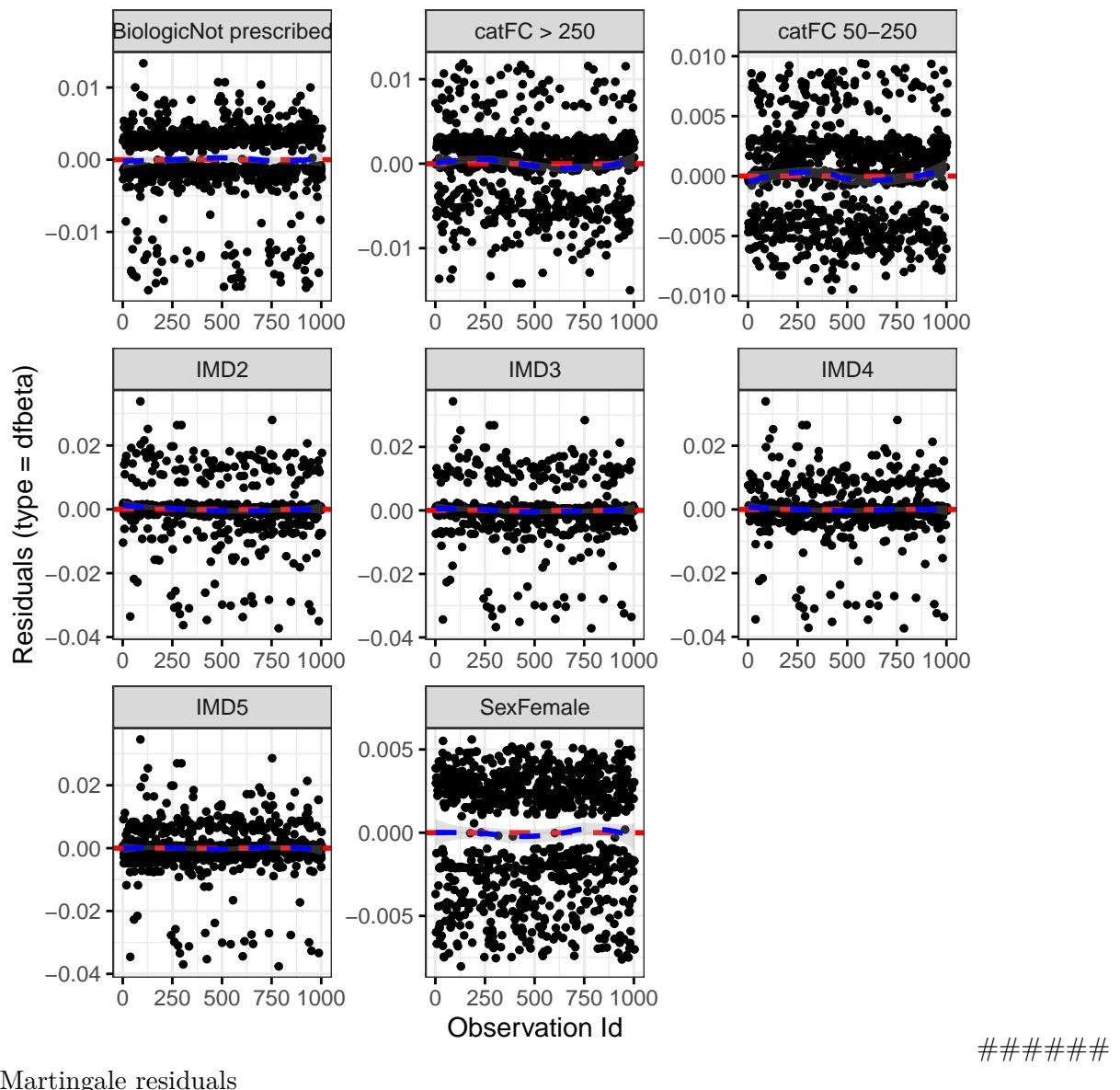
Diagnostics:

### Proportional hazards assumption test

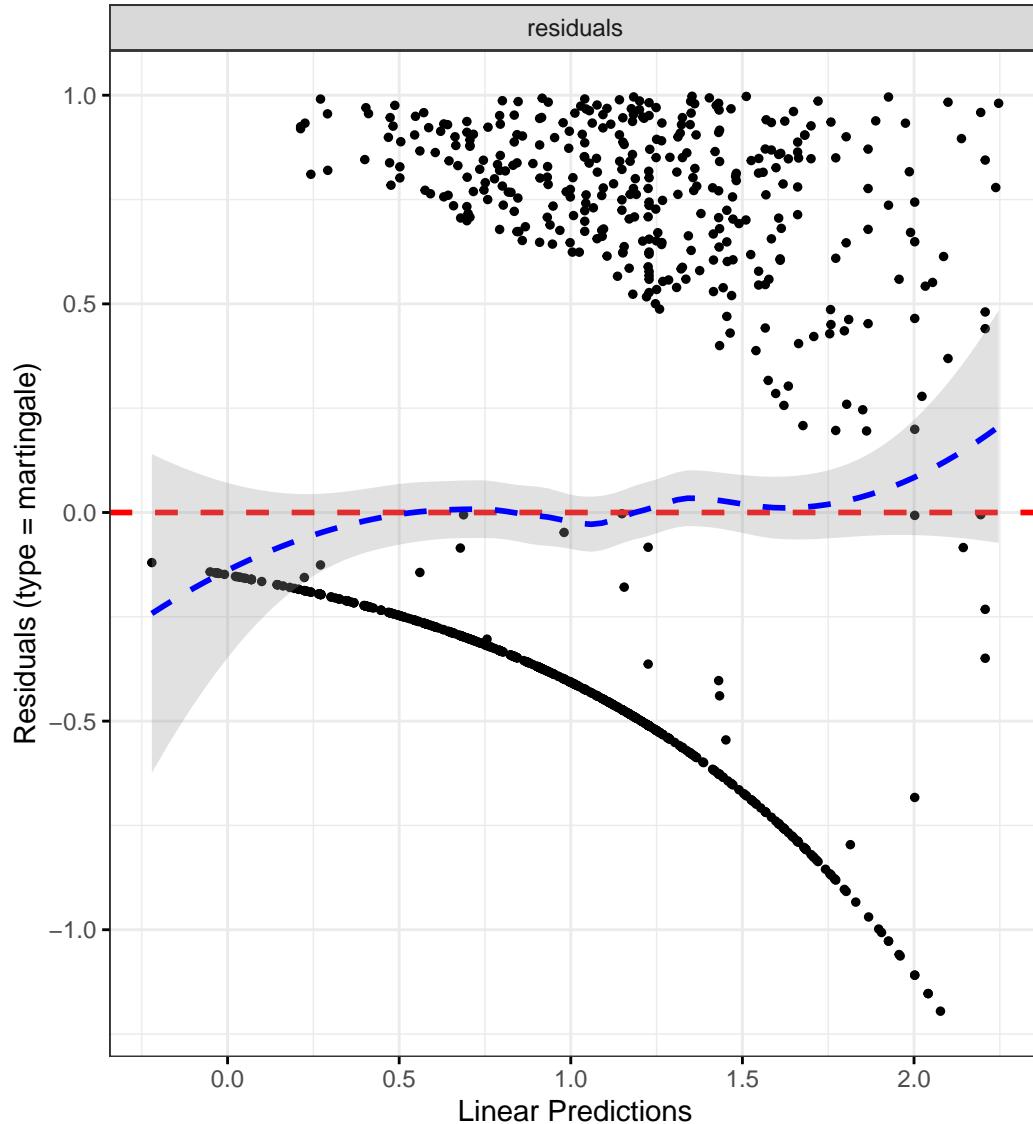
	Chi-squared statistic	DF	P-value
Sex	1.2814	0.9920	0.2554
IMD	4.0168	3.9495	0.3963
cat	5.5881	1.9739	0.0597
Biologic	0.8627	0.9374	0.3312
GLOBAL	12.2828	17.5375	0.8105

### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



```
`geom_smooth()` using formula = 'y ~ x'
```



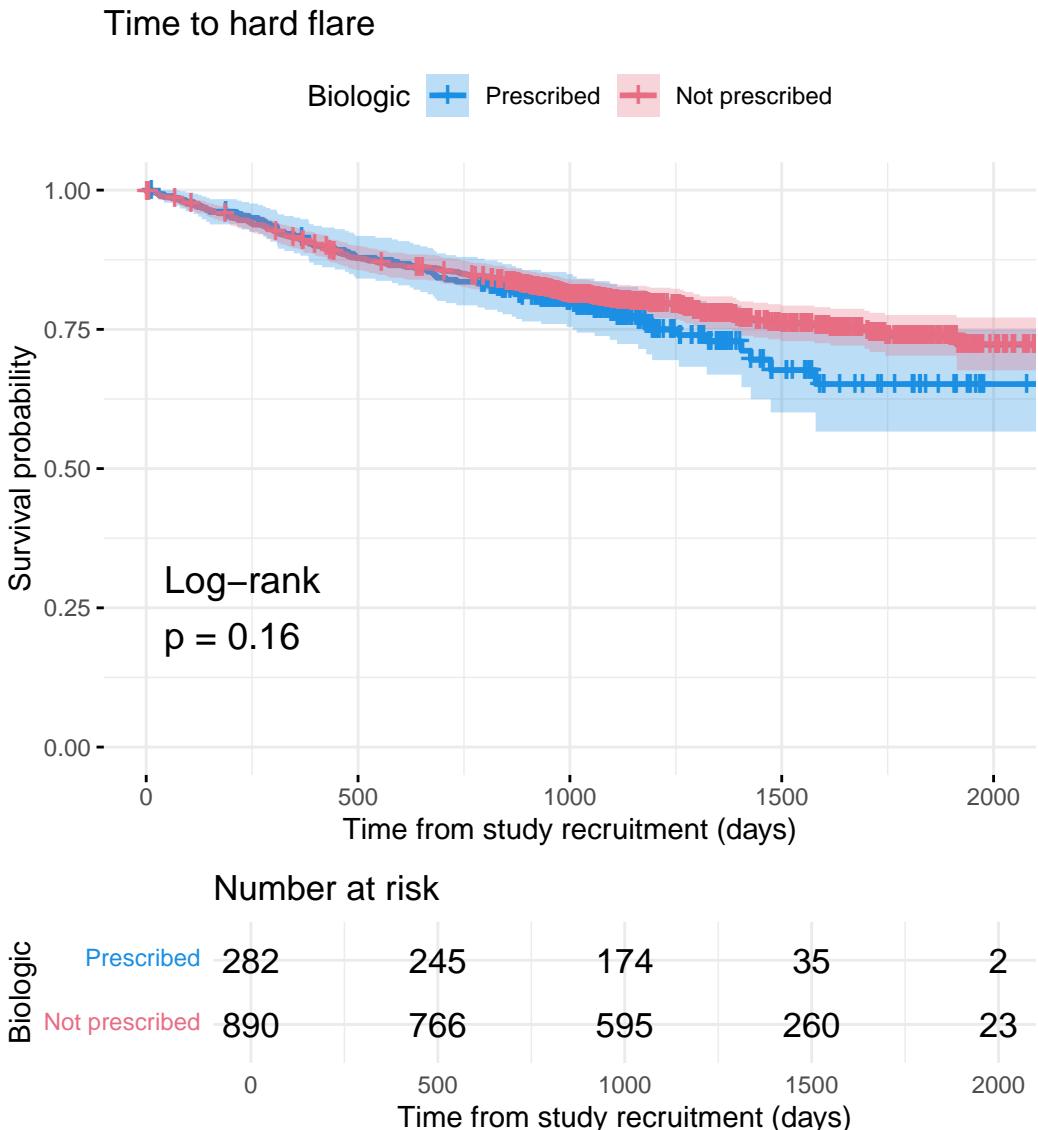
### Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.uc.df,
  var_name = "Biologic",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "Biologic",
```

```
plot_base_path = "plots/uc/hard-flare/ibd/biologic",
break_time_by = 500,
palette = c("#1A8FE3", "#E76D83")
)

# Cox model
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + IMD + cat + Biologic + frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

# Display plot and model summary
analysis_result$plot
```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.3299	1.0239	1.7273	0.0326
IMD2	1.4037	0.7829	2.5167	0.2549
IMD3	1.3959	0.7936	2.4554	0.2471

Variable	HR	Lower 95%	Upper 95%	P-value
IMD4	1.7610	1.0200	3.0403	0.0422
IMD5	1.3247	0.7702	2.2783	0.3095
catFC 50-250	2.0523	1.5024	2.8035	0.0000
catFC > 250	3.2198	2.3238	4.4613	0.0000
BiologicNot prescribed	0.8410	0.6128	1.1542	0.2838

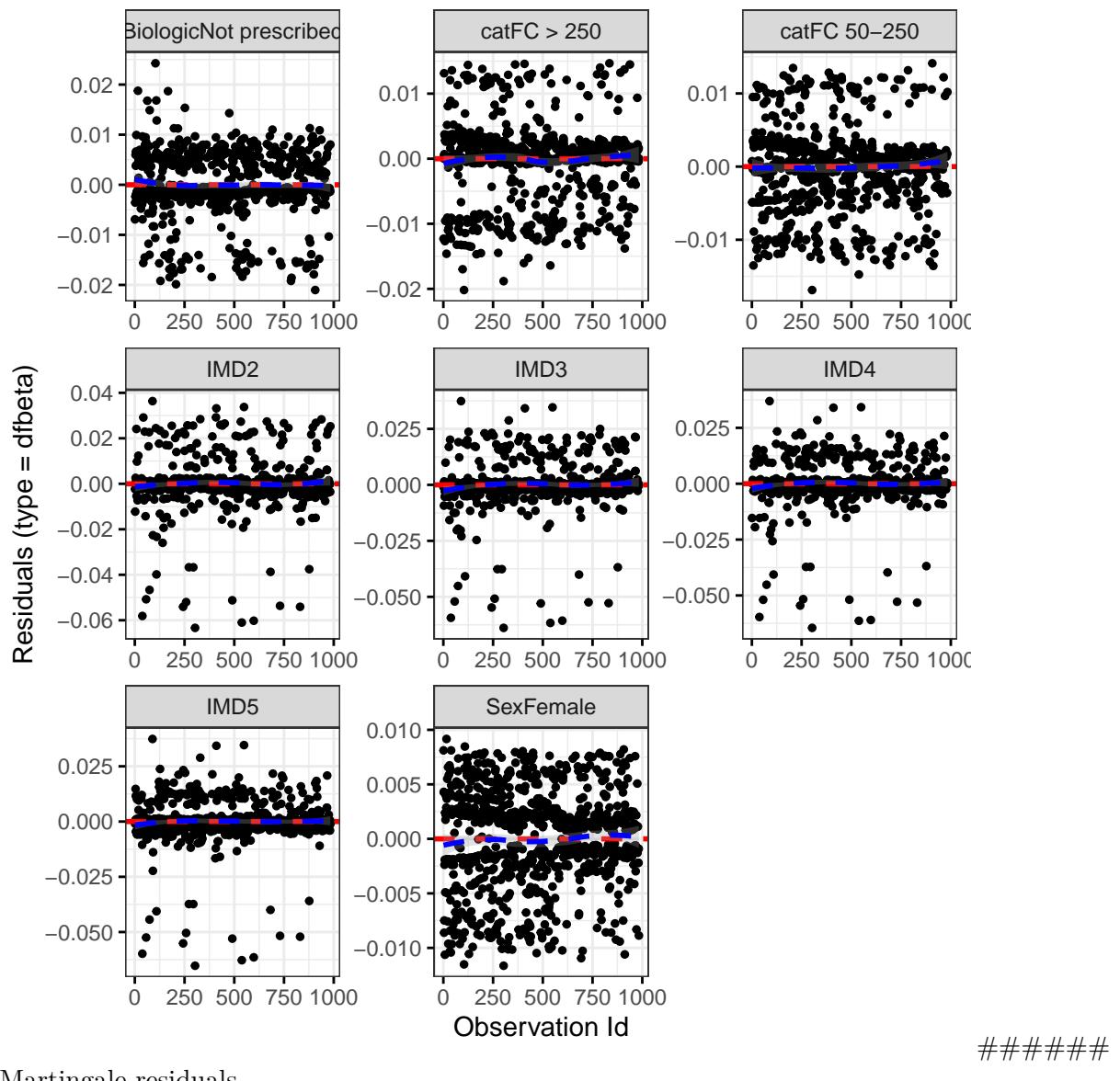
Diagnostics:

#### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.1497	0.9849	0.6927
IMD	2.5933	3.9365	0.6182
cat	4.3484	1.9676	0.1106
Biologic	3.8871	0.9212	0.0431
GLOBAL	11.4032	24.4735	0.9884

#### DF betas

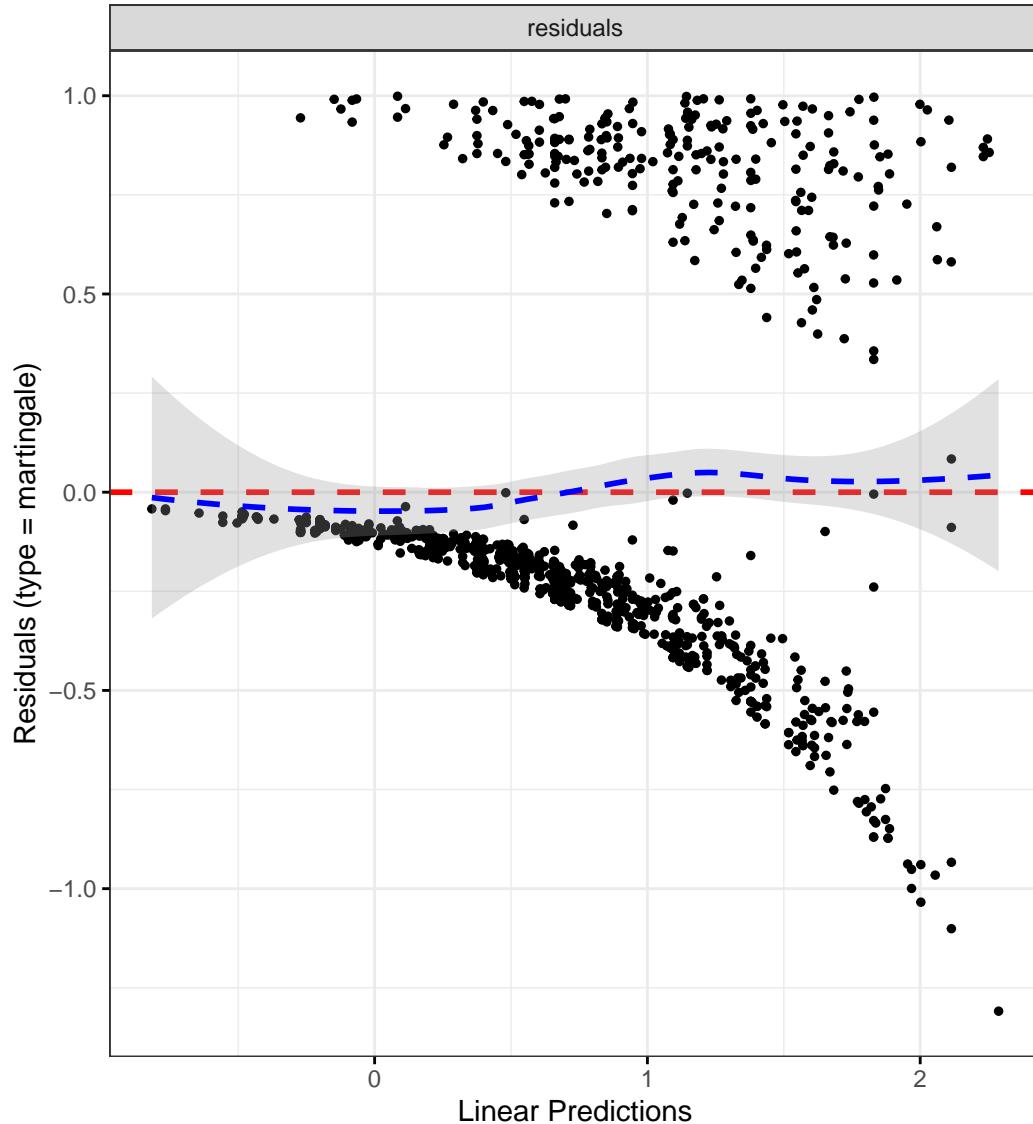
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



### Variables only relevant to Crohn's disease

```
demo.cd <- readRDS(paste0(paths$outdir, "demo-cd.RDS"))
demo.cd <- demo.cd[, c(
  "ParticipantNo",
  "Surgery",
  "Location",
  "L4",
```

```

    "Behaviour",
    "Perianal",
    "HBI",
    "PRO2"
  )]
flare.cd.df <- merge(flare.cd.df,
  demo.cd,
  by = "ParticipantNo",
  all.x = TRUE,
  all.y = FALSE
)

# Create categorized versions for survival analysis
flare.cd.df$Surgery_cat <- flare.cd.df$Surgery
flare.cd.df$Location_cat <- flare.cd.df$Location
flare.cd.df$L4_cat <- flare.cd.df$L4
flare.cd.df$Behaviour_cat <- flare.cd.df$Behaviour
flare.cd.df$Perianal_cat <- flare.cd.df$Perianal

```

## Surgery

### Patient-reported flare

```

# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Surgery",
  outcome_time = "softflare_time",
  outcome_event = "softflare",
  legend_title = "Previous surgery",
  plot_base_path = "plots/cd/soft-flare/ibd/surgery",
  break_time_by = 200,
  palette = c("#3DDC97", "#DD7373")
)

# Cox model
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + cat + IMD + Surgery +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),

```

```

  data = flare.cd.df
)
cd.clin.forest <- rbind(cd.clin.forest, get_HR(fit.me, "SurgeryYes"))

