Non-proportional hazards or unobserved heterogeneity in clustered survival data: Can we tell the difference?

Theodor Adrian Balan   Hein Putter

Leiden University Medical Center

ISCB Vigo // 10-13 July 2017

Scenario

Cluster i, individual j. Binary covariate x_{ij}. Non-informative censoring. Hazard of individual ij?

- assume proportional hazards, robust standard errors (`+cluster(id)`):

$$h_{ij}(t) = e^{\beta_0 x_{ij}} h_0(t)$$

  - *marginal* hazard $\equiv$ individual hazard

Scenario

Cluster i, individual j. Binary covariate x_{ij}. Non-informative censoring. Hazard of individual ij?

- assume proportional hazards, robust standard errors (`+cluster(id)`):

$$h_{ij}(t) = e^{\beta_0 x_{ij}} h_0(t)$$

  - *marginal* hazard $\equiv$ individual hazard

- assume *conditional* proportional hazards, using a random intercept (`+frailty(id)`):

$$\lambda_{ij}(t|Z_i) = Z_i e^{\beta_0 x_{ij}} \lambda_0(t)$$

  - *marginal* hazard:

$$\bar{\lambda}_{ij}(t) = \mathbf{E}[Z|O_i(t_-)] e^{\beta_0 x_{ij}} \lambda_0(t)$$

About frailty models

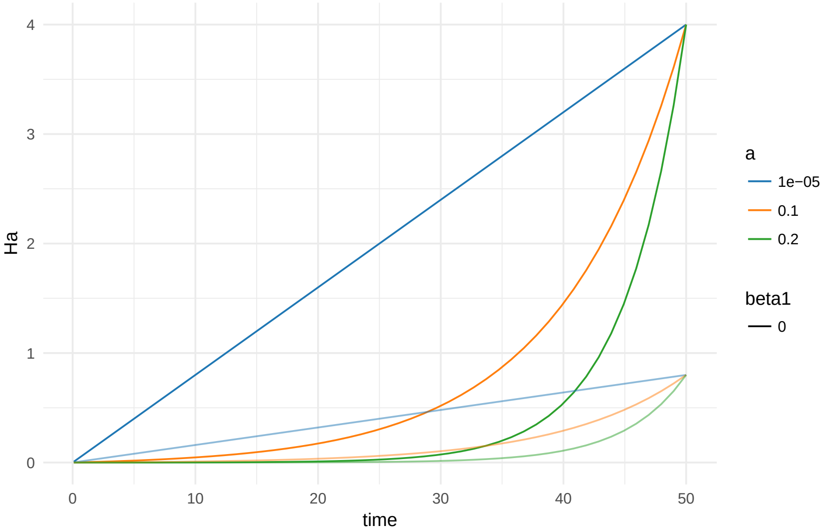
Facts

- If $\mathbf{E}Z < \infty$, then the marginal hazards are not proportional.
- If $\mathbf{E}Z < \infty$ and at least one covariate is present, then the frailty model is identifiable (Elberts & Ridder 1982)
- To a lesser (?) extent, the problem may persist with clustered survival data (Hougaard 2000)

Questions

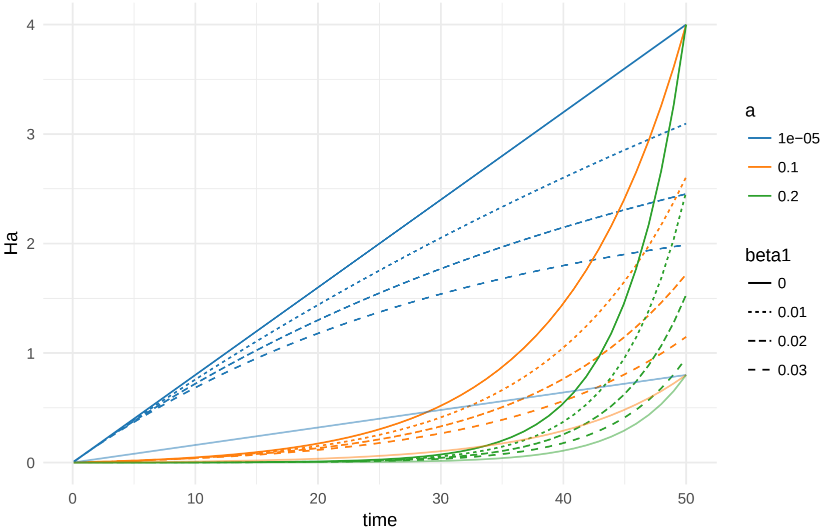
- Non-proportional hazards and no frailty: falsely detect frailty? (type 1 error)
- Non-proportional hazards and frailty: what happens? ("type 3" error)
- How does this depend on sample / cluster size?