# Display plot and model summary
analysis_result$plot

```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.9970	1.5704	2.5394	0.0000
catFC 50-250	1.5774	1.2210	2.0379	0.0005
catFC > 250	2.3834	1.7891	3.1751	0.0000
IMD2	0.8921	0.5692	1.3983	0.6186
IMD3	0.8326	0.5264	1.3170	0.4336
IMD4	0.8750	0.5640	1.3575	0.5512
IMD5	0.9325	0.6106	1.4240	0.7462
SurgeryYes	1.0130	0.8111	1.2652	0.9092

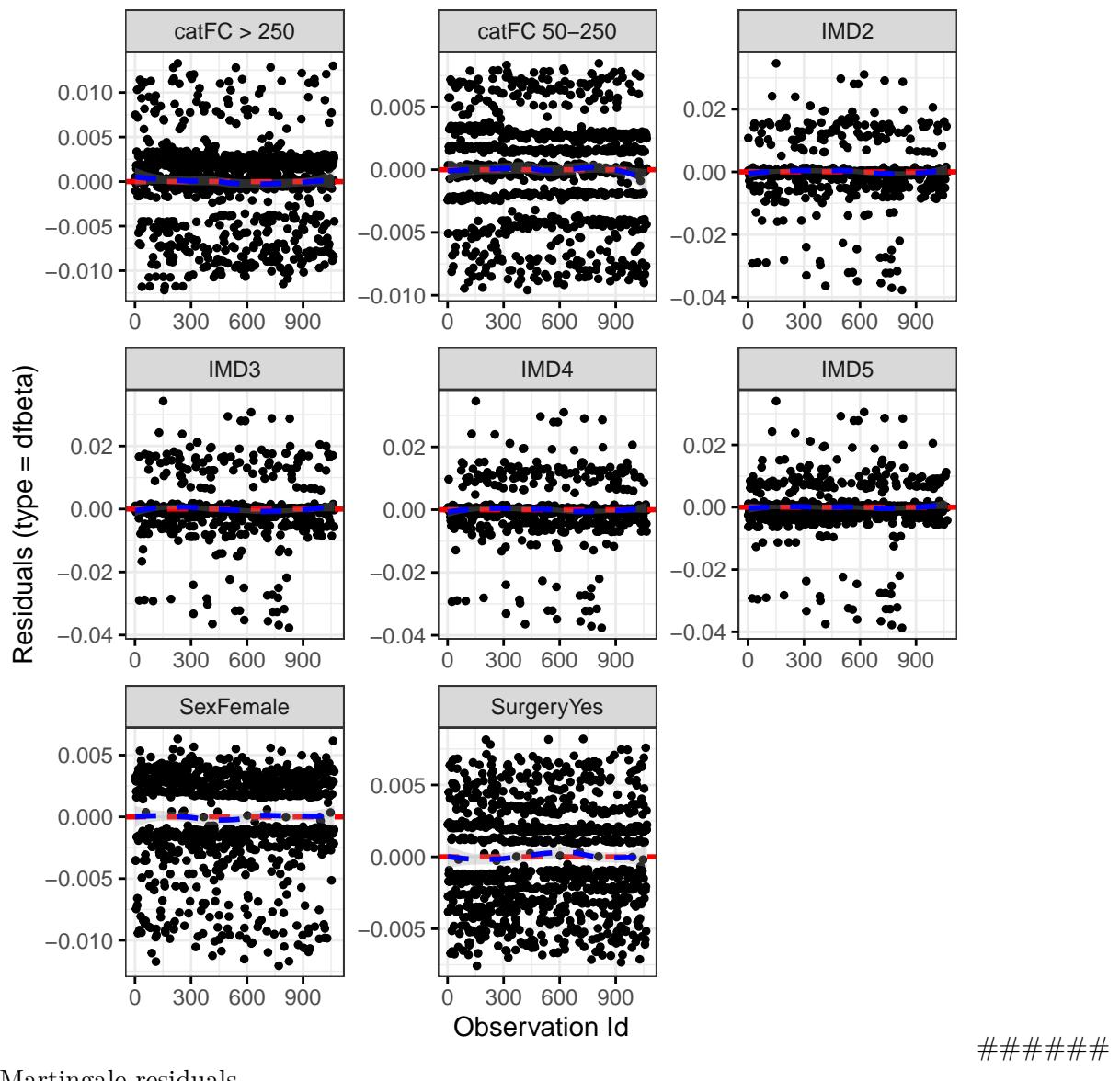
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.3138	0.9943	0.5730
cat	2.4348	1.9874	0.2936
IMD	5.3005	3.9665	0.2539
Surgery	0.6320	0.9915	0.4234
GLOBAL	8.9435	11.8427	0.6958

### DF betas

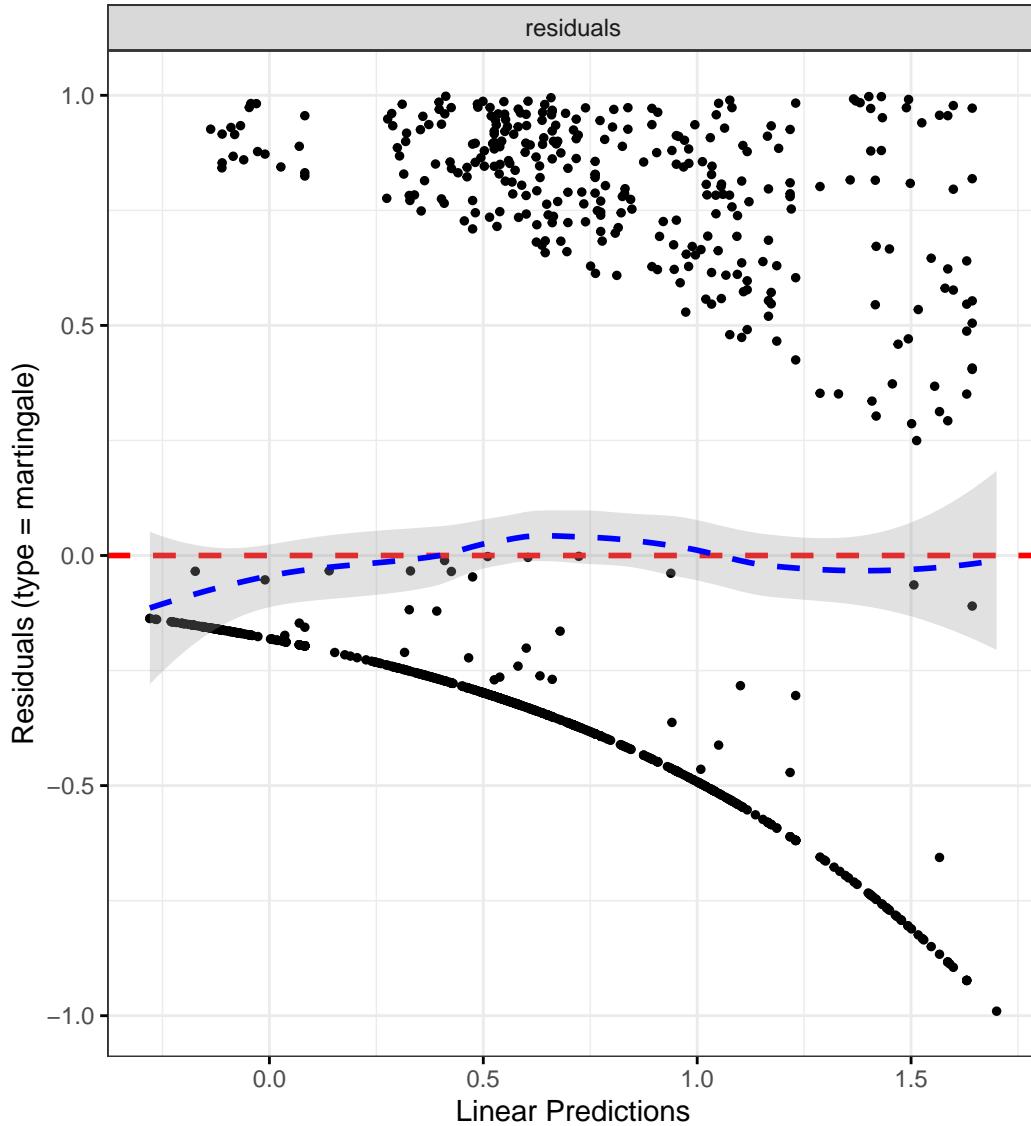
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



## Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Surgery",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "Previous surgery",
```

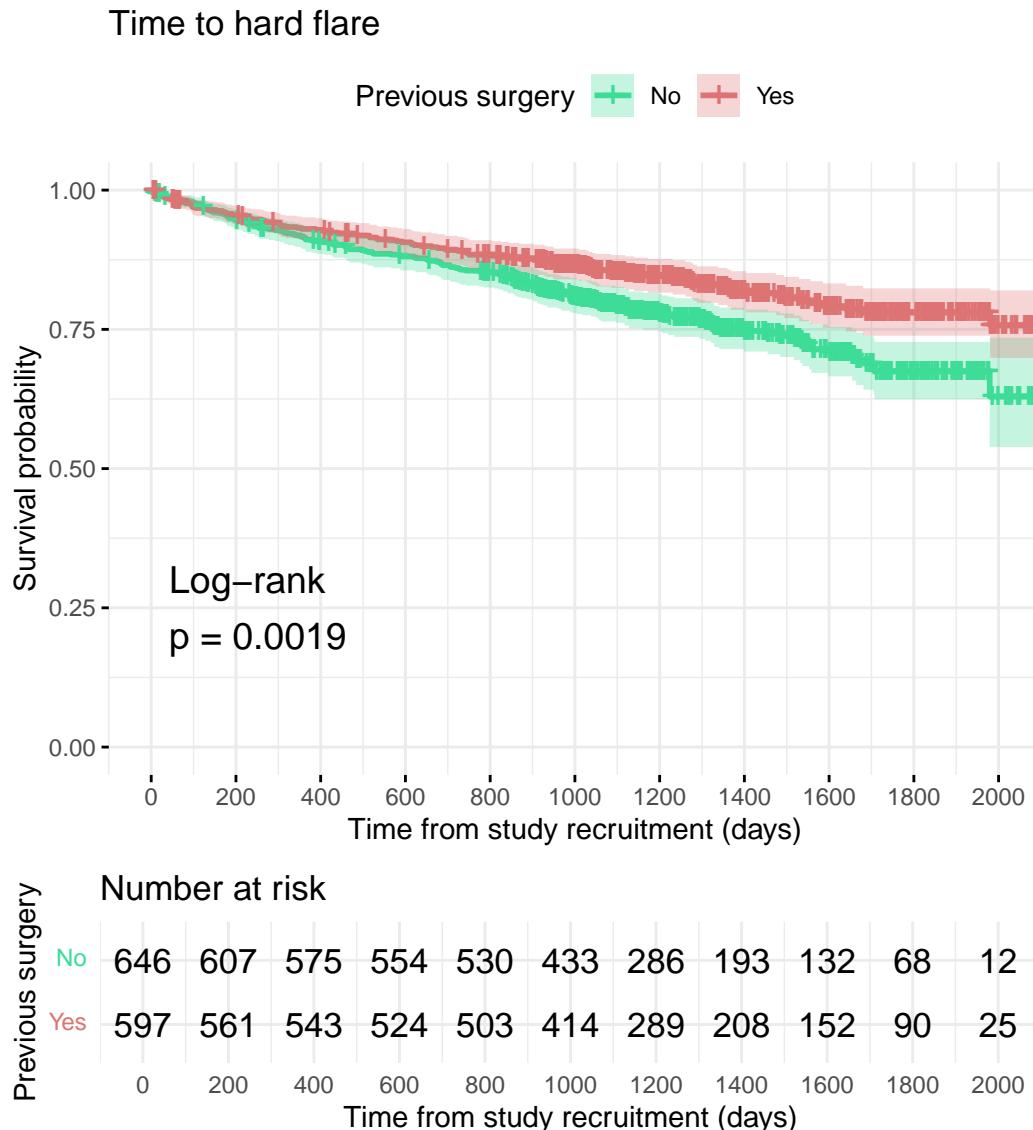
```

plot_base_path = "plots/cd/hard-flare/ibd/surgery",
break_time_by = 200,
palette = c("#3DDC97", "#DD7373")
)

# Cox model
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + cat + IMD + Surgery +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)
cd.hard.forest <- rbind(cd.hard.forest, get_HR(fit.me, "SurgeryYes"))

# Display plot and model summary
analysis_result$plot

```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.4133	1.0743	1.8592	0.0134
catFC 50-250	2.0584	1.4977	2.8292	0.0000
catFC > 250	3.2393	2.2905	4.5812	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	0.8768	0.5106	1.5058	0.6338
IMD3	0.8923	0.5114	1.5569	0.6882
IMD4	0.8420	0.4913	1.4430	0.5314
IMD5	0.8745	0.5201	1.4706	0.6132
SurgeryYes	0.7516	0.5744	0.9835	0.0374

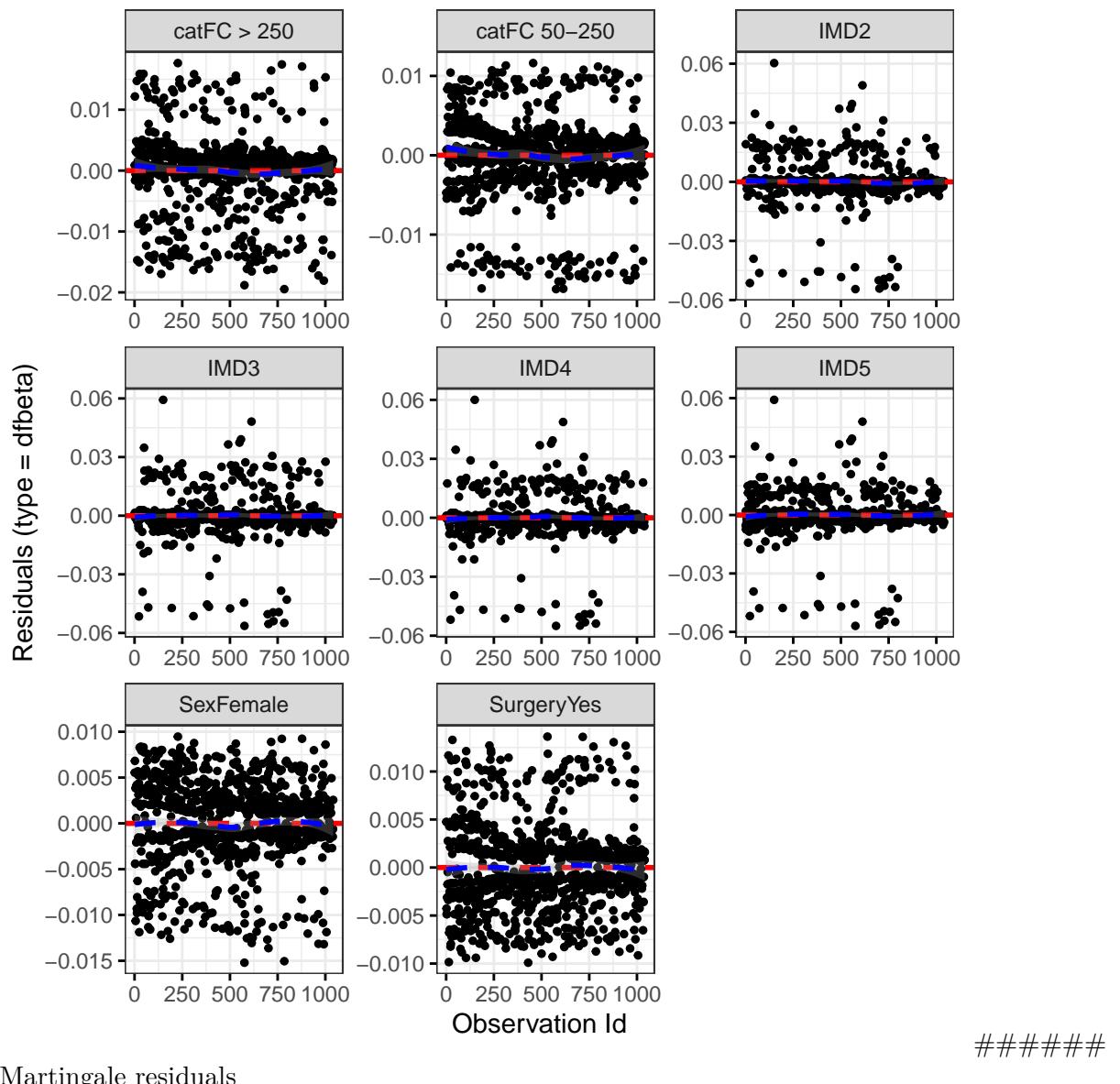
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.3300	0.9858	0.5597
cat	8.5496	1.9843	0.0137
IMD	4.0006	3.9415	0.3973
Surgery	2.5667	0.9872	0.1073
GLOBAL	17.3483	20.3096	0.6492

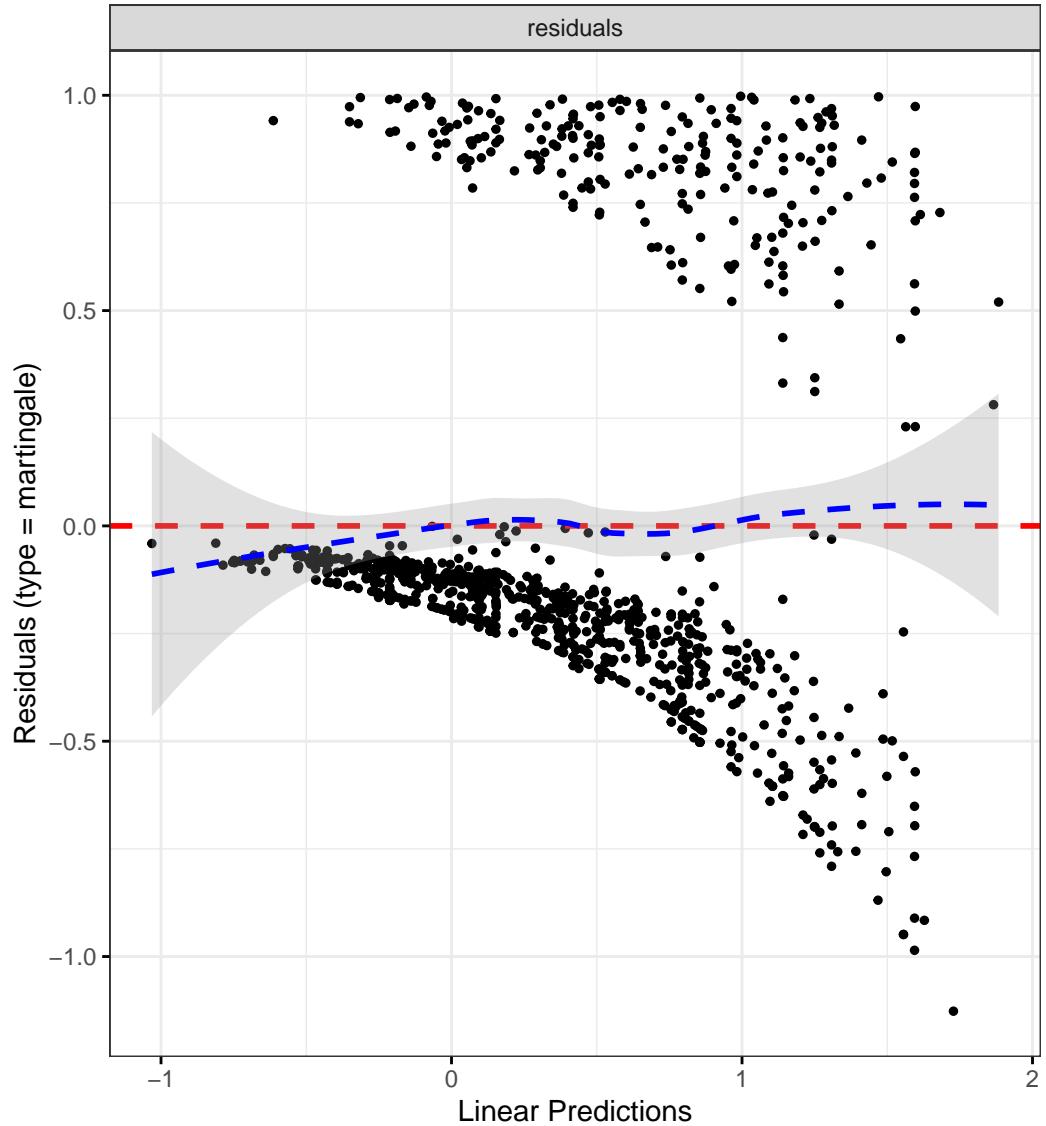
### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



### Montreal location

#### Patient-reported flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Location",
  outcome_time = "softflare_time",
```

```

outcome_event = "softflare",
legend_title = "Montreal location",
plot_base_path = "plots/cd/soft-flare/ibd/location",
break_time_by = 200,
palette = c("#3B3561", "#5BC0EB", "#FFD166", "#E4572E")
)

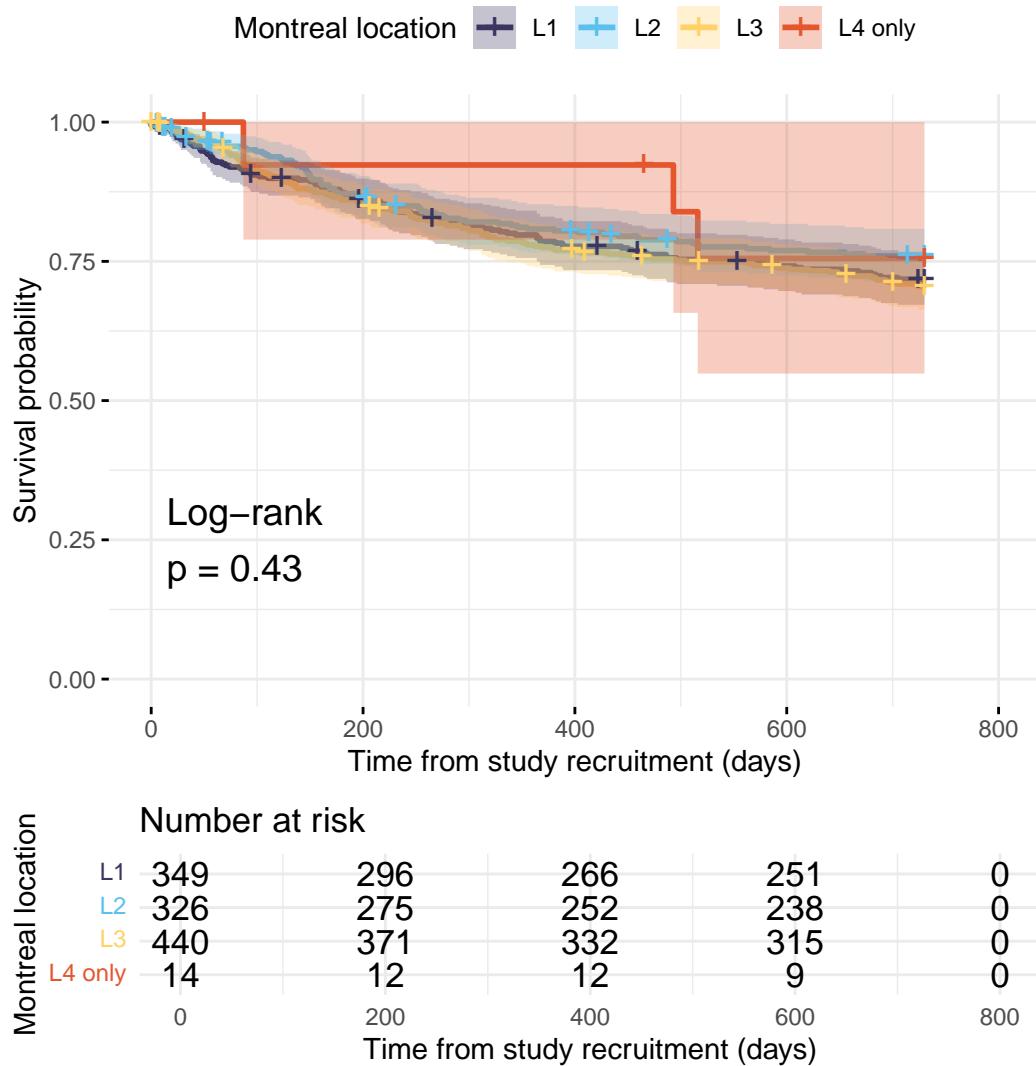
# Cox model
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + cat + IMD + Location +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

cd.clin.forest <- rbind(
  cd.clin.forest,
  get_HR(fit.me, c("LocationL2", "LocationL3"))
)

# Display plot and model summary
analysis_result$plot

```

### Time to patient-reported flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	2.0379	1.5673	2.6498	0.0000
catFC 50-250	1.6798	1.2692	2.2231	0.0003
catFC > 250	2.4535	1.8007	3.3429	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	0.9218	0.5727	1.4836	0.7372
IMD3	0.8187	0.5015	1.3365	0.4237
IMD4	0.8363	0.5225	1.3386	0.4564
IMD5	0.9443	0.6018	1.4816	0.8029
LocationL2	0.8397	0.6095	1.1570	0.2855
LocationL3	1.0929	0.8212	1.4543	0.5425
LocationL4 only	0.9482	0.2984	3.0126	0.9281

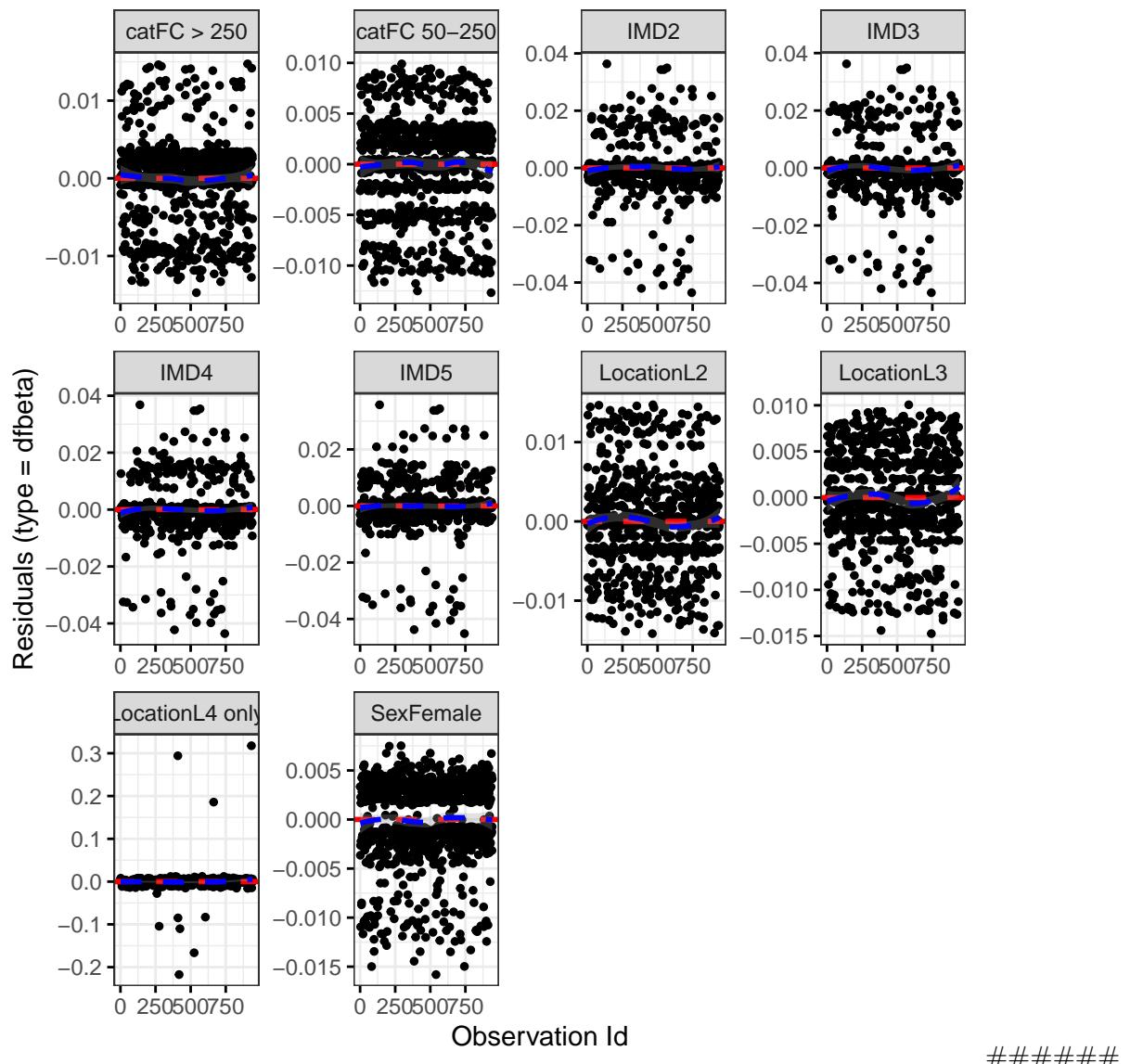
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.2323	0.9980	0.6290
cat	3.2675	1.9968	0.1947
IMD	5.1401	3.9898	0.2720
Location	1.1604	2.9932	0.7614
GLOBAL	9.8488	10.7869	0.5251

### DF betas

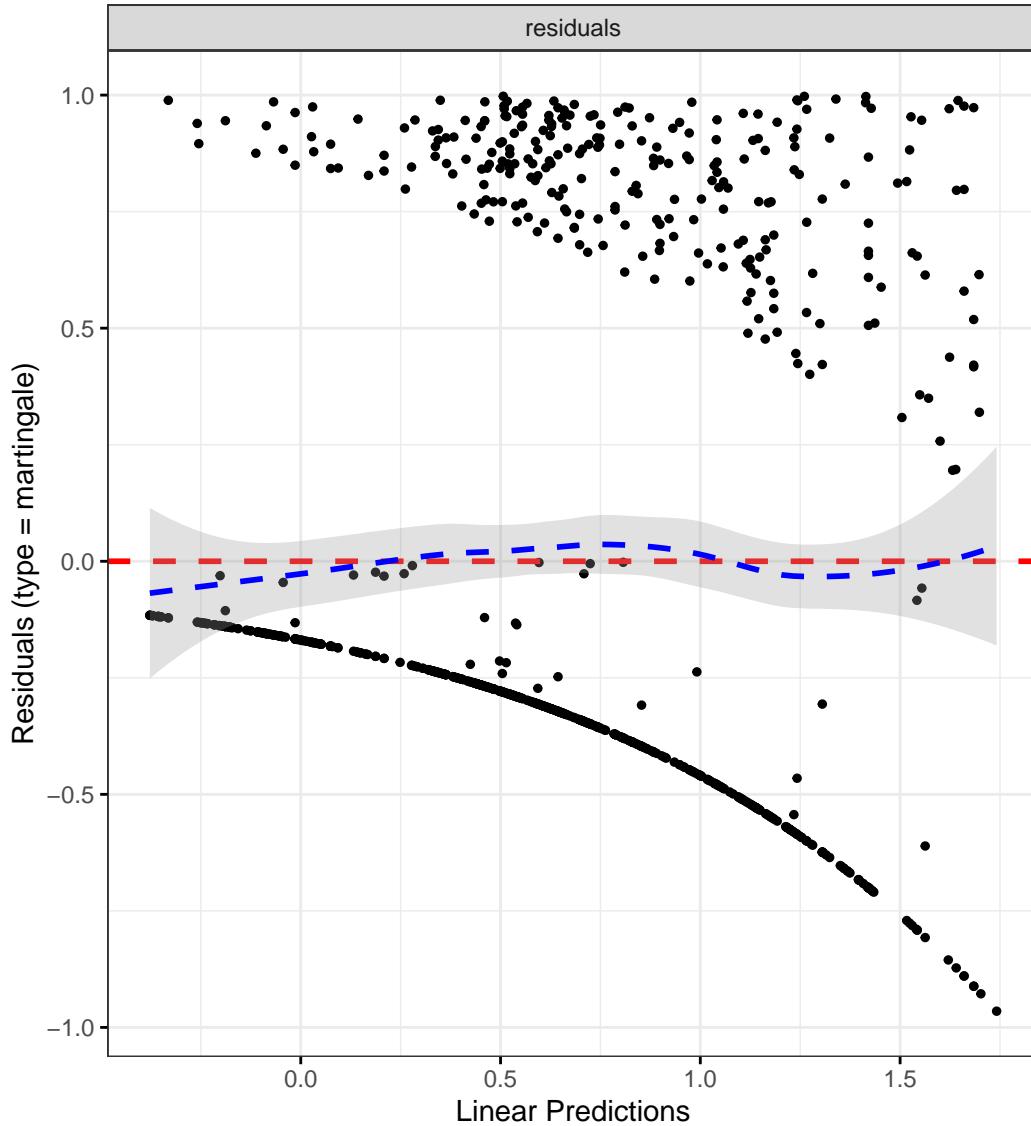
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



## Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Location",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "Montreal location",
```

```

plot_base_path = "plots/cd/hard-flare/ibd/location",
break_time_by = 200,
palette = c("#3B3561", "#5BC0EB", "#FFD166", "#E4572E")
)

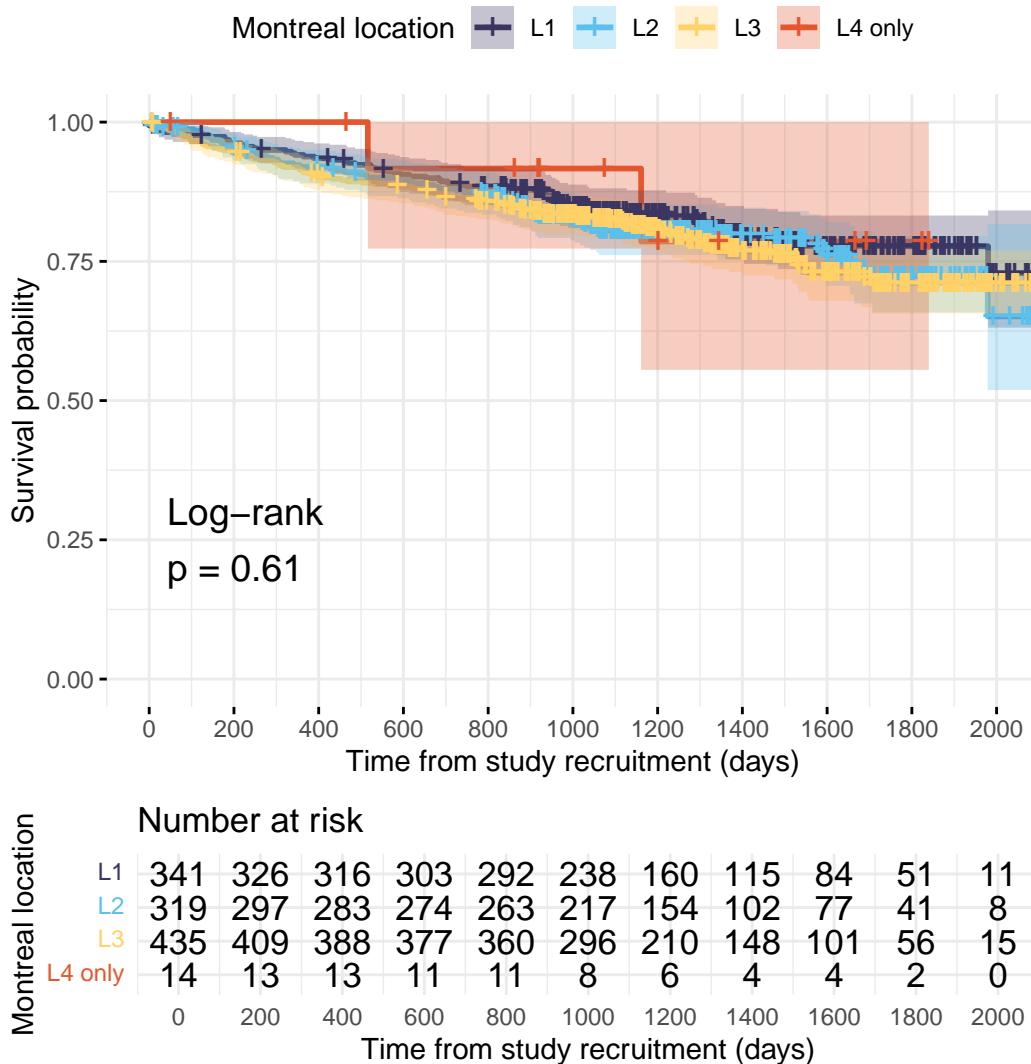
# Cox model
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + cat + IMD + Location +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

cd.hard.forest <- rbind(
  cd.hard.forest,
  get_HR(fit.me, c("LocationL2", "LocationL3"))
)

# Display plot and model summary
analysis_result$plot

```

### Time to hard flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.4122	1.0506	1.8982	0.0222
catFC 50-250	1.9337	1.3740	2.7214	0.0002
catFC > 250	3.1807	2.2094	4.5789	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	0.8819	0.4941	1.5742	0.6707
IMD3	0.9504	0.5256	1.7185	0.8663
IMD4	0.9099	0.5129	1.6144	0.7470
IMD5	0.9891	0.5687	1.7202	0.9691
LocationL2	1.2964	0.8834	1.9025	0.1846
LocationL3	1.3328	0.9285	1.9132	0.1193
LocationL4 only	1.2113	0.2871	5.1108	0.7941

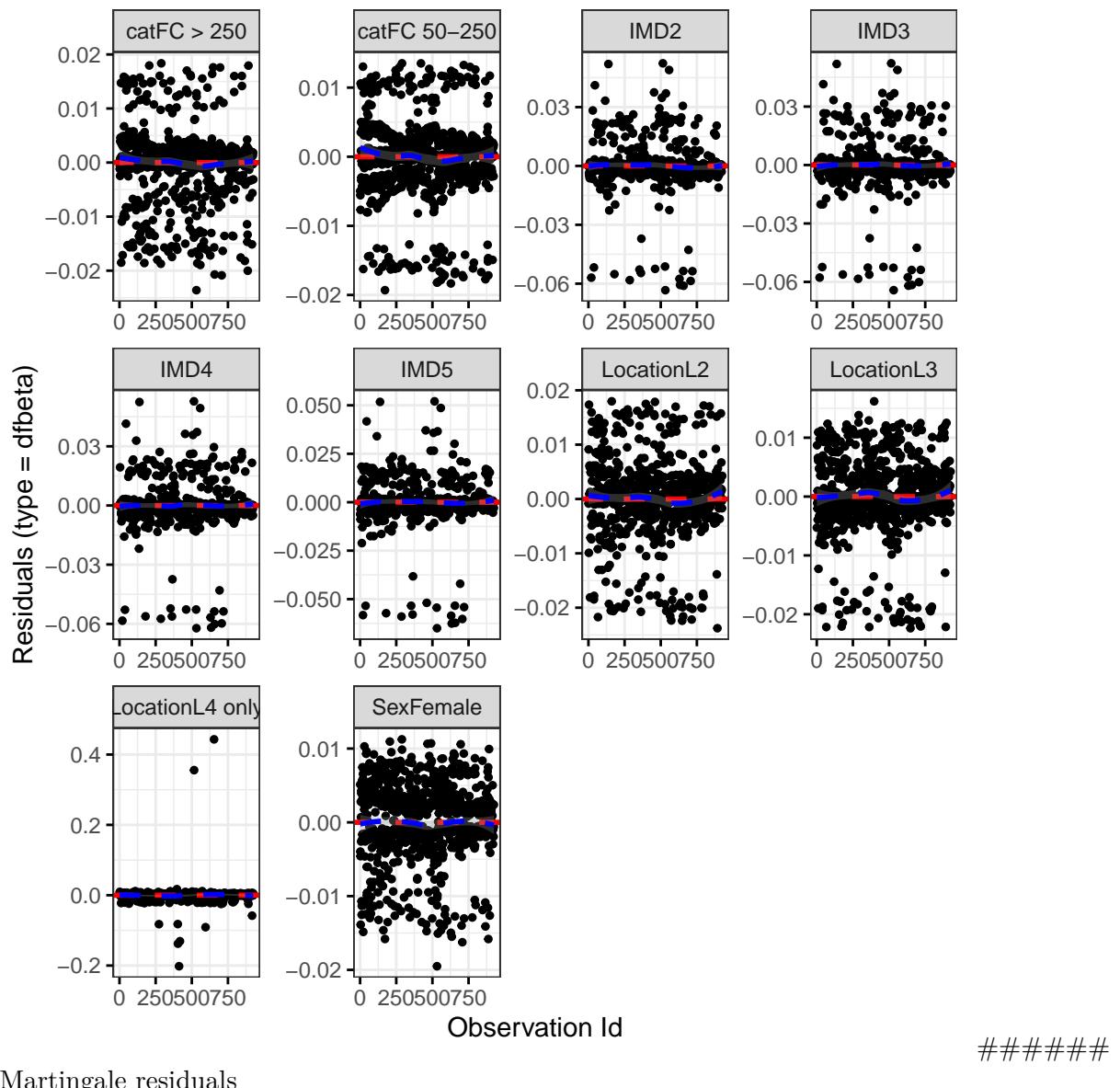
Diagnostics:

### Proportional hazards assumption test

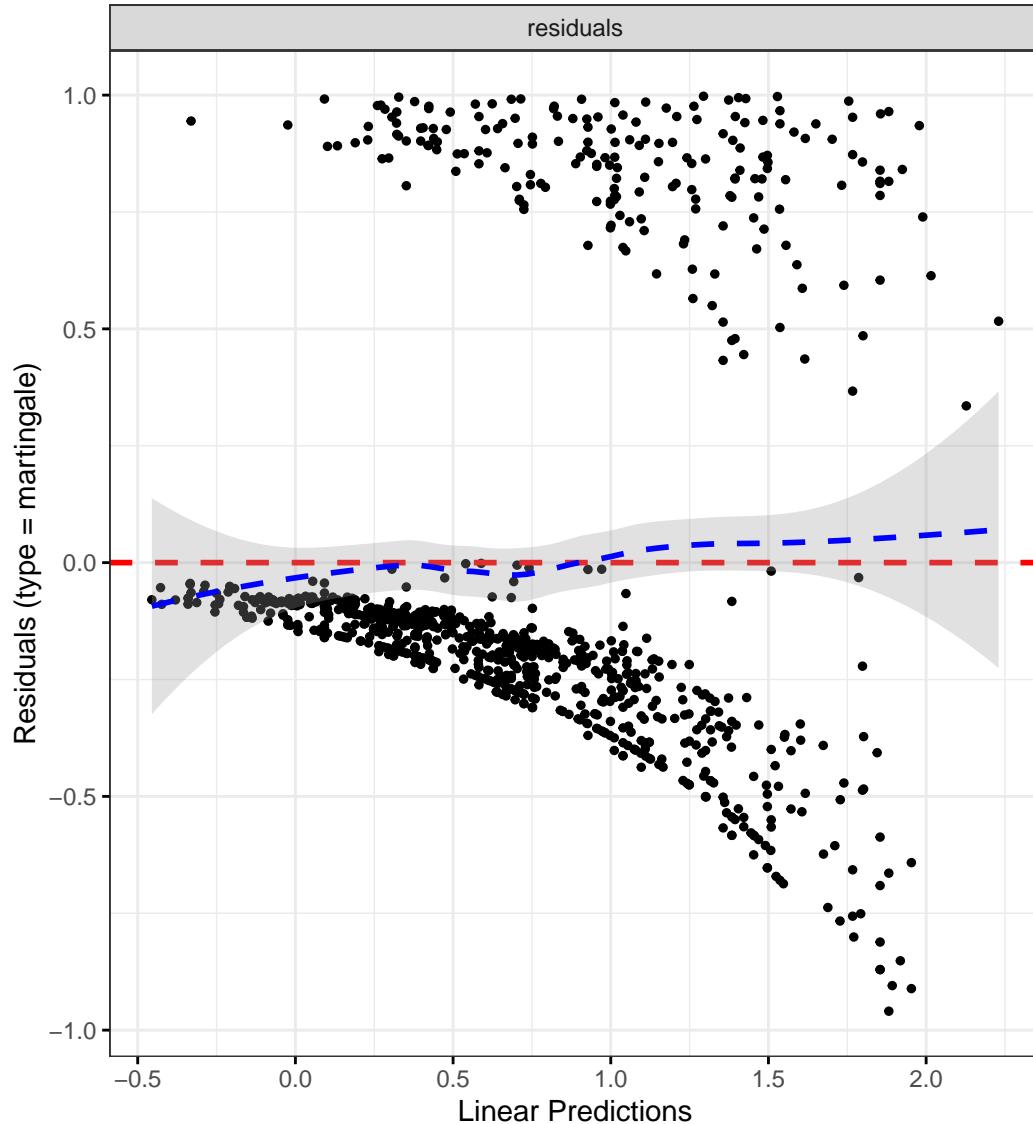
	Chi-squared statistic	DF	P-value
Sex	0.1424	0.9830	0.6991
cat	6.3620	1.9837	0.0409
IMD	3.8329	3.9412	0.4203
Location	0.5903	2.9550	0.8941
GLOBAL	11.3705	21.5180	0.9629

### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



```
`geom_smooth()` using formula = 'y ~ x'
```



## Montreal L4

### Patient-reported flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "L4",
  outcome_time = "softflare_time",
```

```

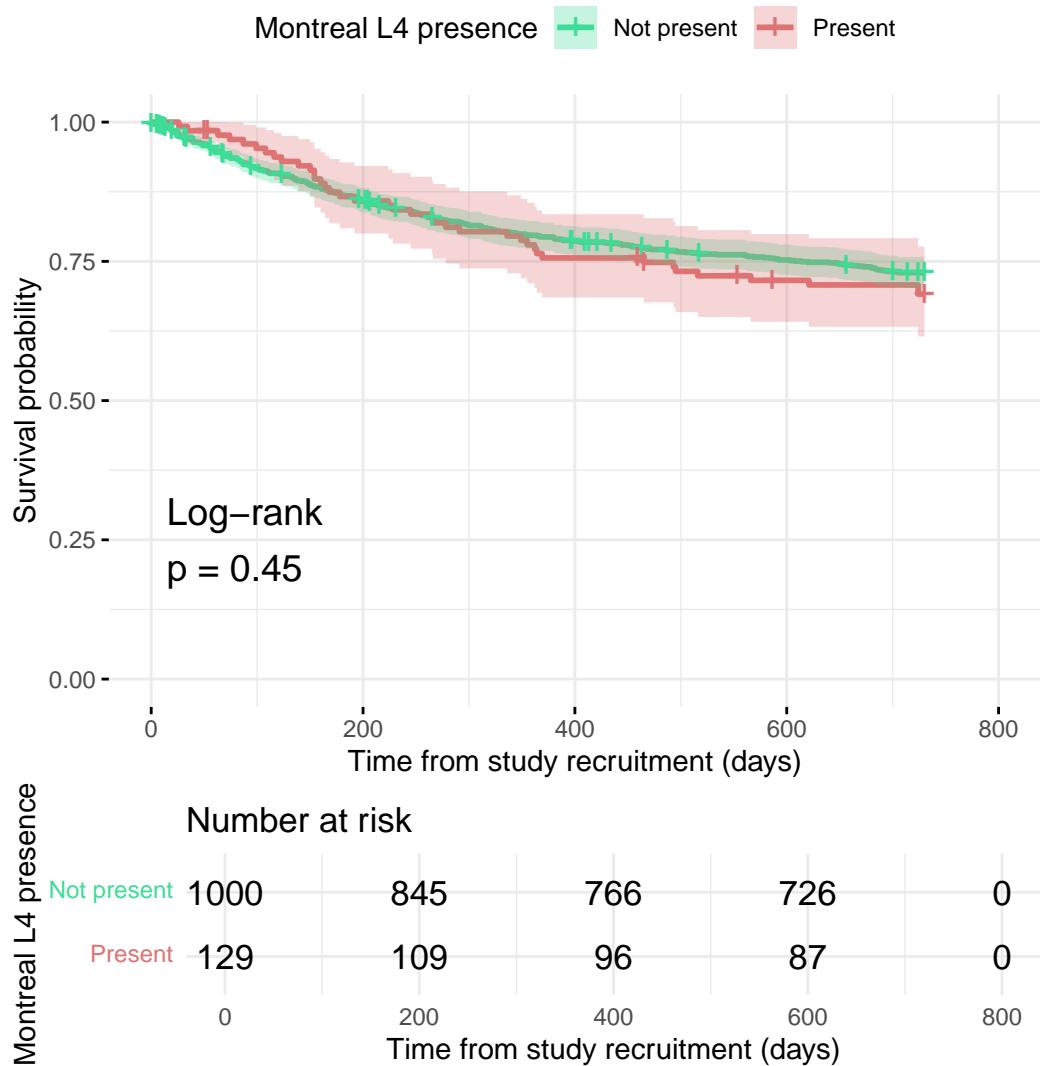
outcome_event = "softflare",
legend_title = "Montreal L4 presence",
plot_base_path = "plots/cd/soft-flare/ibd/l4",
break_time_by = 200,
palette = c("#3DDC97", "#DD7373")
)

# Cox model
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + cat + IMD + L4 +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)
cd.clin.forest <- rbind(
  cd.clin.forest,
  get_HR(fit.me, "L4Present")
)