Hazard with time dependent effect



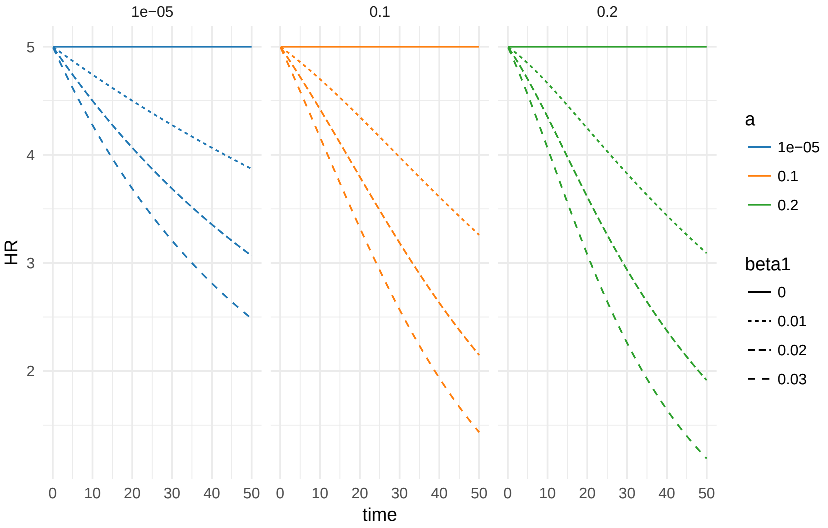
$h_0(t) = b\exp(a * t),\ h_A(t) = e^{\log 5 + \beta_1 t} h_0(t).$

Hazard with time-dependent effect



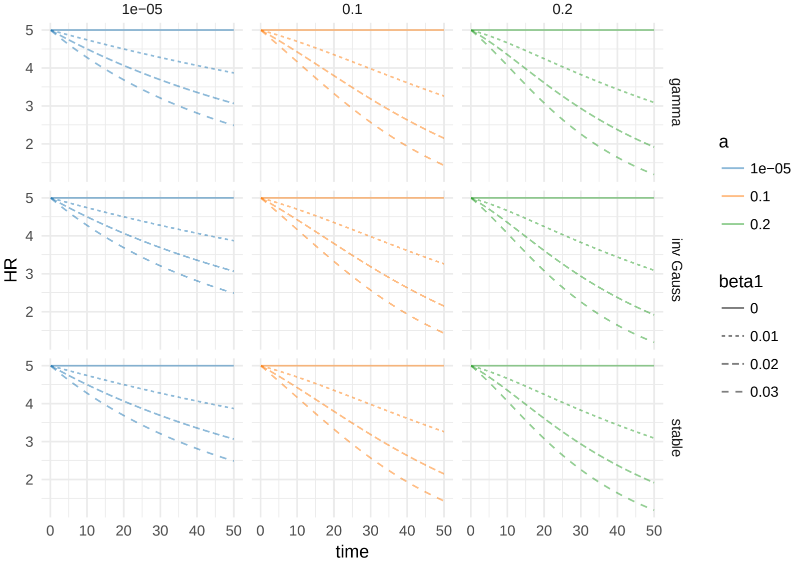
$h_0(t) = b \exp(a * t),\; h_A(t) = e^{\log 5 + \beta_1 t} h_0(t).$