# Display plot and model summary
analysis_result$plot

```

### Time to patient-reported flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	2.0268	1.5589	2.6350	0.0000
catFC 50-250	1.6892	1.2767	2.2352	0.0002
catFC > 250	2.3738	1.7423	3.2342	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	0.8815	0.5472	1.4199	0.6040
IMD3	0.7833	0.4794	1.2799	0.3296
IMD4	0.8069	0.5031	1.2941	0.3733
IMD5	0.9046	0.5765	1.4193	0.6625
L4Present	1.2936	0.9172	1.8244	0.1423

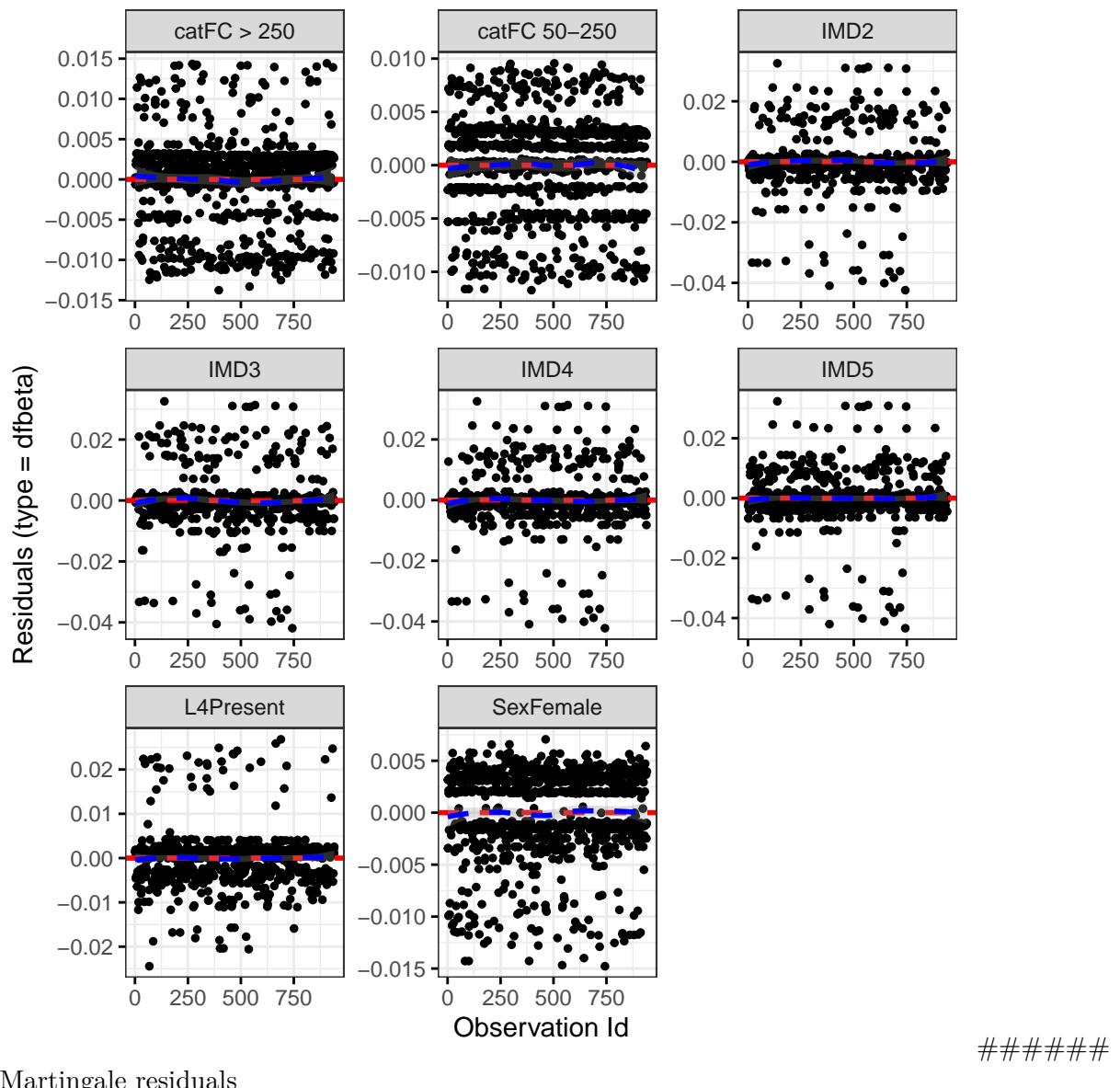
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.2367	0.9981	0.6258
cat	3.4174	1.9965	0.1806
IMD	5.2745	3.9899	0.2591
L4	2.9817	0.9968	0.0838
GLOBAL	12.5997	8.8065	0.1700

### DF betas

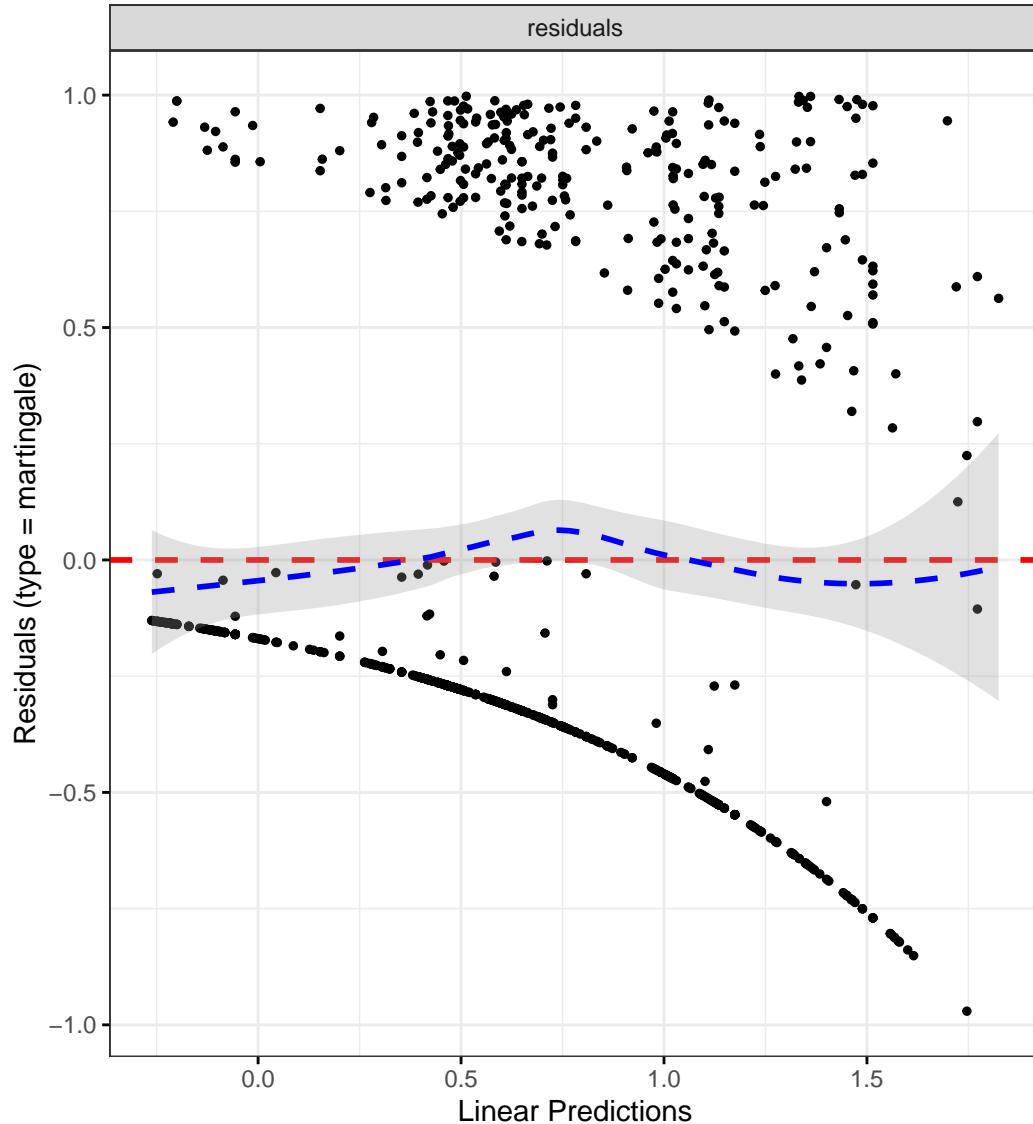
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



## Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "L4",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "Montreal L4 presence",
```

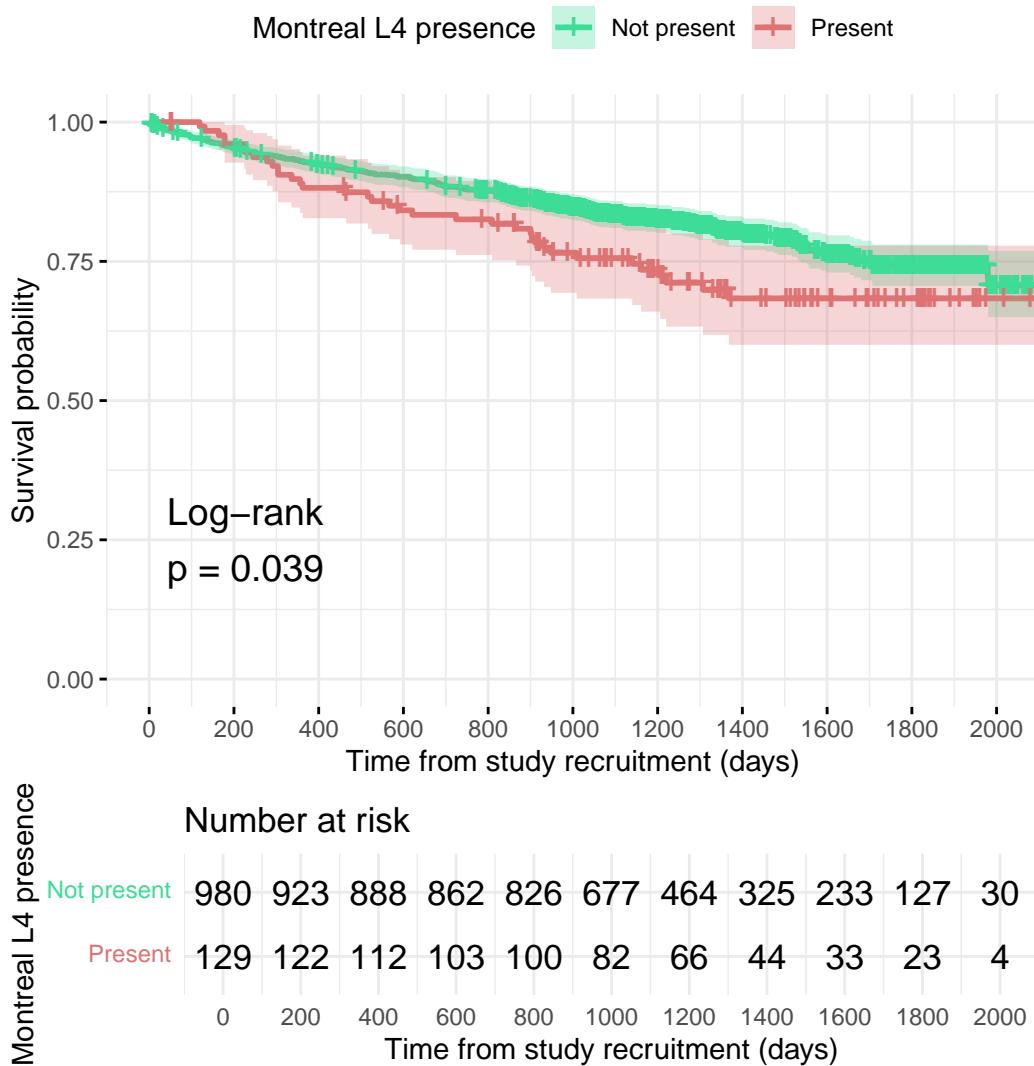
```
plot_base_path = "plots/cd/hard-flare/ibd/14",
break_time_by = 200,
palette = c("#3DDC97", "#DD7373")
)

# Cox model
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + cat + IMD + L4 +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

cd.hard.forest <- rbind(
  cd.hard.forest,
  get_HR(fit.me, "L4Present")
)

# Display plot and model summary
analysis_result$plot
```

### Time to hard flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.4360	1.0684	1.9301	0.0165
catFC 50-250	1.9358	1.3756	2.7240	0.0002
catFC > 250	3.1320	2.1758	4.5085	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	0.8257	0.4627	1.4736	0.5170
IMD3	0.8677	0.4791	1.5715	0.6396
IMD4	0.8321	0.4680	1.4794	0.5314
IMD5	0.9026	0.5188	1.5705	0.7169
L4Present	1.5676	1.0693	2.2981	0.0213

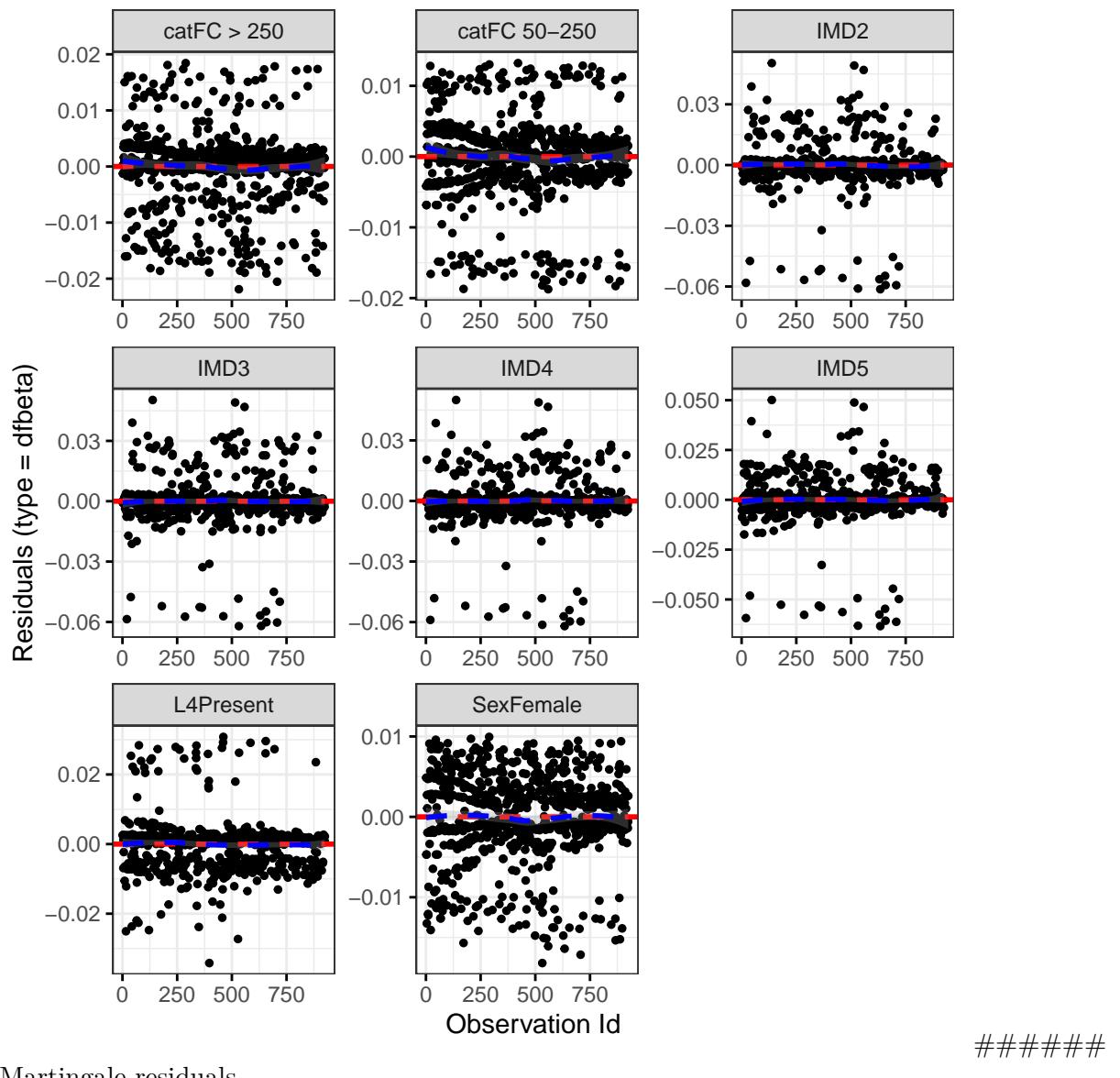
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.1595	0.9826	0.6825
cat	6.4740	1.9836	0.0386
IMD	4.0148	3.9425	0.3956
L4	0.0375	0.9606	0.8338
GLOBAL	10.7469	19.2888	0.9386

### DF betas

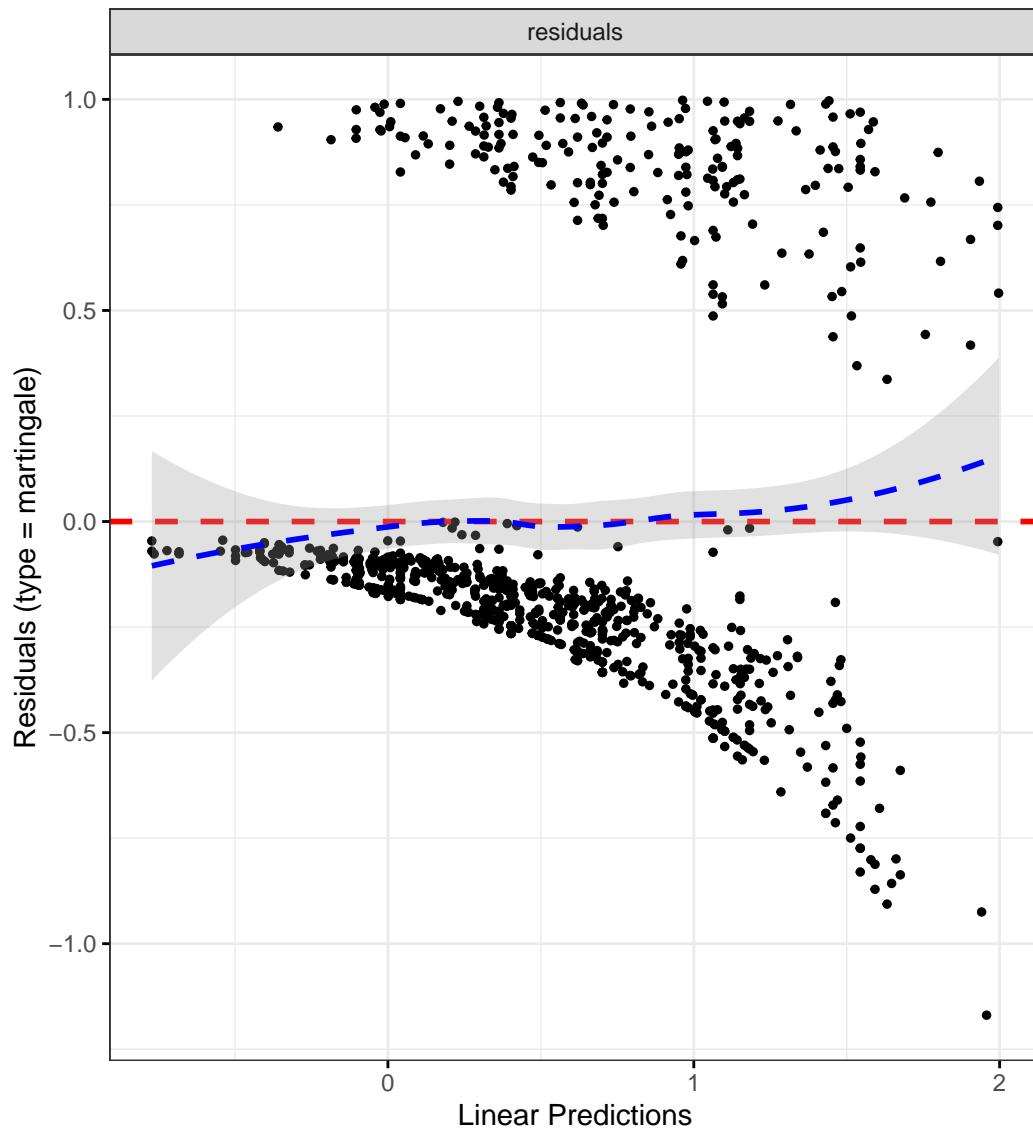
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



## Montreal Behaviour

### Patient-reported flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Behaviour",
  outcome_time = "softflare_time",
```

```

outcome_event = "softflare",
legend_title = "Montreal behaviour",
plot_base_path = "plots/cd/soft-flare/ibd/behaviour",
break_time_by = 200,
palette = c("#3DDC97", "#FFD166", "#DD7373")
)

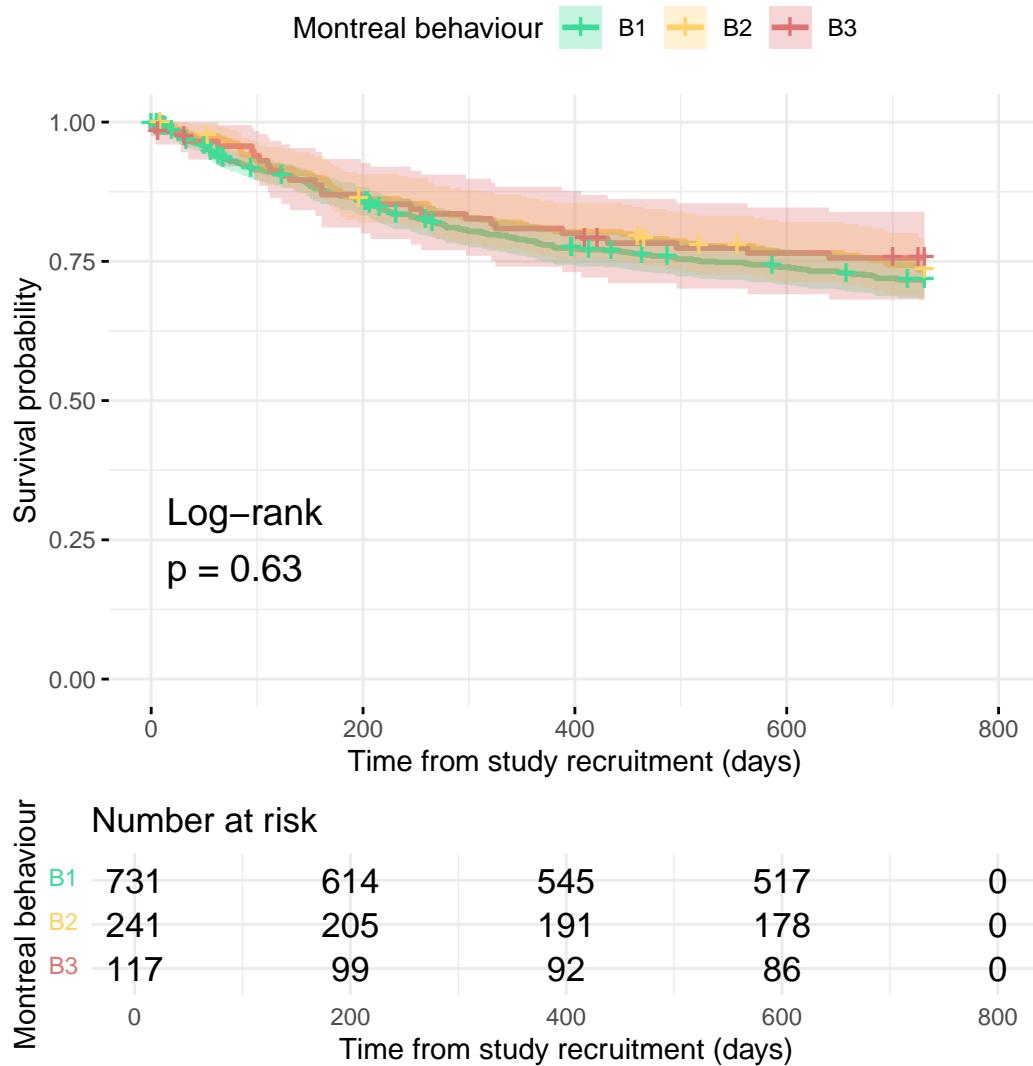
# Cox model
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + cat + IMD + Behaviour +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

cd.clin.forest <- rbind(
  cd.clin.forest,
  get_HR(fit.me, c("BehaviourB2", "BehaviourB3"))
)