Hazard ratio with time-dependent effect

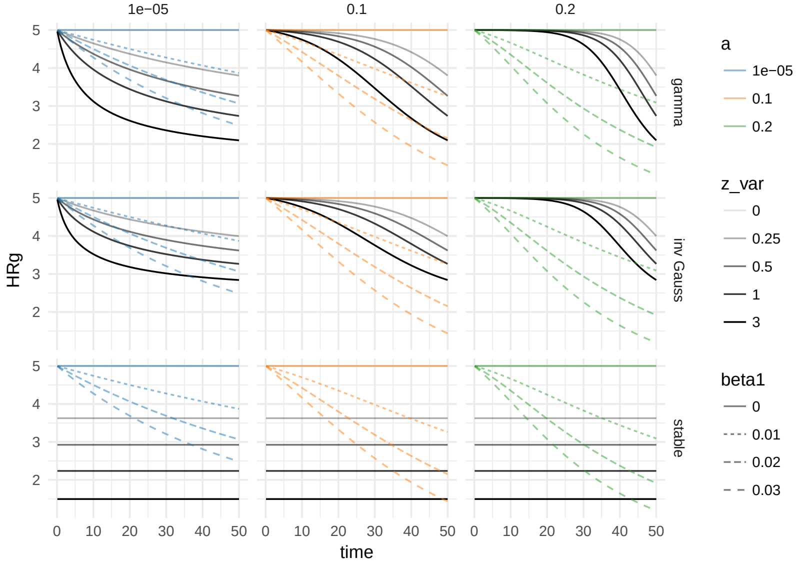


$h_0(t) = b\exp(a * t),\ h_A(t) = e^{\log 5 + \beta_1 t} h_0(t).$

Hazard ratios from frailty models



Hazard ratios from frailty models



The big question

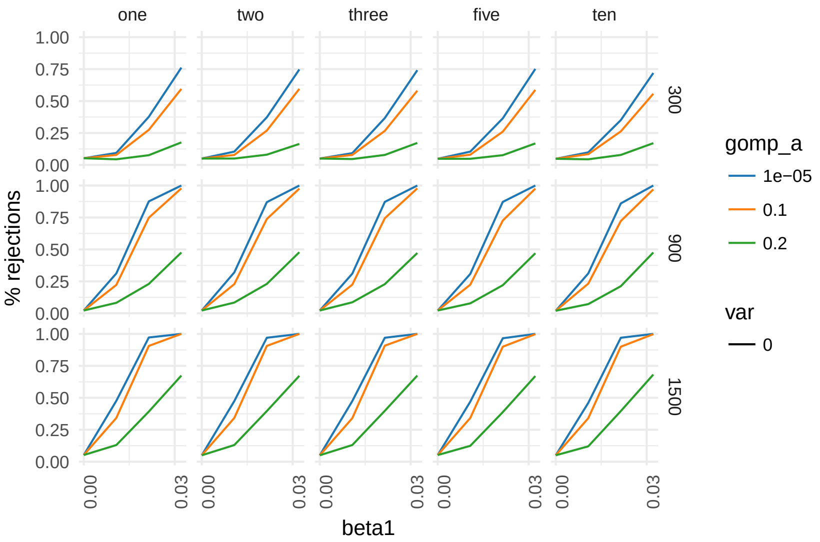
Non-proportional hazards or unobserved heterogeneity in clustered survival data: Can we tell the difference?

Simulation framework:

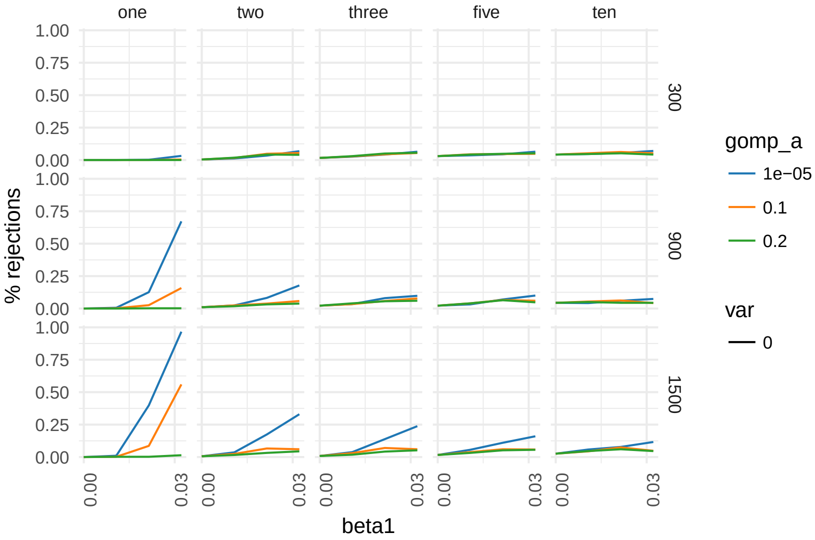
Take $\beta_0 = \log(5) \approx 1.6$. Gompertz distribution with $\beta_1(t) = \beta_1 t$
Also look in combination with a log-normal Z with variance 0, 0.25, 0.5.

- Marginal (Cox) model (`cox.zph` test & estimates)
  - `cox.zph` test & estimates
- Semi-parametric shared frailty models: **gamma**, **inverse Gaussian**, **positive stable**
  - Score test for heterogeneity (Commenges & Andersen 1995)
  - Likelihood ratio test for the presence of frailty
  - estimates of frailty variance (gamma, inverse Gaussian)
- `survival` and `frailtyEM` packages

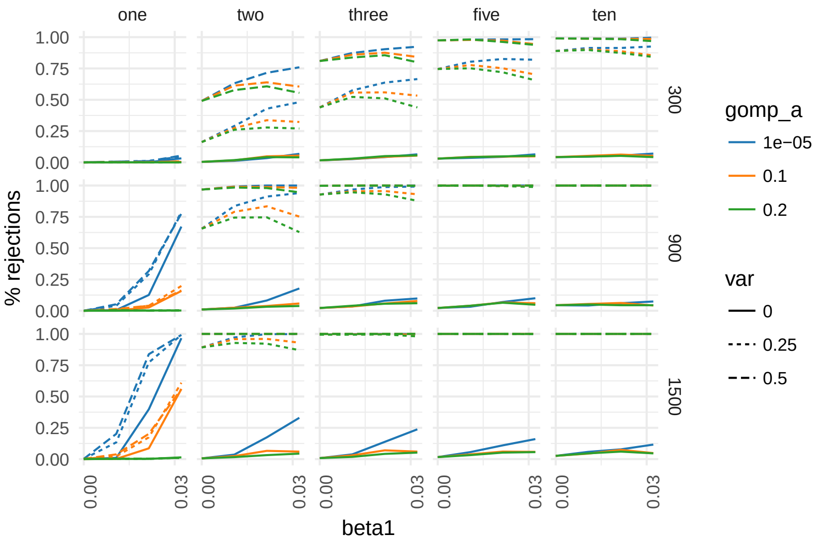
How non-proportional? - cox.zph



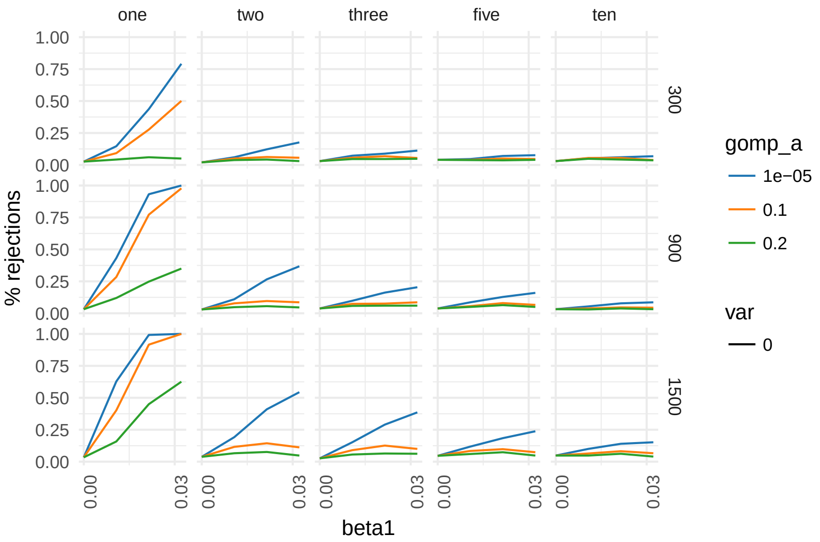
Commenges-Andersen score test for heterogeneity