# Display plot and model summary
analysis_result$plot

```

### Time to patient-reported flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	2.0640	1.5766	2.7020	0.0000
catFC 50-250	1.6340	1.2279	2.1746	0.0008
catFC > 250	2.3258	1.6953	3.1909	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	0.9230	0.5693	1.4965	0.7452
IMD3	0.7493	0.4518	1.2430	0.2637
IMD4	0.8865	0.5480	1.4342	0.6236
IMD5	0.9870	0.6239	1.5614	0.9554
BehaviourB2	1.0662	0.7953	1.4294	0.6683
BehaviourB3	0.7879	0.4947	1.2549	0.3155

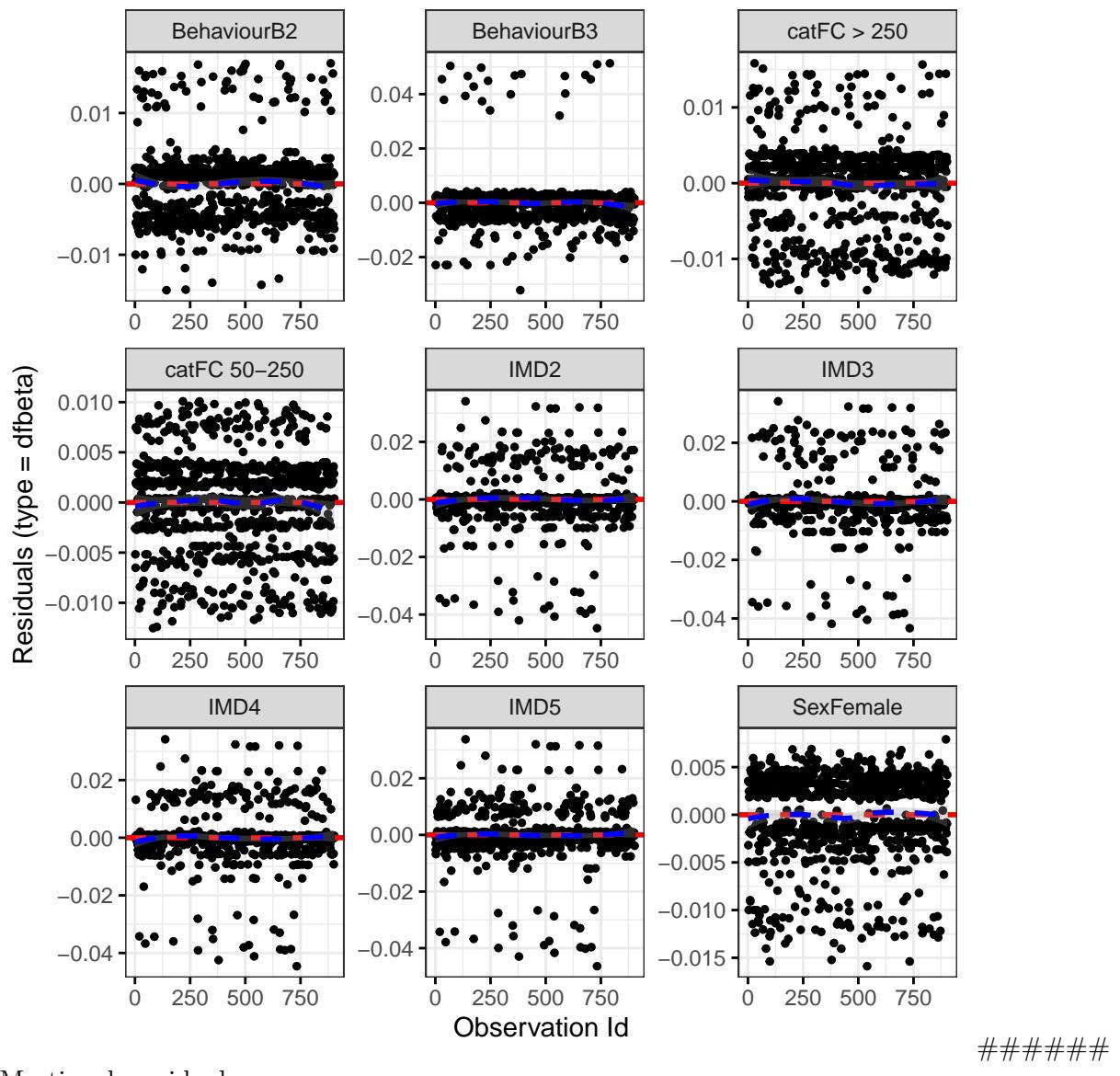
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.1946	0.9978	0.6582
cat	3.9115	1.9961	0.1410
IMD	4.3369	3.9877	0.3606
Behaviour	0.9019	1.9953	0.6359
GLOBAL	9.4596	10.0044	0.4895

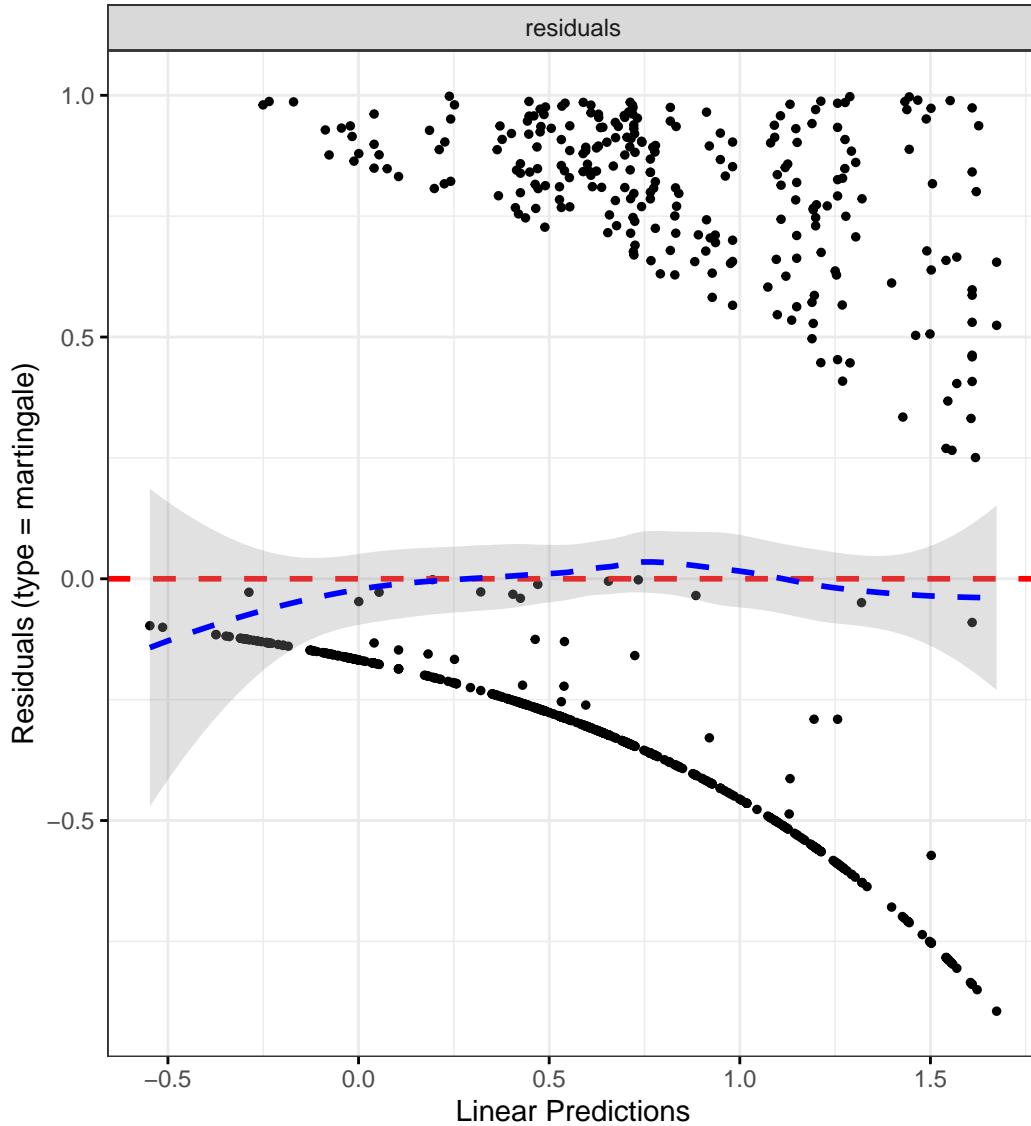
### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



## Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Behaviour",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "Montreal behaviour",
```

```

plot_base_path = "plots/cd/hard-flare/ibd/behaviour",
break_time_by = 200,
palette = c("#3DDC97", "#FFD166", "#DD7373")
)

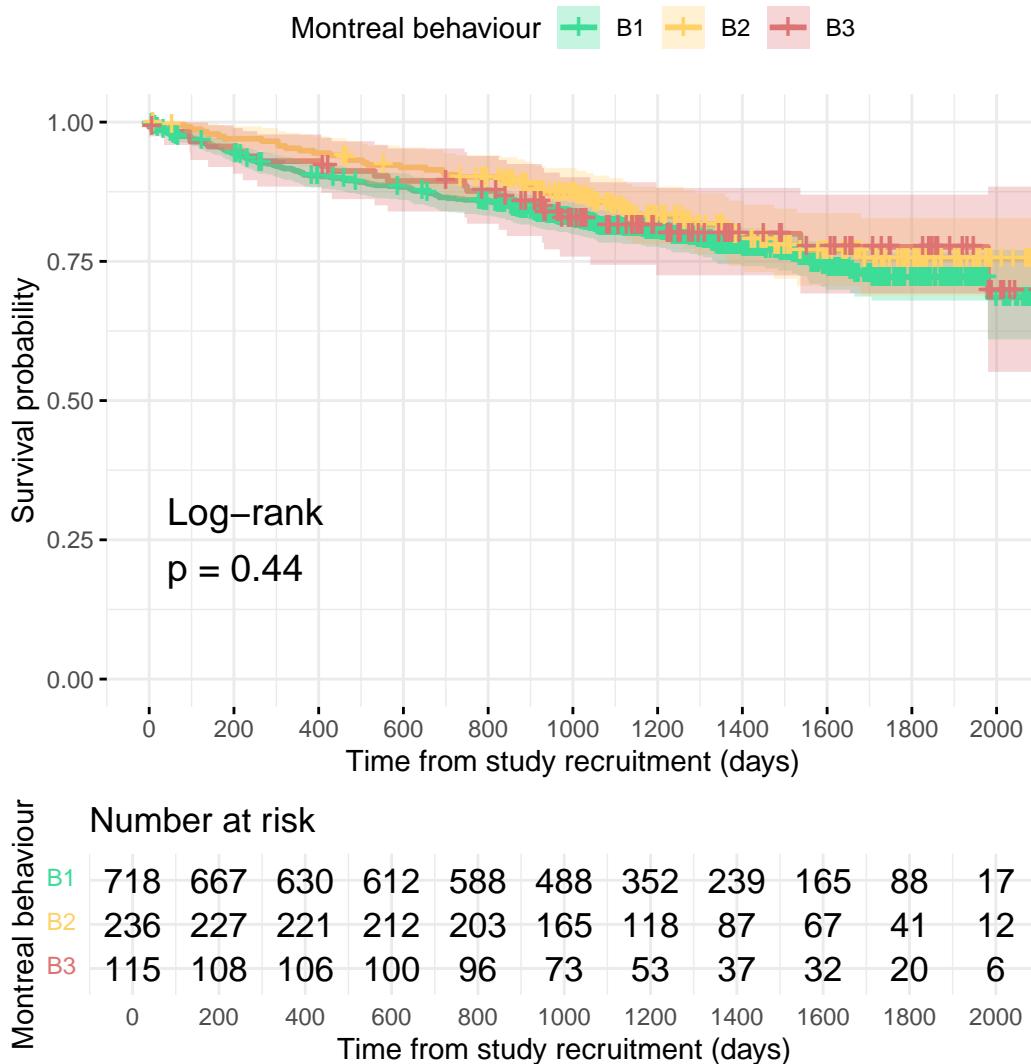
# Cox model
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + cat + IMD + Behaviour +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

cd.hard.forest <- rbind(
  cd.hard.forest,
  get_HR(fit.me, c("BehaviourB2", "BehaviourB3"))
)

# Display plot and model summary
analysis_result$plot

```

### Time to hard flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.5091	1.1174	2.0383	0.0073
catFC 50-250	2.0811	1.4681	2.9499	0.0000
catFC > 250	3.4647	2.3857	5.0316	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	0.9419	0.5228	1.6971	0.8421
IMD3	0.8970	0.4872	1.6518	0.7272
IMD4	0.9794	0.5448	1.7605	0.9444
IMD5	1.0317	0.5866	1.8145	0.9137
BehaviourB2	0.8913	0.6267	1.2677	0.5221
BehaviourB3	0.8505	0.5012	1.4431	0.5483

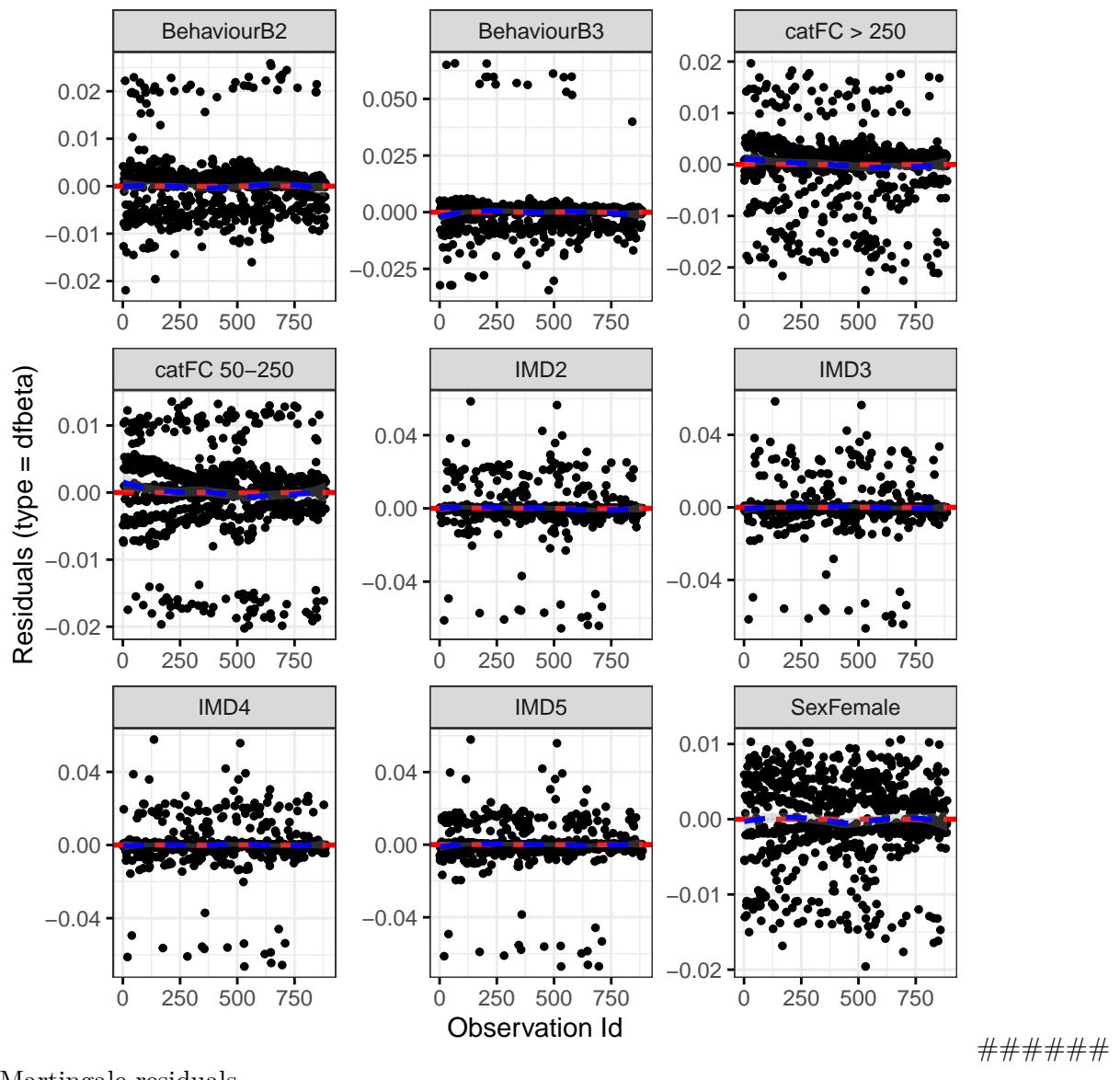
Diagnostics:

#### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.0070	0.9845	0.9303
cat	5.6993	1.9835	0.0570
IMD	2.8336	3.9427	0.5771
Behaviour	2.7982	1.9761	0.2427
GLOBAL	12.1681	20.1919	0.9154

#### DF betas

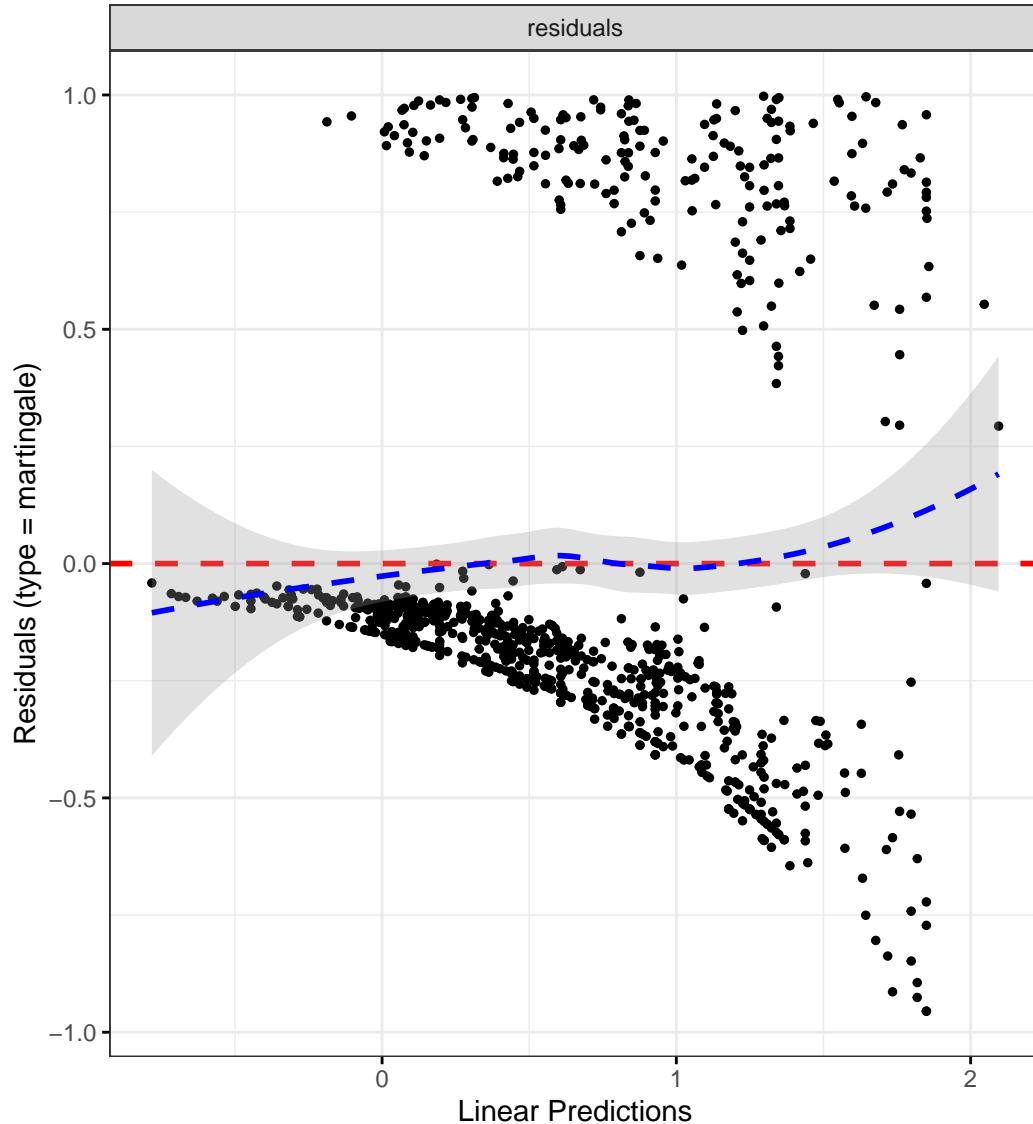
```
`geom_smooth()` using formula = 'y ~ x'
```



Martingale residuals

#####

```
`geom_smooth()` using formula = 'y ~ x'
```



### Perianal disease

#### Patient-reported flare

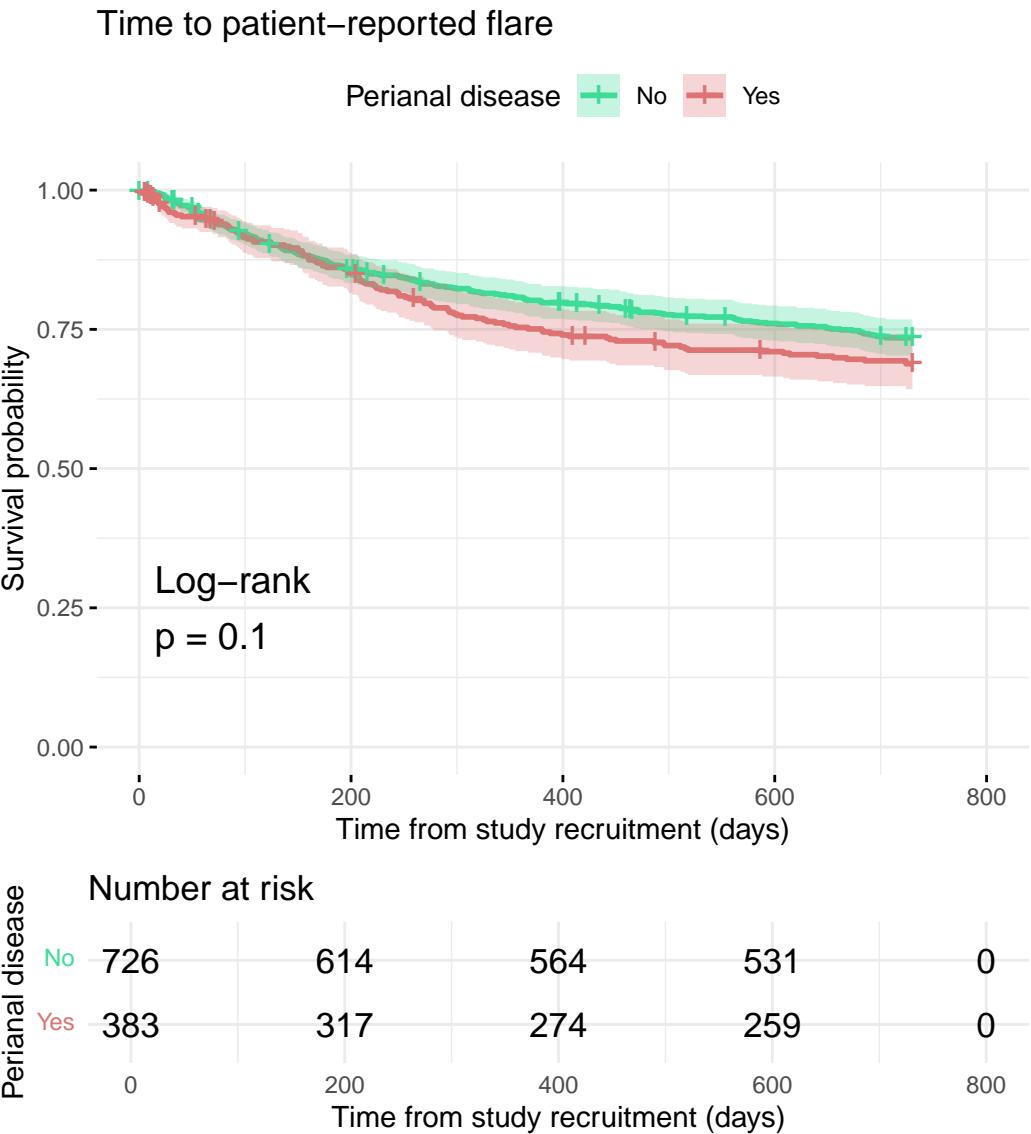
```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Perianal",
  outcome_time = "softflare_time",
```

```
outcome_event = "softflare",
legend_title = "Perianal disease",
plot_base_path = "plots/cd/soft-flare/ibd/perinal",
break_time_by = 200,
palette = c("#3DDC97", "#DD7373")
)

# Cox model
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + cat + IMD + Perianal +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

cd.clin.forest <- rbind(
  cd.clin.forest,
  get_HR(fit.me, "PerianalYes")
)

# Display plot and model summary
analysis_result$plot
```



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	2.0572	1.5795	2.6794	0.0000
catFC 50-250	1.6277	1.2282	2.1573	0.0007
catFC > 250	2.3910	1.7543	3.2587	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	0.9586	0.5934	1.5485	0.8628
IMD3	0.8177	0.4975	1.3441	0.4274
IMD4	0.8333	0.5158	1.3464	0.4564
IMD5	0.8992	0.5679	1.4237	0.6503
PerianalYes	1.3022	1.0174	1.6667	0.0360

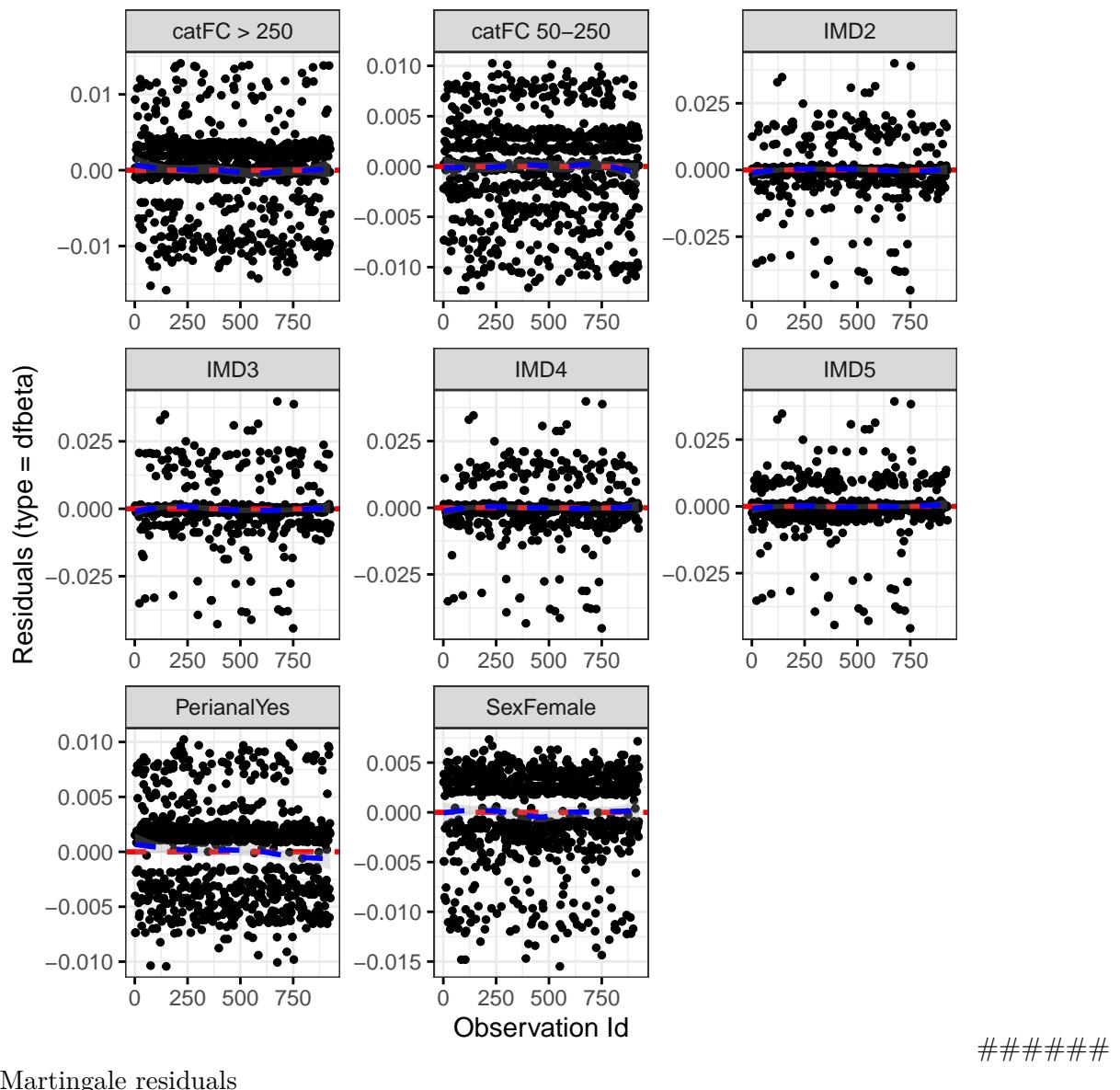
Diagnostics:

### Proportional hazards assumption test

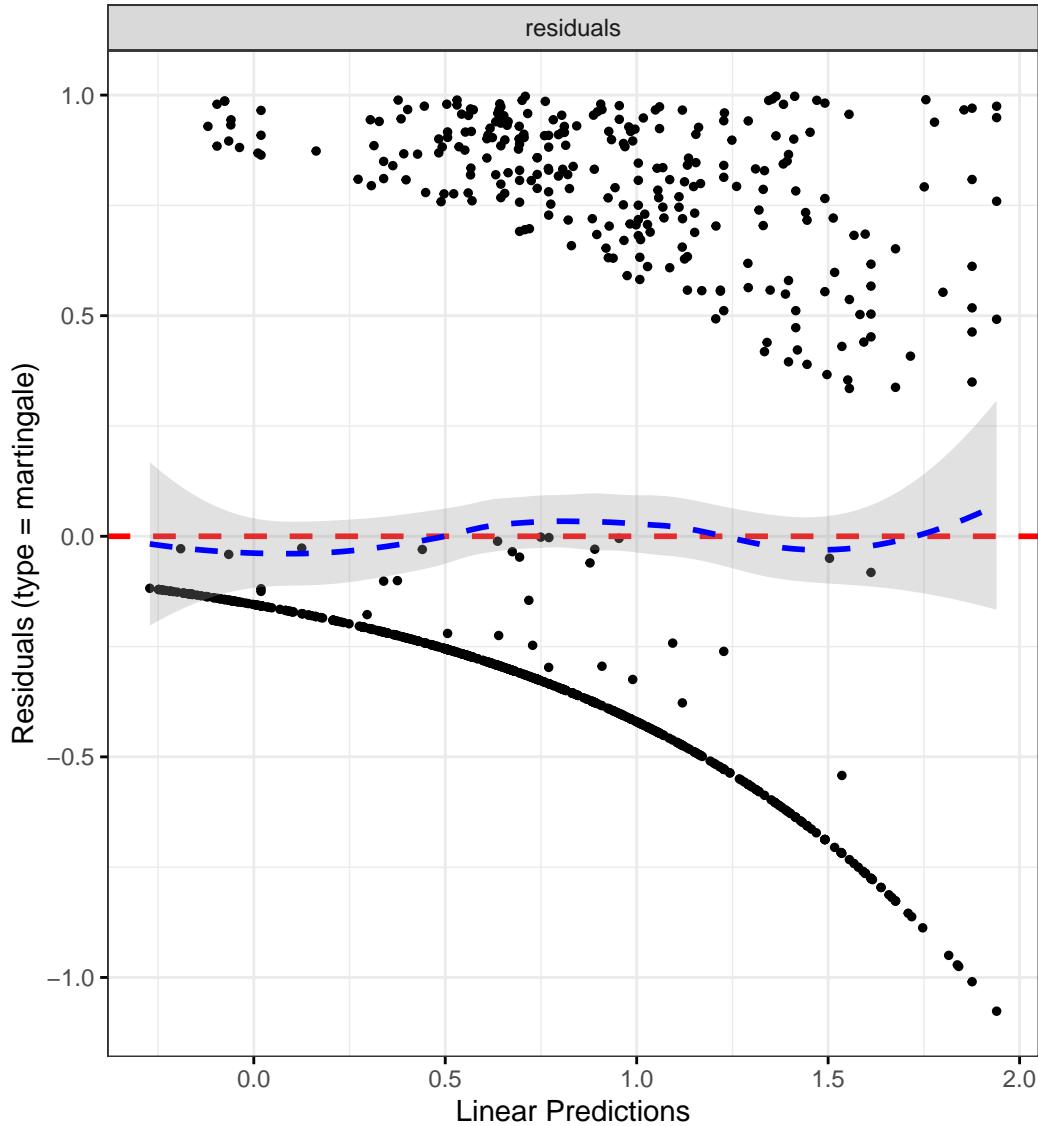
	Chi-squared statistic	DF	P-value
Sex	0.0684	0.9950	0.7920
cat	4.5984	1.9892	0.0994
IMD	5.5262	3.9719	0.2343
Perianal	0.2214	0.9919	0.6346
GLOBAL	10.2702	10.9257	0.4997

### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



```
`geom_smooth()` using formula = 'y ~ x'
```



## Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.cd.df,
  var_name = "Perianal",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "Perianal disease",
```

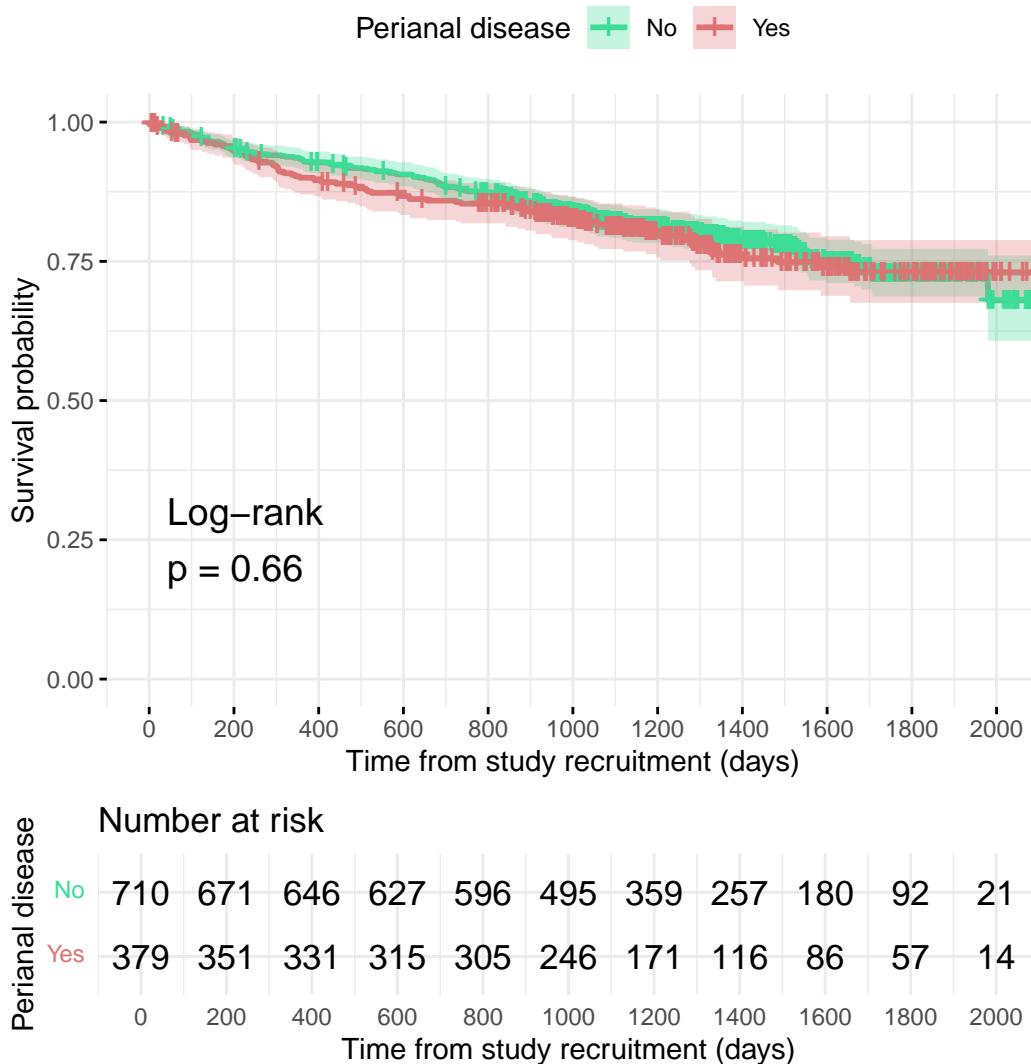
```
plot_base_path = "plots/cd/hard-flare/ibd/perinal",
break_time_by = 200,
palette = c("#3DDC97", "#DD7373")
)

# Cox model
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + cat + IMD + Perianal +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.cd.df
)

cd.hard.forest <- rbind(
  cd.hard.forest,
  get_HR(fit.me, "PerianalYes")
)

# Display plot and model summary
analysis_result$plot
```

### Time to hard flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.4036	1.0483	1.8795	0.0228
catFC 50-250	1.8951	1.3518	2.6569	0.0002
catFC > 250	3.1094	2.1650	4.4659	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	0.9969	0.5566	1.7855	0.9918
IMD3	1.0334	0.5690	1.8768	0.9140
IMD4	0.9261	0.5160	1.6621	0.7969
IMD5	0.9379	0.5339	1.6477	0.8235
PerianalYes	1.1428	0.8537	1.5299	0.3697

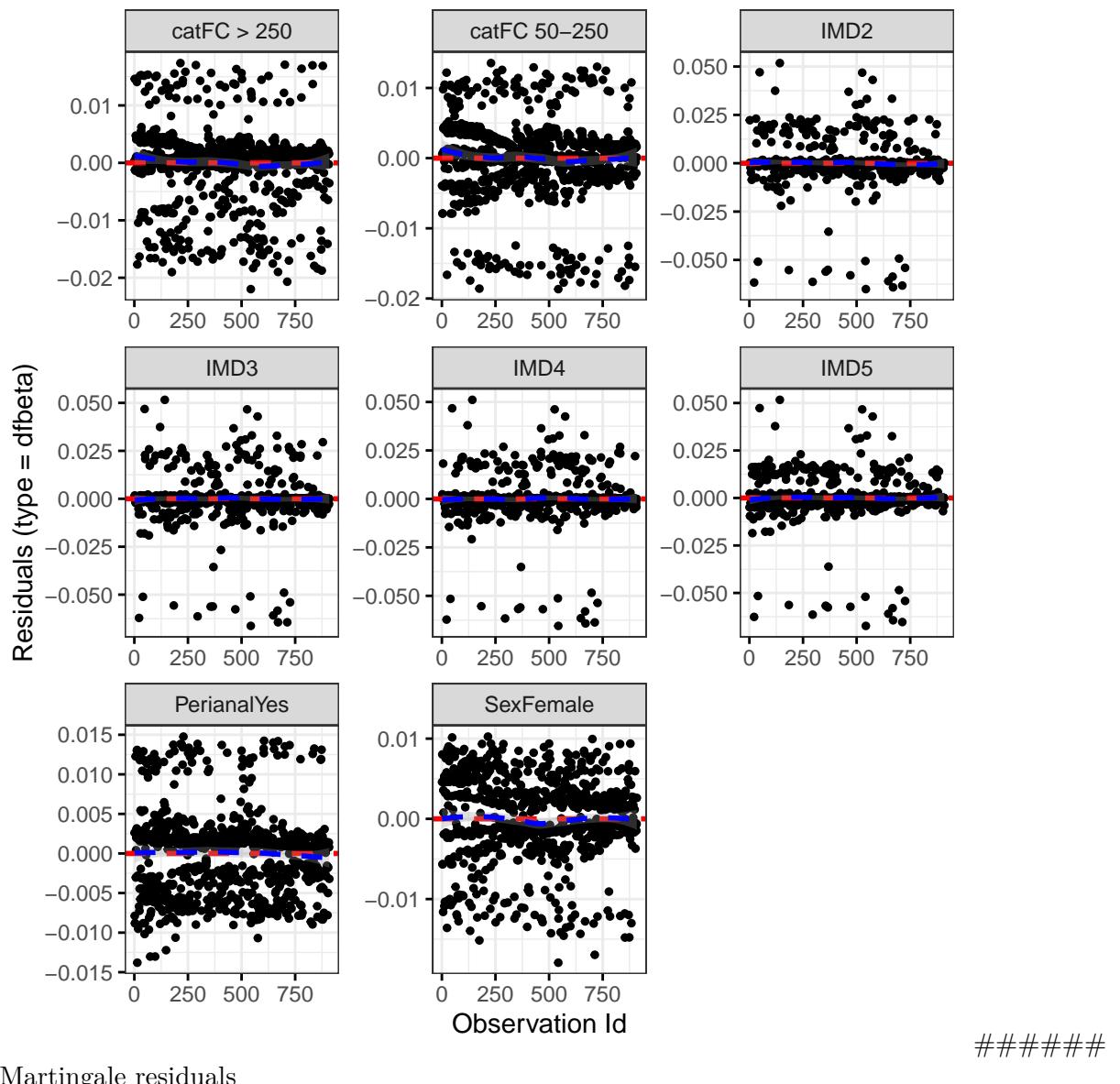
Diagnostics:

#### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.1937	0.9841	0.6533
cat	7.0942	1.9838	0.0283
IMD	2.3289	3.9479	0.6677
Perianal	1.9313	0.9841	0.1614
GLOBAL	11.5538	19.8863	0.9279

#### DF betas

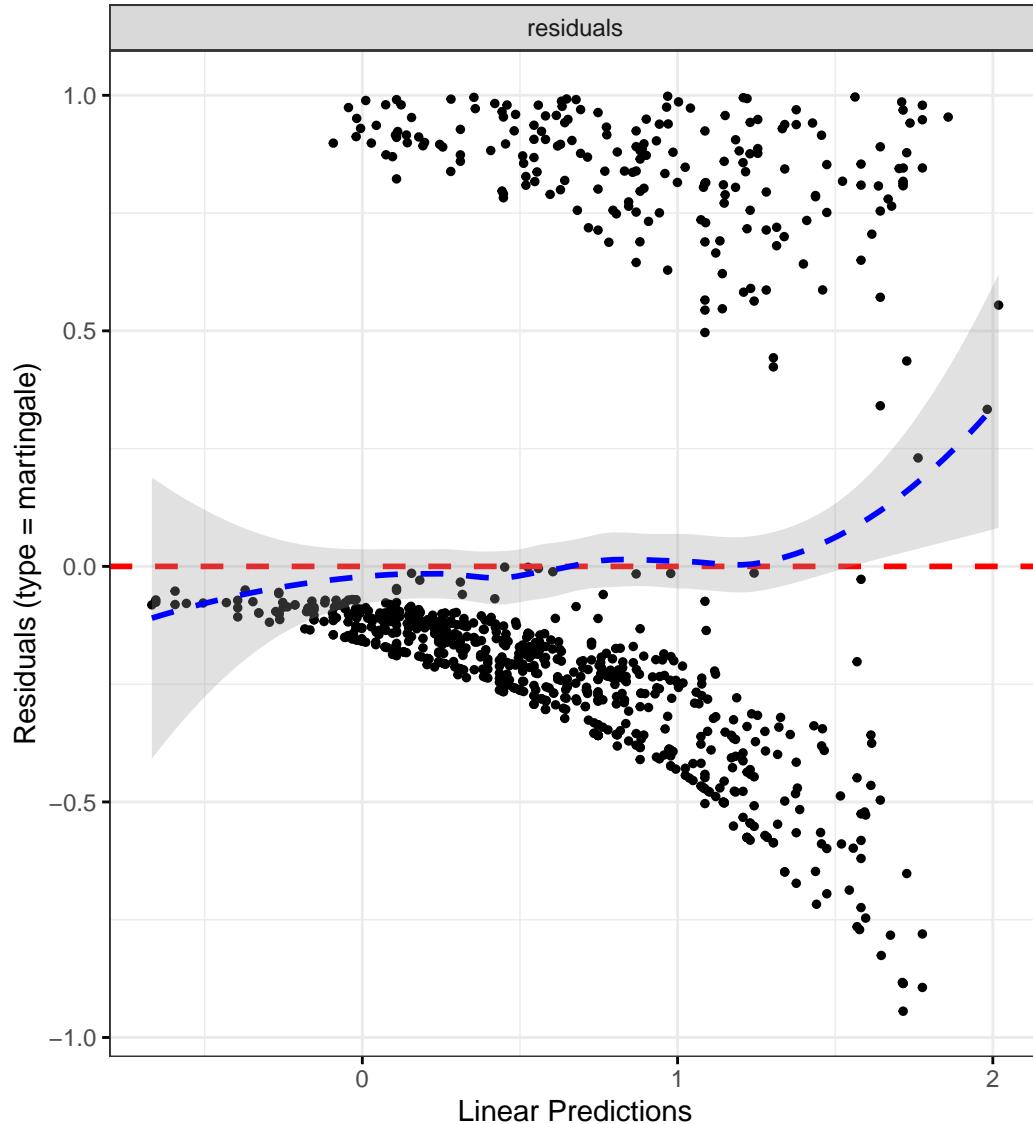
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



### Variables only relevant to UC/IBDU

```

demo.uc <- readRDS(paste0(paths$outdir, "demo-uc.RDS"))
demo.uc <- demo.uc[, c("ParticipantNo", "Extent", "Mayo")]
flare.uc.df <- merge(flare.uc.df,
  demo.uc,
  by = "ParticipantNo",
  all.x = TRUE,
  
```

```

    all.y = FALSE
)

# Create categorized version for survival analysis
flare.uc.df$Extent_cat <- flare.uc.df$Extent

```

## Montreal extent

### Patient-reported flare

```

# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.uc.df,
  var_name = "Extent",
  outcome_time = "softflare_time",
  outcome_event = "softflare",
  legend_title = "Montreal extent",
  plot_base_path = "plots/uc/soft-flare/ibd/extents",
  break_time_by = 200,
  palette = c("#3DDC97", "orange", "#DD7373")
)

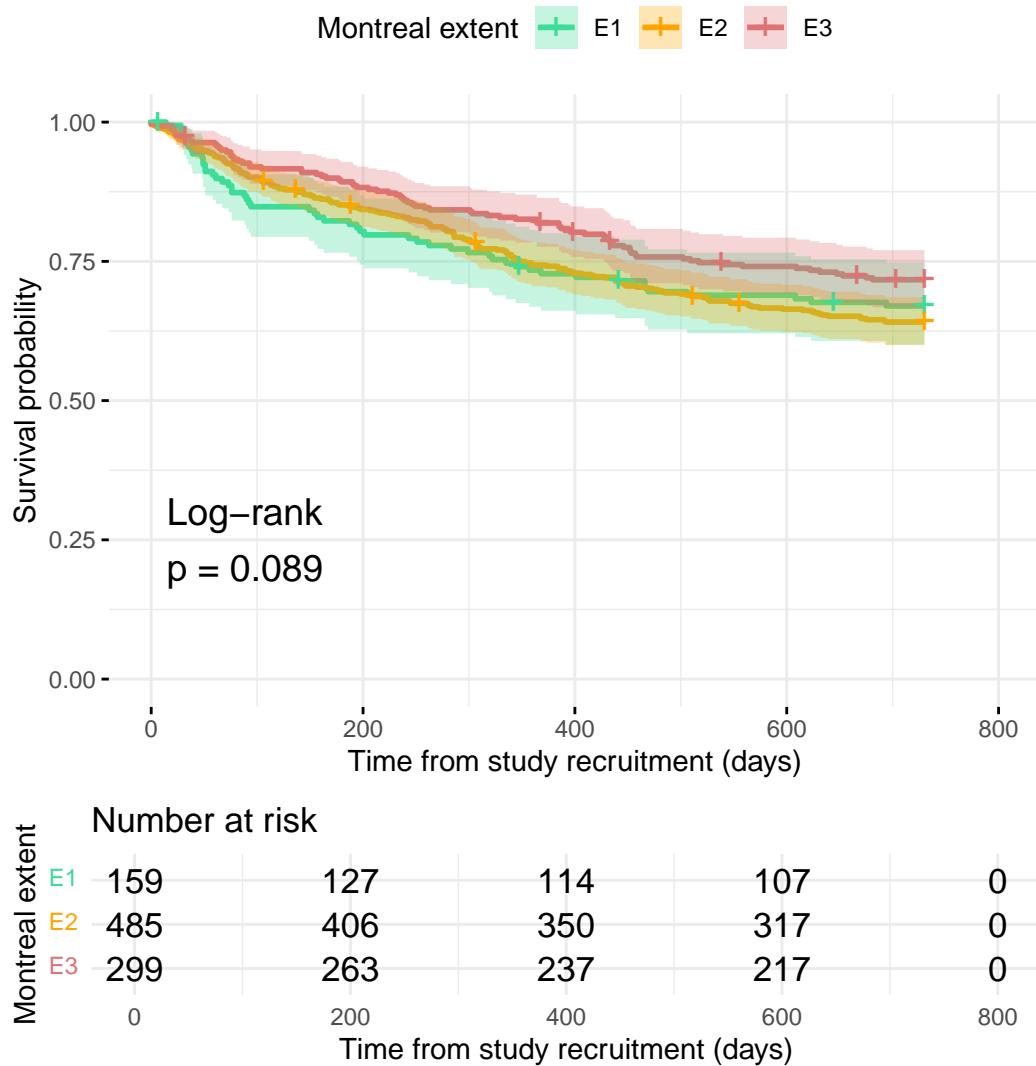
# Cox model
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + cat + IMD + Extent +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

uc.clin.forest <- rbind(
  uc.clin.forest,
  get_HR(fit.me, c("ExtentE2", "ExtentE3"))
)