Commenges-Andersen score test for heterogeneity

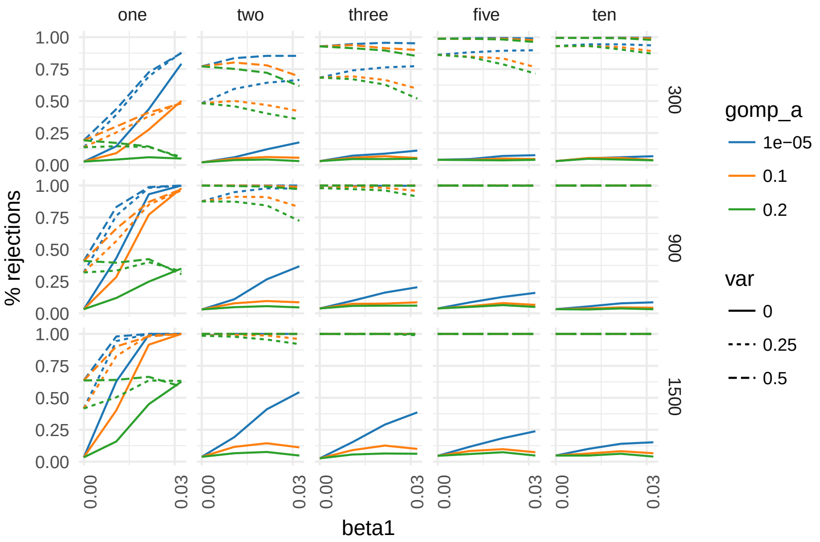


LRT - gamma



one
two
three
five
ten
300
900
1500
1.00
0.75
0.50
0.25
0.00
% rejections
0.00
0.03
beta1
gomp_a
1e−05
0.1
0.2
var
0

LRT - gamma

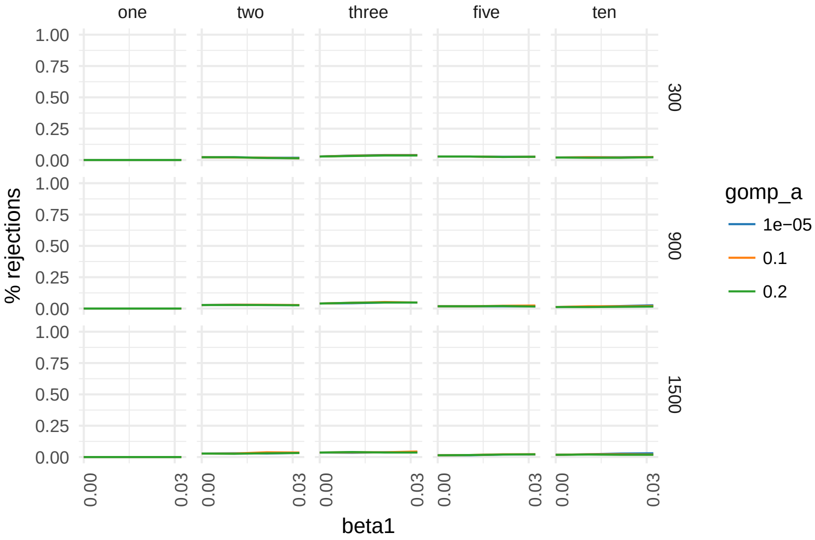


one
two
three
five
ten
1.00
0.75
0.50
0.25
0.00
% rejections
300
900
1500
0.00
0.03
beta1
gomp_a
1e−05
0.1
0.2
var
0
0.25
0.5