# Display plot and model summary
analysis_result$plot

```

### Time to patient-reported flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.4261	1.1158	1.8227	0.0046
catFC 50-250	1.6346	1.2328	2.1673	0.0006
catFC > 250	2.2553	1.6710	3.0440	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	1.2814	0.7812	2.1019	0.3262
IMD3	1.0697	0.6587	1.7371	0.7854
IMD4	1.4373	0.9127	2.2636	0.1175
IMD5	1.1749	0.7483	1.8449	0.4838
ExtentE2	0.9196	0.6607	1.2801	0.6196
ExtentE3	0.6967	0.4810	1.0090	0.0558

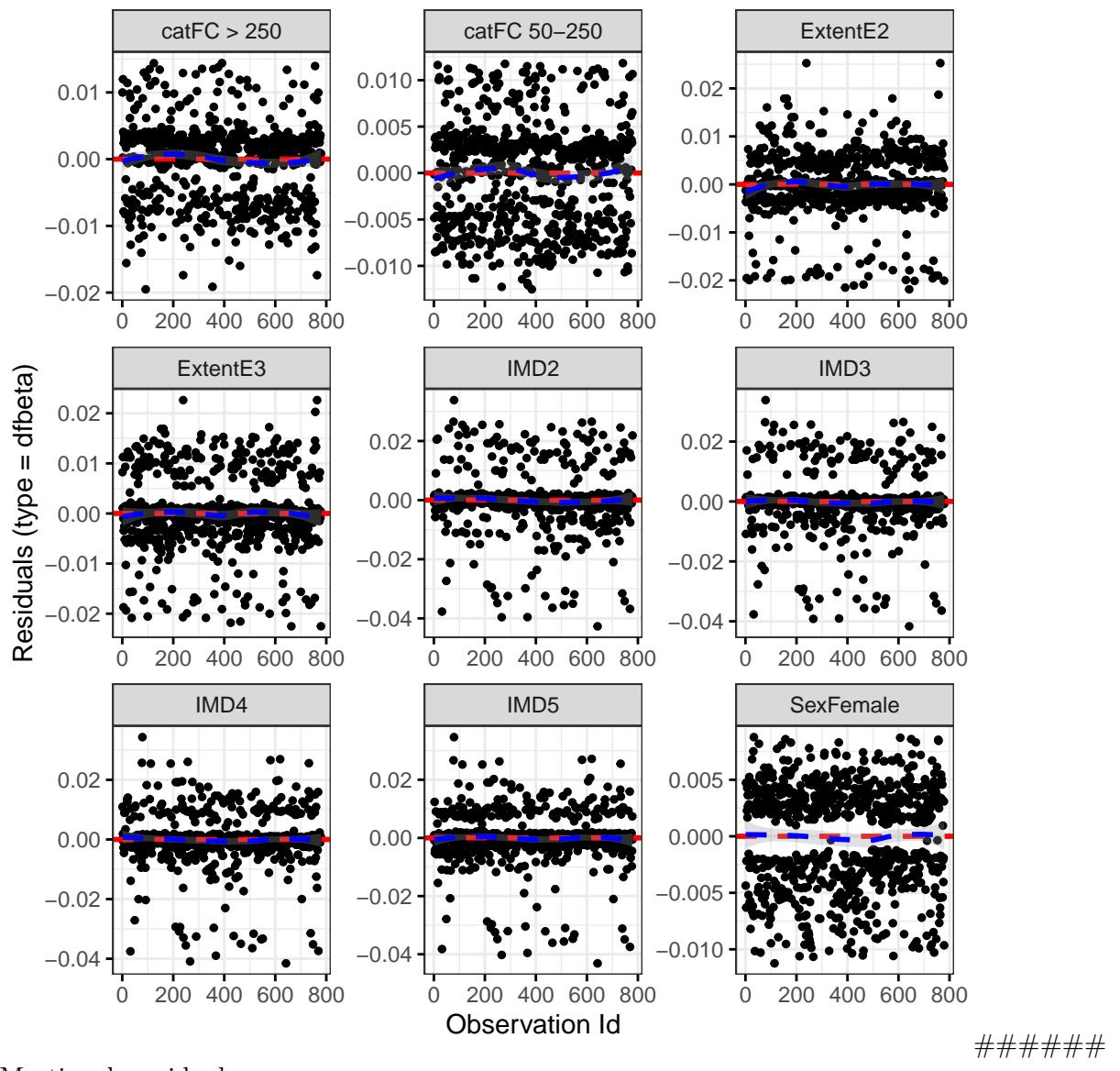
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.0215	0.9891	0.8804
cat	4.6159	1.9587	0.0959
IMD	2.9011	3.9332	0.5640
Extent	4.1788	1.9657	0.1202
GLOBAL	12.0732	21.9250	0.9547

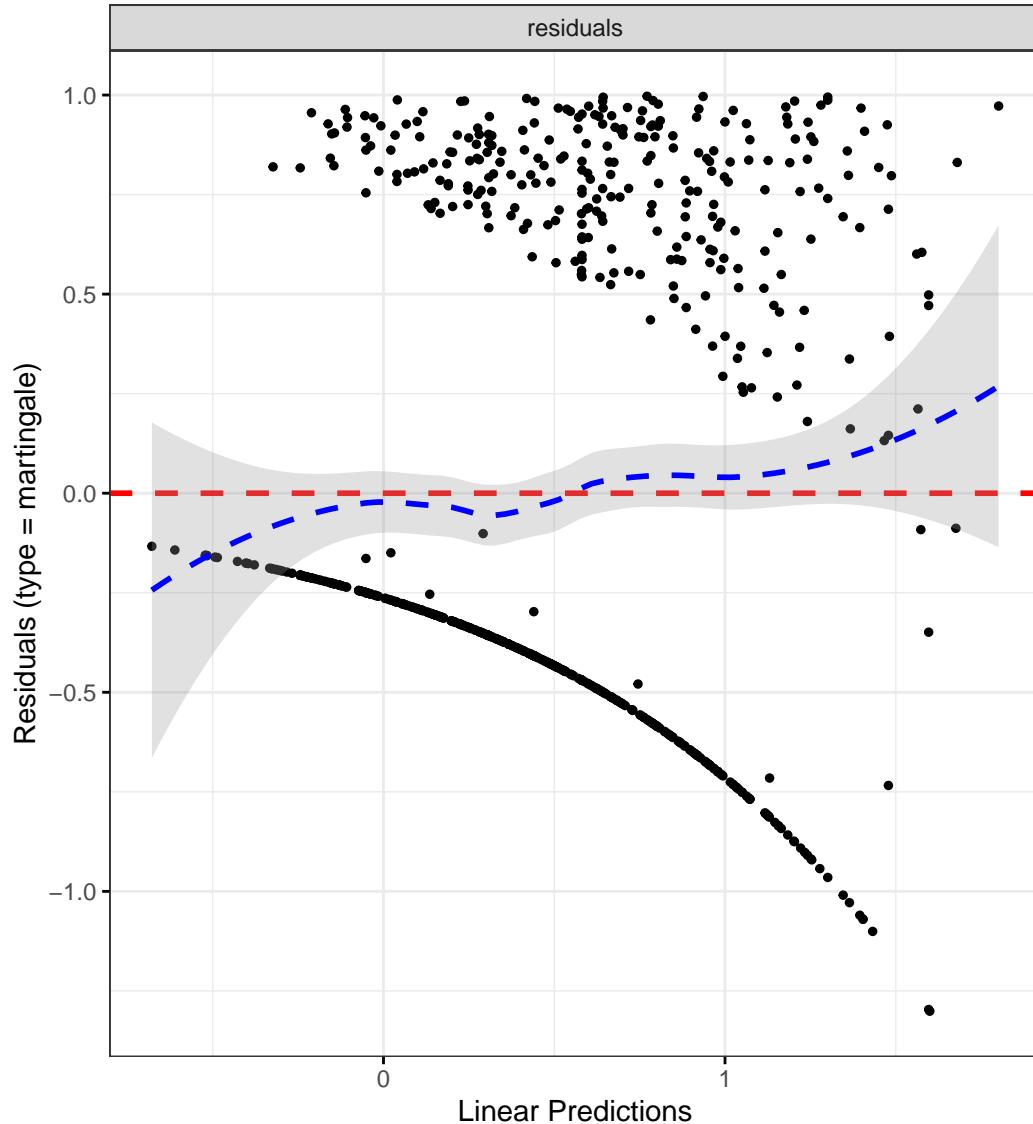
### DF betas

```
`geom_smooth()` using formula = 'y ~ x'
```



Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



## Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.uc.df,
  var_name = "Extent",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "Montreal extent",
```

```

plot_base_path = "plots/uc/hard-flare/ibd/extents",
break_time_by = 200,
palette = c("#3DDC97", "orange", "#DD7373")
)

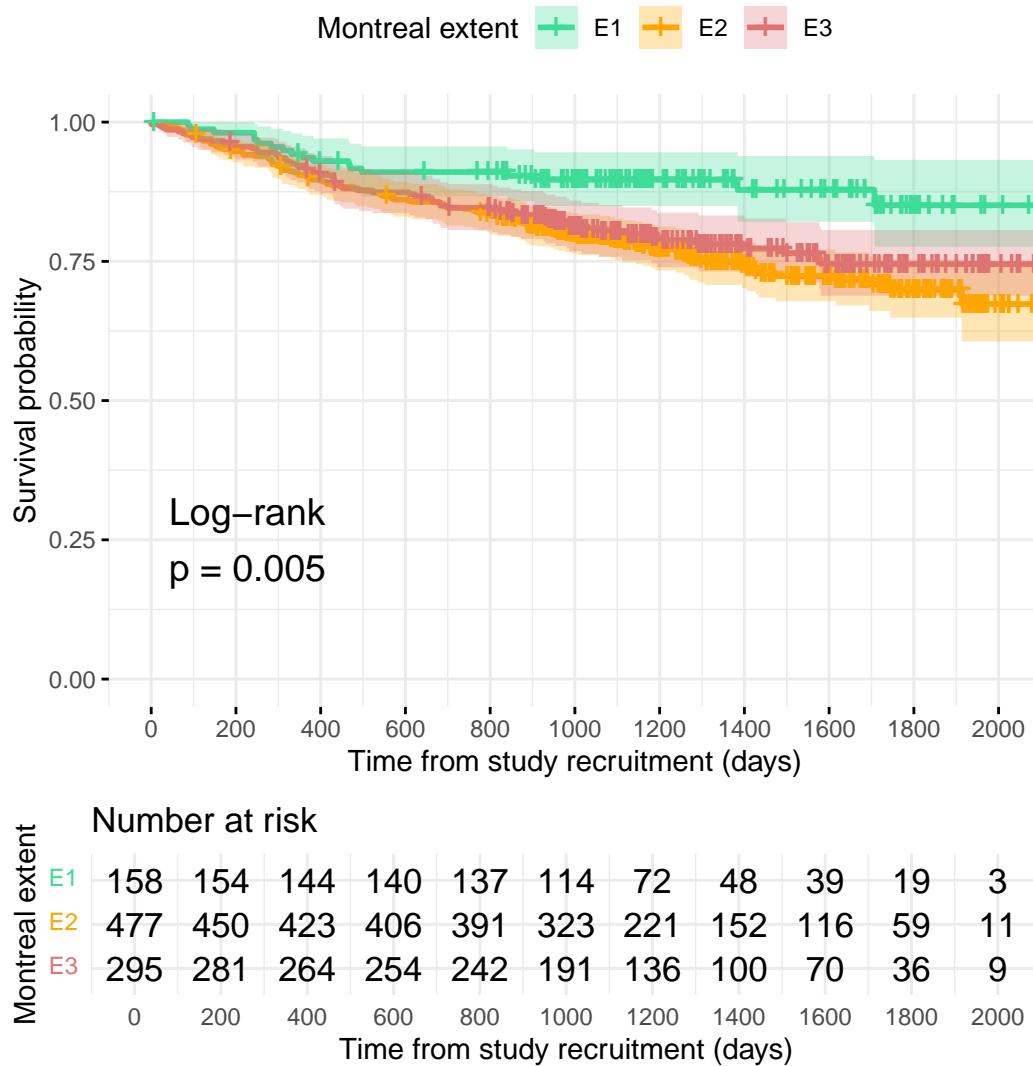
# Cox model
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + cat + IMD + Extent +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

uc.hard.forest <- rbind(
  uc.hard.forest,
  get_HR(fit.me, c("ExtentE2", "ExtentE3"))
)

# Display plot and model summary
analysis_result$plot

```

### Time to hard flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.4242	1.0492	1.9334	0.0234
catFC 50-250	1.9574	1.3666	2.8037	0.0002
catFC > 250	3.2458	2.2460	4.6907	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	1.6852	0.8594	3.3047	0.1288
IMD3	1.6813	0.8824	3.2033	0.1142
IMD4	2.2154	1.1918	4.1182	0.0119
IMD5	1.6230	0.8722	3.0201	0.1264
ExtentE2	2.1493	1.2632	3.6570	0.0048
ExtentE3	1.8507	1.0529	3.2532	0.0324

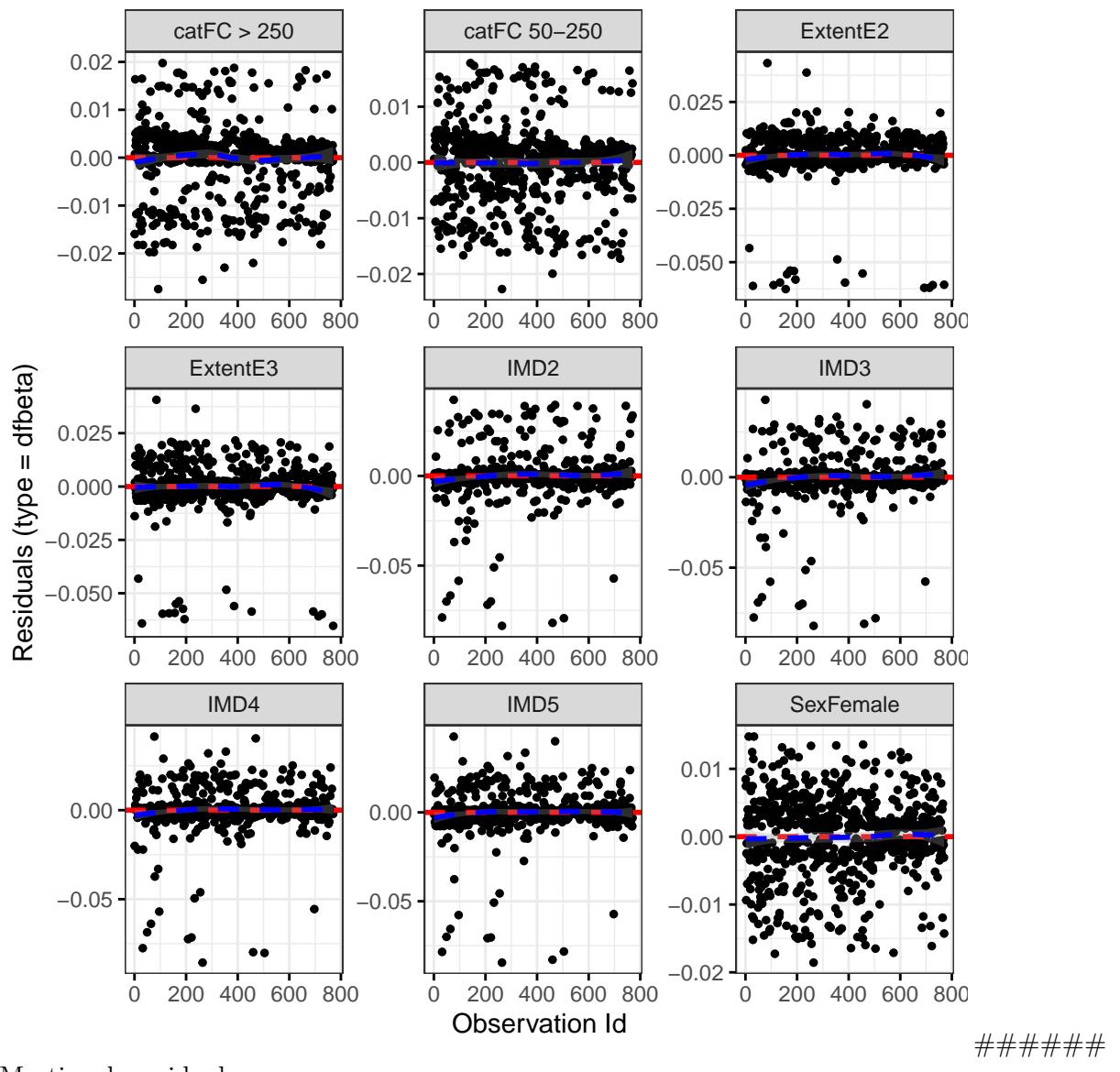
Diagnostics:

#### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.6040	0.9857	0.4315
cat	5.9334	1.9616	0.0496
IMD	2.5189	3.9364	0.6315
Extent	0.7654	1.9746	0.6760
GLOBAL	9.5946	21.8882	0.9891

#### DF betas

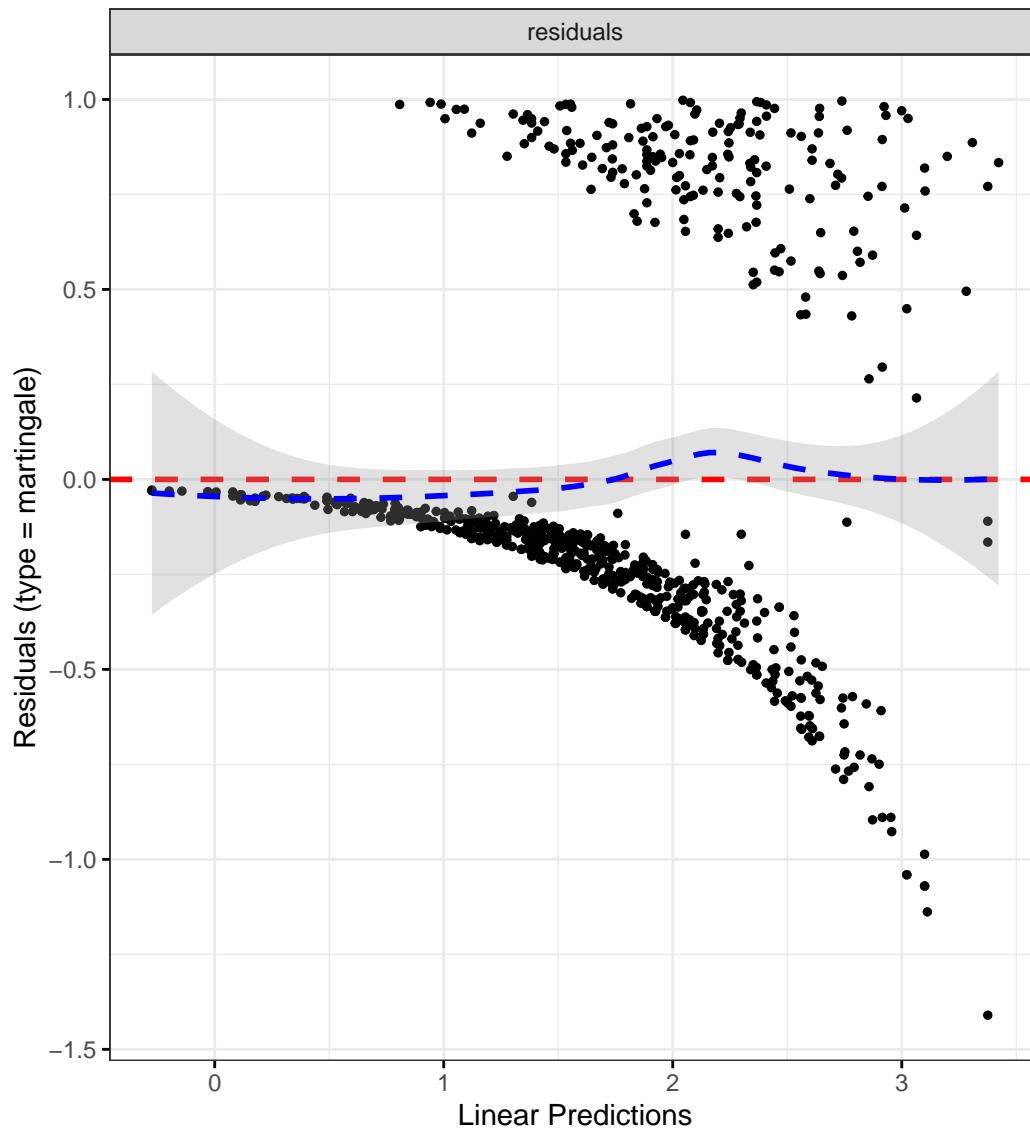
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



### Mayo score

```
flare.uc.df$Mayo_cat <- cut(flare.uc.df$Mayo,
  c(0, 2, 5, 9),
  labels = c("Inactive", "Mild", "Moderate"),
  right = TRUE
)
```

## Patient-reported flare

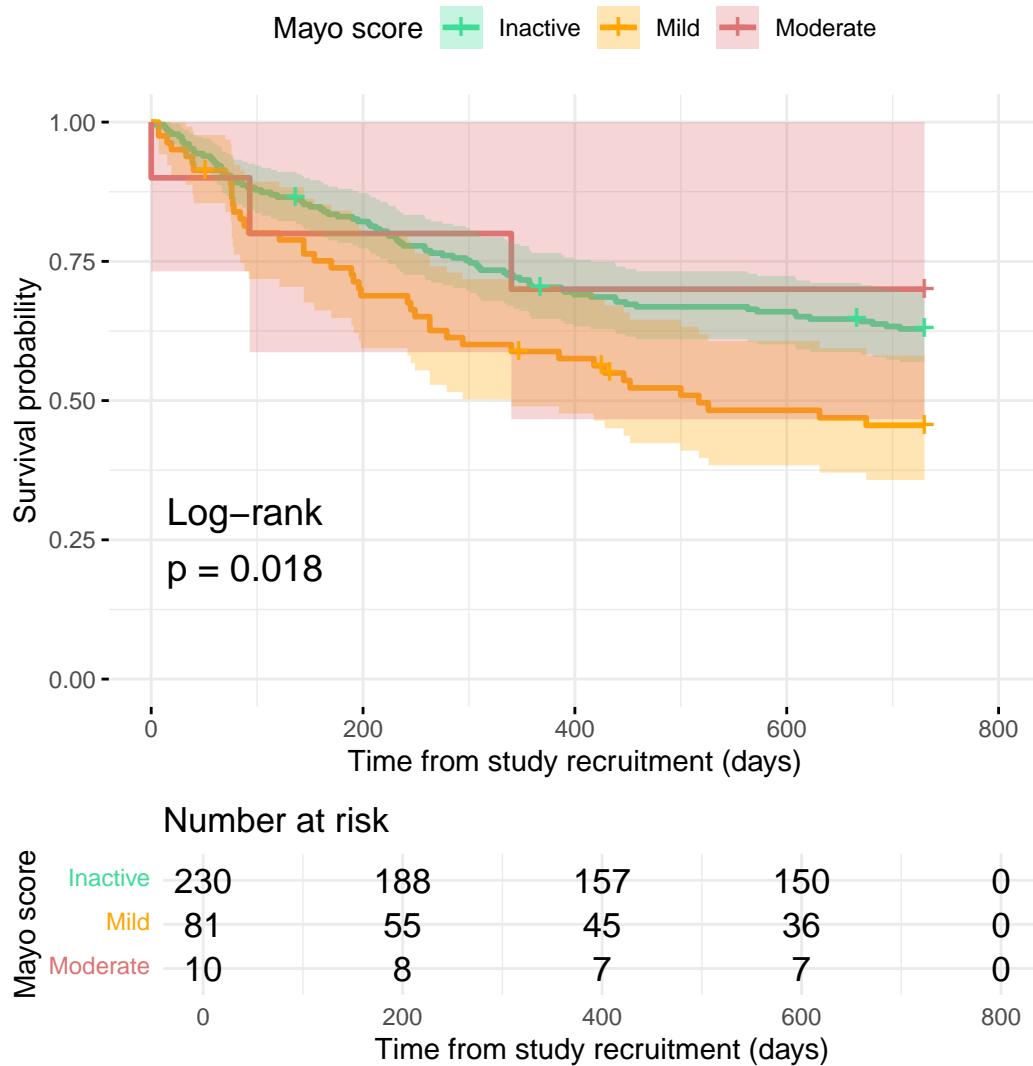
```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.uc.df,
  var_name = "Mayo",
  outcome_time = "softflare_time",
  outcome_event = "softflare",
  legend_title = "Mayo score",
  plot_base_path = "plots/uc/soft-flare/ibd/mayo",
  break_time_by = 200,
  palette = c("#3DDC97", "orange", "#DD7373")
)

# Cox model using continuous Mayo variable
fit.me <- coxph(
  Surv(softflare_time, softflare) ~
    Sex + cat + Mayo + Extent +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

uc.clin.forest <- rbind(
  uc.clin.forest,
  get_HR(fit.me, "Mayo")
)

# Display plot and model summary
analysis_result$plot
```

### Time to patient-reported flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.4049	1.0761	1.8343	0.0125
catFC 50-250	1.6646	1.2283	2.2557	0.0010
catFC > 250	2.4320	1.7451	3.3892	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
Mayo	1.1117	1.0094	1.2243	0.0316
ExtentE2	1.0050	0.6957	1.4518	0.9789
ExtentE3	0.7477	0.4960	1.1274	0.1652

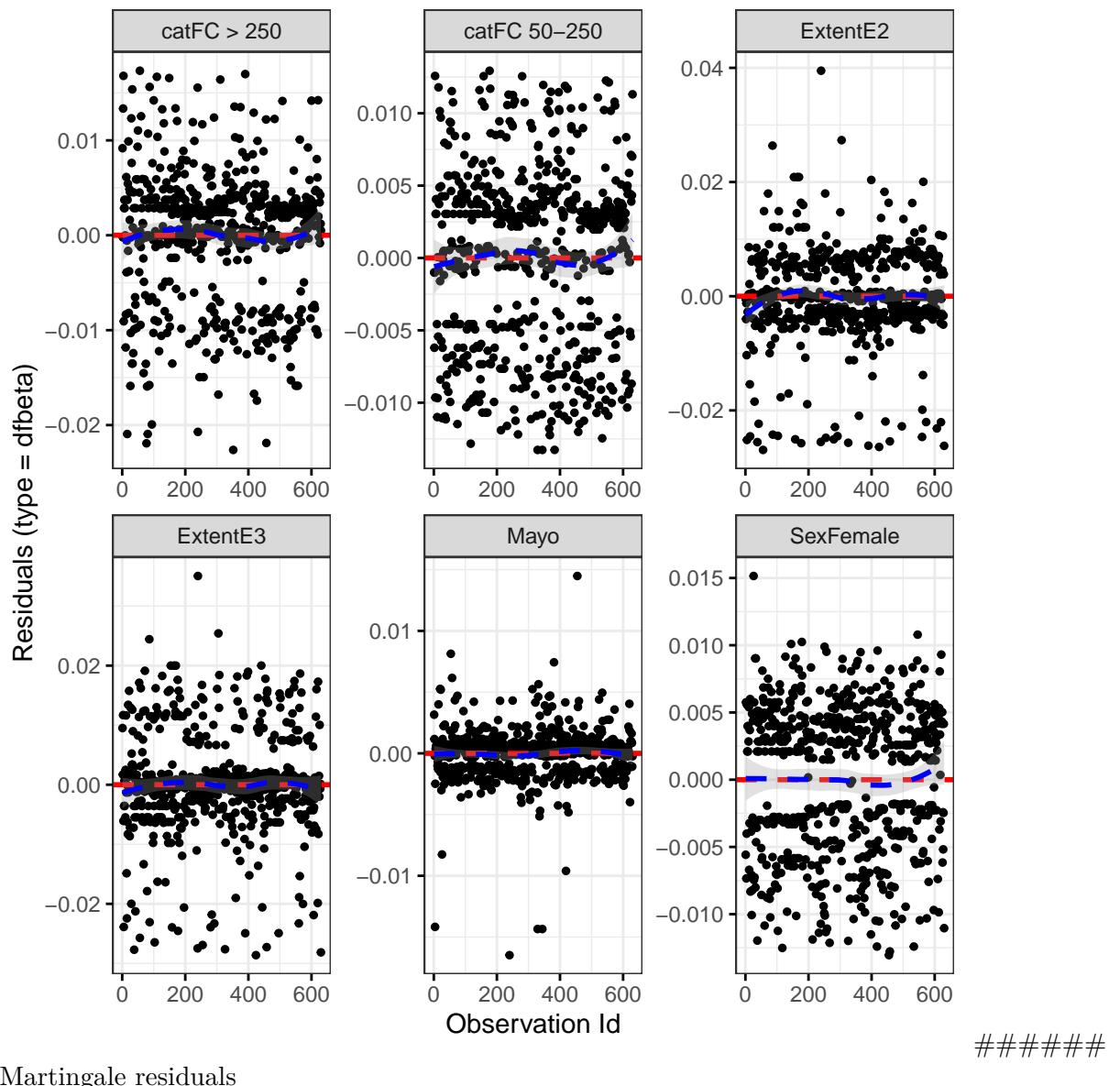
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.2221	0.9943	0.6351
cat	3.2655	1.9791	0.1924
Mayo	2.8510	0.9265	0.0823
Extent	4.4854	1.9674	0.1032
GLOBAL	10.5608	13.9227	0.7149

### DF betas

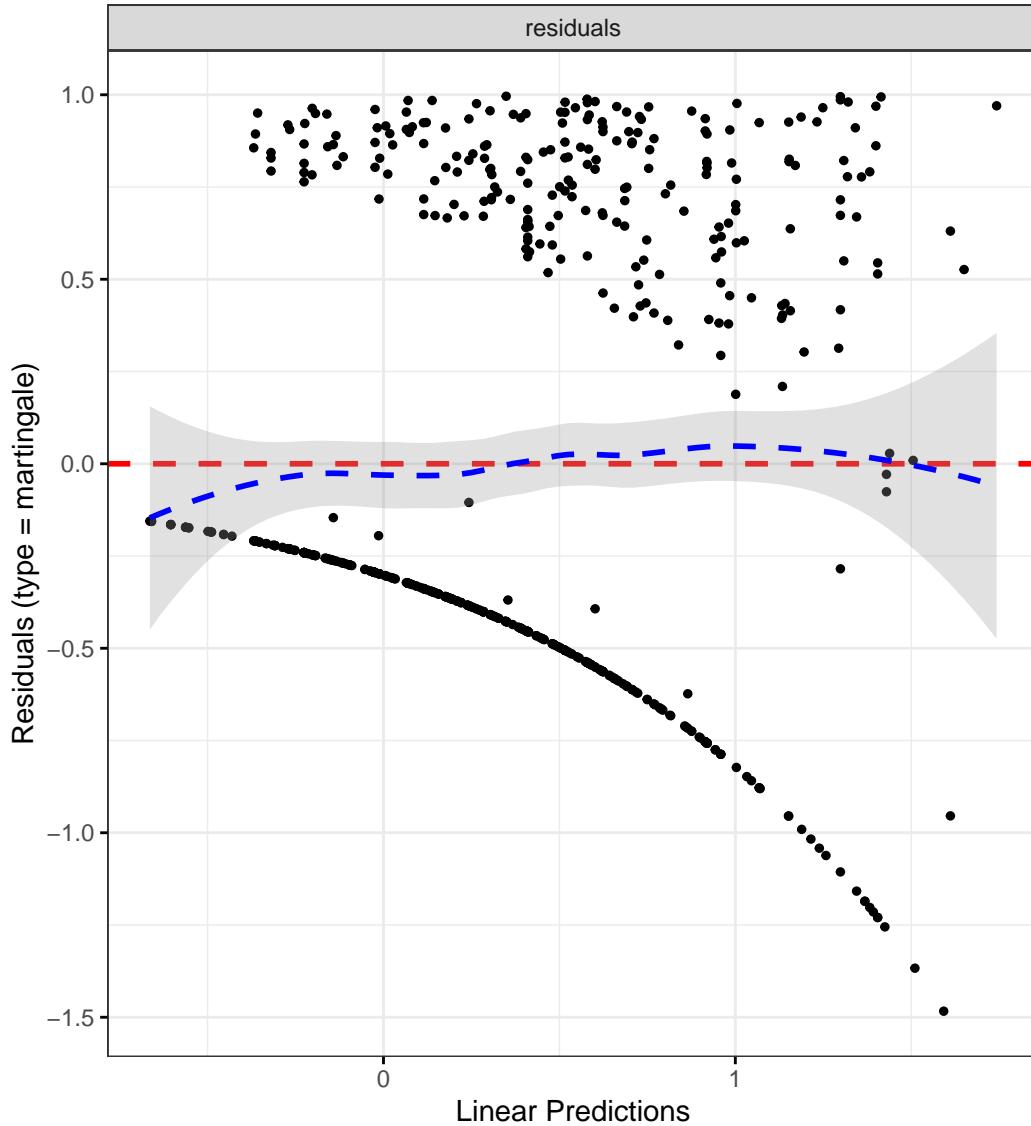
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



### Hard flare

```
# Run survival analysis using utility function
analysis_result <- run_survival_analysis(
  data = flare.uc.df,
  var_name = "Mayo",
  outcome_time = "hardflare_time",
  outcome_event = "hardflare",
  legend_title = "Mayo score",
```

```

plot_base_path = "plots/uc/hard-flare/ibd/mayo",
break_time_by = 200,
palette = c("#3DDC97", "orange", "#DD7373")
)

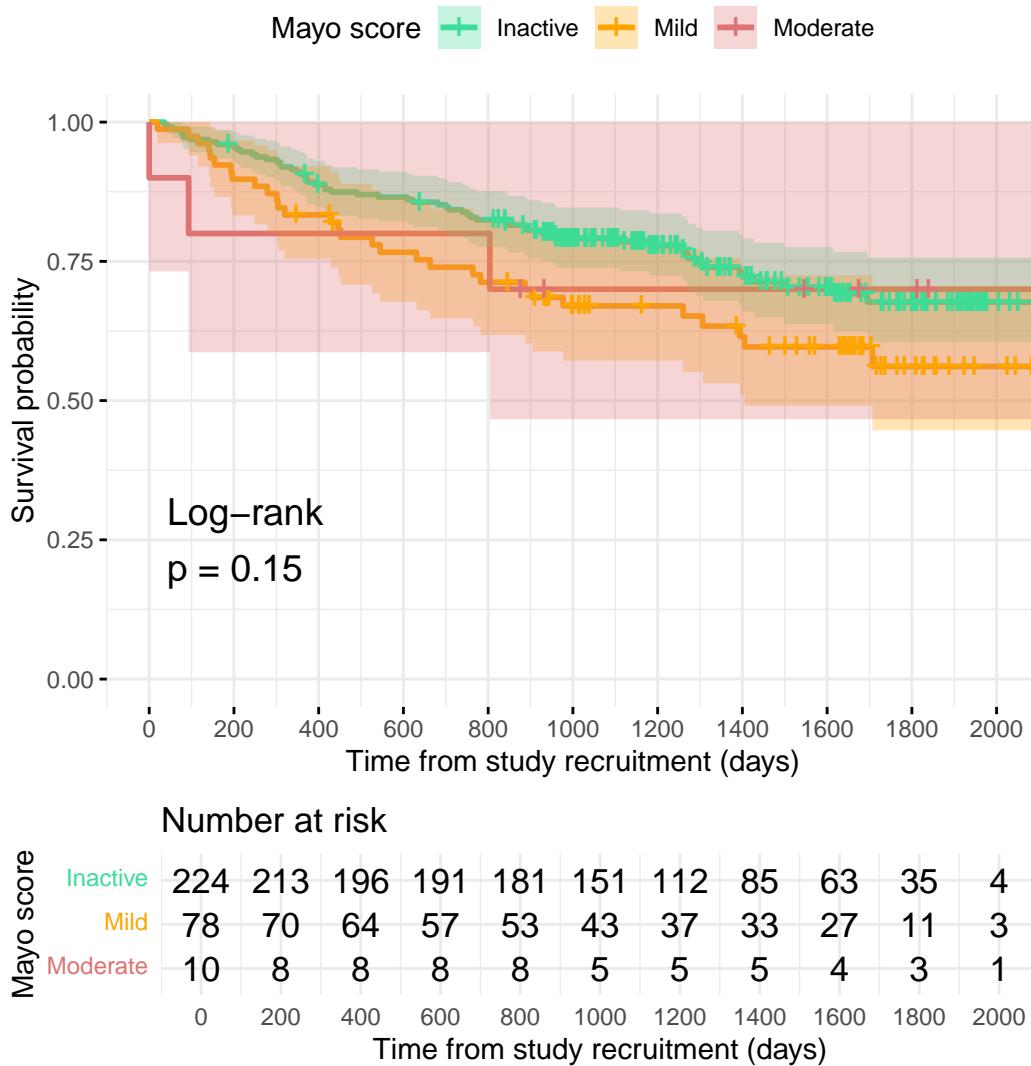
# Cox model using continuous Mayo variable
fit.me <- coxph(
  Surv(hardflare_time, hardflare) ~
    Sex + cat + IMD + Mayo +
    frailty(SiteNo),
  control = coxph.control(outer.max = 20),
  data = flare.uc.df
)

uc.hard.forest <- rbind(
  uc.hard.forest,
  get_HR(fit.me, "Mayo")
)

# Display plot and model summary
analysis_result$plot

```

### Time to hard flare



```
invisible(cox_summary(fit.me))
```

Cox model summary:

Variable	HR	Lower 95%	Upper 95%	P-value
SexFemale	1.2672	0.9438	1.7015	0.1152
catFC 50-250	1.8912	1.3331	2.6830	0.0004
catFC > 250	3.1128	2.1355	4.5373	0.0000

Variable	HR	Lower 95%	Upper 95%	P-value
IMD2	1.6242	0.8553	3.0844	0.1383
IMD3	1.1995	0.6373	2.2574	0.5729
IMD4	1.5787	0.8645	2.8828	0.1373
IMD5	1.2277	0.6754	2.2314	0.5011
Mayo	1.0890	0.9753	1.2161	0.1297

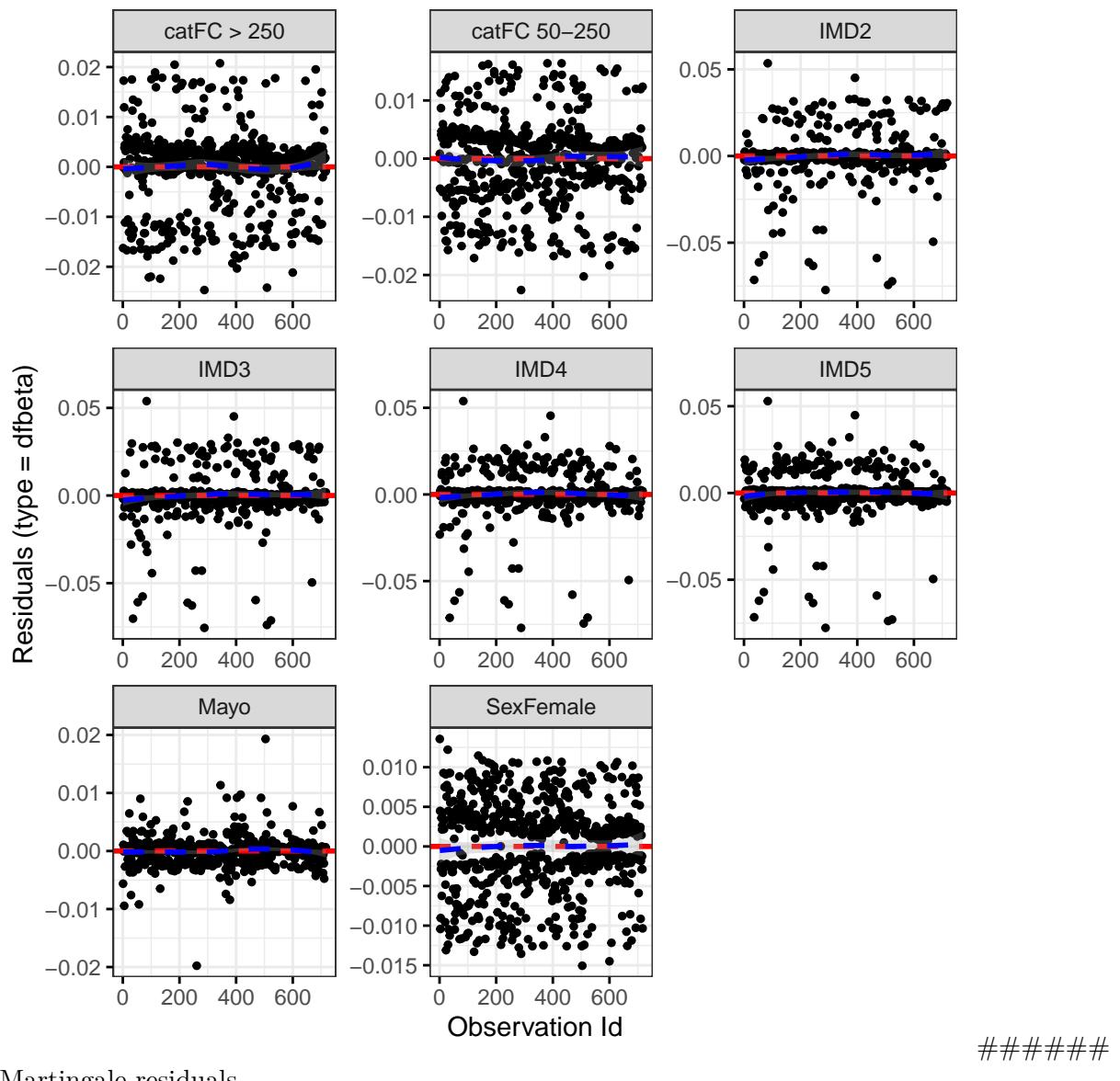
Diagnostics:

### Proportional hazards assumption test

	Chi-squared statistic	DF	P-value
Sex	0.0043	0.9878	0.9457
cat	3.1681	1.9834	0.2026
IMD	1.4273	3.9518	0.8342
Mayo	0.1658	0.9504	0.6632
GLOBAL	4.6901	17.1518	0.9987

### DF betas

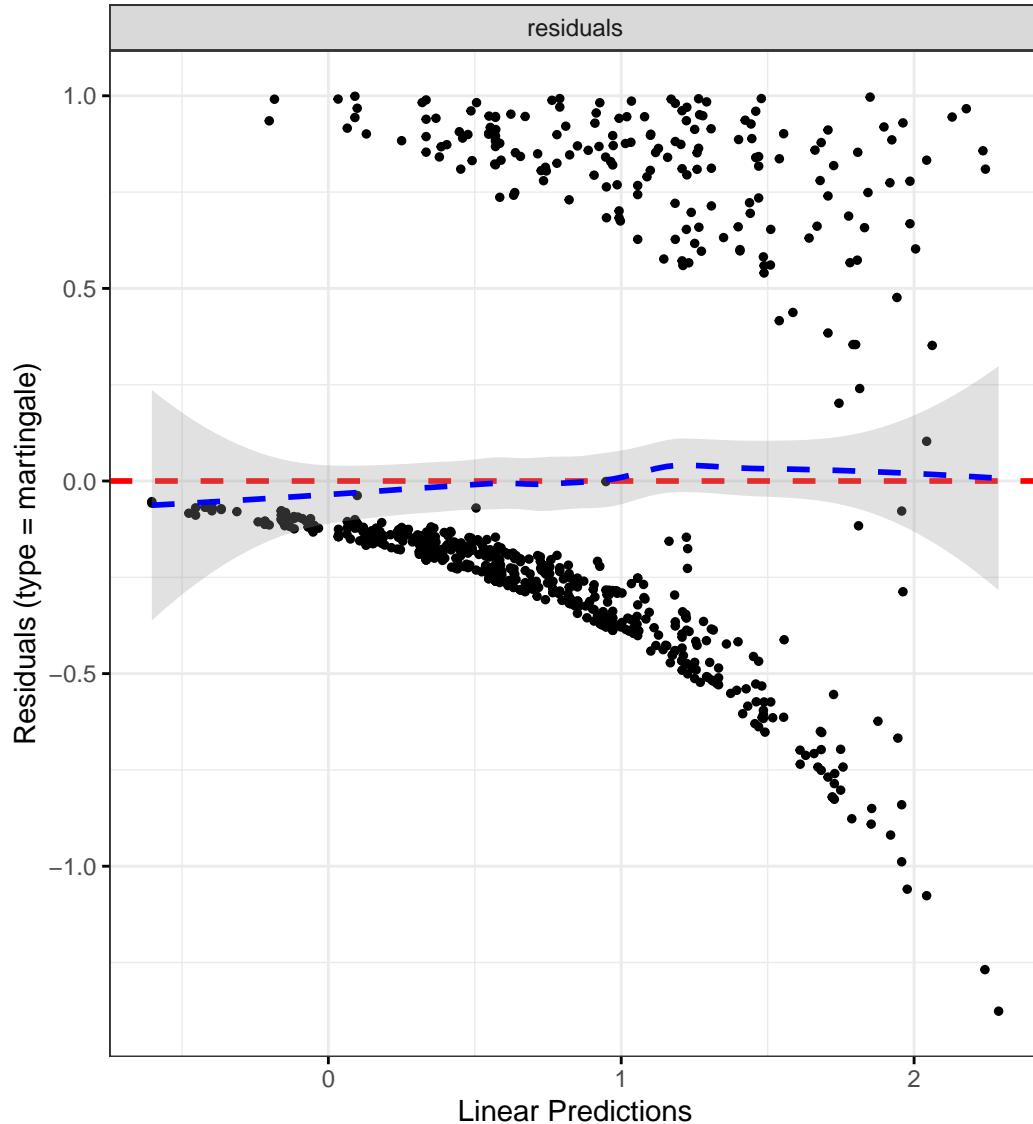
```
`geom_smooth()` using formula = 'y ~ x'
```



#####

Martingale residuals

```
`geom_smooth()` using formula = 'y ~ x'
```



```

saveRDS(flare.df, paste0(paths$outdir, "flares-IBD.RDS"))
saveRDS(flare.cd.df, paste0(paths$outdir, "flares-IBD-cd.RDS"))
saveRDS(flare.uc.df, paste0(paths$outdir, "flares-IBD-uc.RDS"))

saveRDS(cd.clin.forest, paste0(paths$outdir, "cd-clin-IBD.RDS"))
saveRDS(cd.hard.forest, paste0(paths$outdir, "cd-hard-IBD.RDS"))
saveRDS(uc.clin.forest, paste0(paths$outdir, "uc-clin-IBD.RDS"))
saveRDS(uc.hard.forest, paste0(paths$outdir, "uc-hard-IBD.RDS"))

```

## Reproduction and reproducibility

Session info

**R version 4.4.0 (2024-04-24)**

**Platform:** aarch64-unknown-linux-gnu

**locale:** *LC\_CTYPE=en\_US.UTF-8, LC\_NUMERIC=C, LC\_TIME=en\_US.UTF-8, LC\_COLLATE=en\_US.UTF-8, LC\_MONETARY=en\_US.UTF-8, LC\_MESSAGES=en\_US.UTF-8, LC\_PAPER=en\_US.UTF-8, LC\_NAME=C, LC\_ADDRESS=C, LC\_TELEPHONE=C, LC\_MEASUREMENT=en\_US.UTF-8 and LC\_IDENTIFICATION=C*

**attached base packages:** *stats, graphics, grDevices, utils, datasets, methods and base*

**other attached packages:** *gtsummary(v.1.7.2), DescTools(v.0.99.54), finalfit(v.1.0.7), coxme(v.2.2-20), bdsmatrix(v.1.3-7), pander(v.0.6.5), survminer(v.0.4.9), ggpubr(v.0.6.0), survival(v.3.5-8), datefixR(v.1.6.1), lubridate(v.1.9.3), forcats(v.1.0.0), stringr(v.1.5.1), dplyr(v.1.1.4), purrr(v.1.0.2), readr(v.2.1.5), tidyverse(v.1.3.1), tibble(v.3.2.1), ggplot2(v.3.5.1), tidyverse(v.2.0.0) and readxl(v.1.4.3)*

**loaded via a namespace (and not attached):** *gridExtra(v.2.3), gld(v.2.6.6), rlang(v.1.1.3), magrittr(v.2.0.3), e1071(v.1.7-14), compiler(v.4.4.0), mgcv(v.1.9-1), vctrs(v.0.6.5), pkgconfig(v.2.0.3), shape(v.1.4.6.1), fastmap(v.1.2.0), backports(v.1.5.0), labeling(v.0.4.3), KMsurv(v.0.1-5), utf8(v.1.2.4), rmarkdown(v.2.27), markdown(v.1.12), tzdb(v.0.4.0), nloptr(v.2.0.3), xfun(v.0.44), glmnet(v.4.1-8), jomo(v.2.7-6), jsonlite(v.1.8.8), pan(v.1.9), broom(v.1.0.6), R6(v.2.5.1), stringi(v.1.8.4), car(v.3.1-2), boot(v.1.3-30), rpart(v.4.1.23), cellranger(v.1.1.0), Rcpp(v.1.0.12), iterators(v.1.0.14), knitr(v.1.47), zoo(v.1.8-12), Matrix(v.1.7-0), splines(v.4.4.0), nnet(v.7.3-19), timechange(v.0.3.0), tidyselect(v.1.2.1), rstudioapi(v.0.16.0), abind(v.1.4-5), yaml(v.2.3.8), ggtext(v.0.1.2), codetools(v.0.2-20), plyr(v.1.8.9), lattice(v.0.22-6), withr(v.3.0.0), evaluate(v.0.23), proxy(v.0.4-27), xml2(v.1.3.6), survMisc(v.0.5.6), pillar(v.1.9.0), carData(v.3.0-5), mice(v.3.16.0), foreach(v.1.5.2), generics(v.0.1.3), hms(v.1.1.3), commonmark(v.1.9.1), munsell(v.0.5.1), scales(v.1.3.0), rootSolve(v.1.8.2.4), minqa(v.1.2.7), xtable(v.1.8-4), class(v.7.3-22), glue(v.1.7.0), lmom(v.3.0), tools(v.4.4.0), data.table(v.1.15.4), lme4(v.1.1-35.3), ggsignif(v.0.6.4), Exact(v.3.2), mvtnorm(v.1.2-5), grid(v.4.4.0), colorspace(v.2.1-0), nlme(v.3.1-164), cli(v.3.6.2), km.ci(v.0.5-6), fansi(v.1.0.6), expm(v.0.999-9), broom.helpers(v.1.15.0), gt(v.0.10.1), gtable(v.0.3.5), rstatix(v.0.7.2), digest(v.0.6.35), farver(v.2.1.2), htmltools(v.0.5.8.1), lifecycle(v.1.0.4), httr(v.1.4.7), mitml(v.0.4-5), gridtext(v.0.1.5) and MASS(v.7.3-60.2)*

Licensed by CC BY unless otherwise stated.