LRT - IG

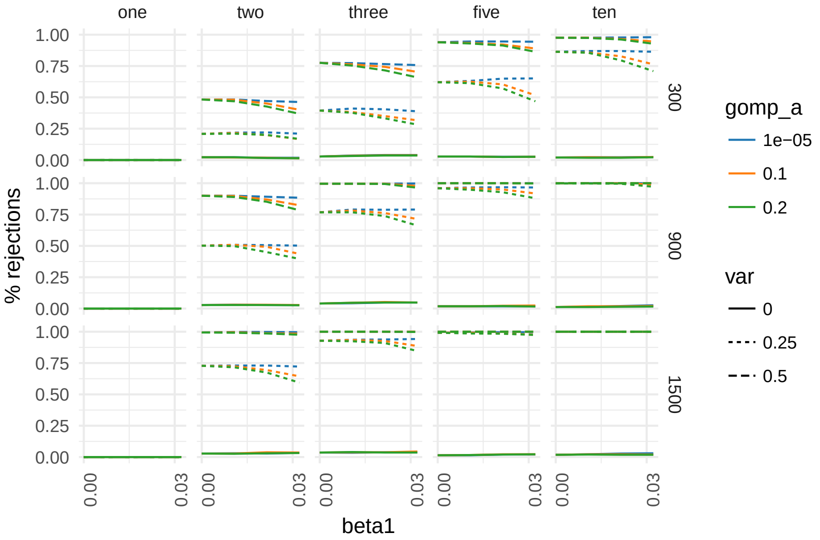
Almost just like the gamma!

LRT - stable



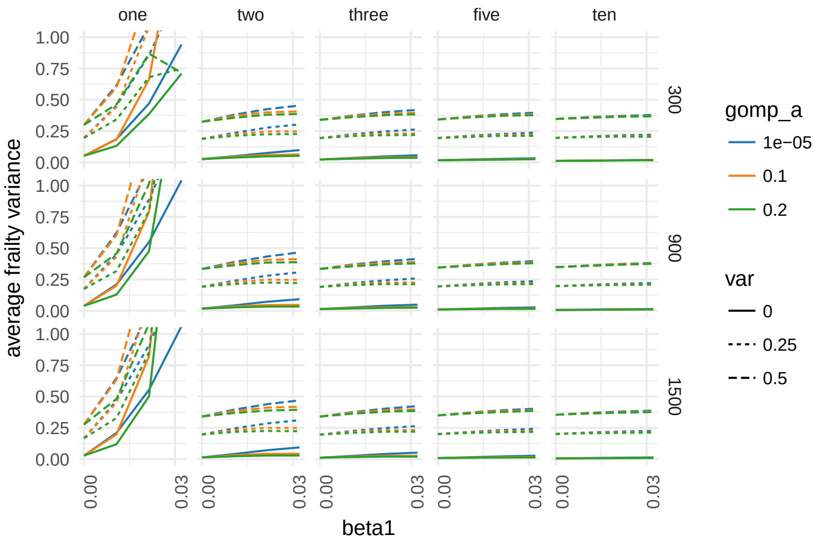
one
two
three
five
ten
1.00
0.75
0.50
0.25
0.00
300
900
1500
% rejections
0.00
0.03
beta1
gomp_a
1e−05
0.1
0.2

LRT - stable



one
two
three
five
ten
300
900
1500
1.00
0.75
0.50
0.25
0.00
% rejections
0.00
0.03
beta1
gomp_a
1e−05
0.1
0.2
var
0
0.25
0.5

estimated gamma frailty variance



Data example

Kidney data

Data on the recurrence times to infection, at the point of insertion of the catheter, for kidney patients using portable dialysis equipment. Catheters may be removed for reasons other than infection, in which case the observation is censored.

- Each patient has exactly 2 observations.
- Used for demonstrating shared frailty models in numerous places (incl. Therneau & Grambsch 2000).

```
## # A tibble: 76 × 6
##      id  time status   age    sex frail
##   <dbl> <dbl>  <dbl> <dbl>  <chr> <dbl>
## 1     1     8      1    28   male   2.3
## 2     1    16      1    28   male   2.3
## 3     2    23      1    48 female   1.9
## 4     2    13      0    48 female   1.9
## 5     3    22      1    32   male   1.2
## 6     3    28      1    32   male   1.2
## # ... with 70 more rows
```

Results that are usually presented

```
Call:
emfrail(formula = Surv(time, status) ~ age + sex + cluster(id),
    data = kdn2)

Regression coefficients:
             coef exp(coef)  se(coef) adjusted se         z      p
age     0.0054372 1.0054520 0.0115813   0.0116976 0.4694816 0.6387
sexmale 1.5528409 4.7248738 0.4451768   0.4995171 3.4881440 0.0005
Estimated distribution: gamma / left truncation: FALSE

Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.0245
(marginal) no-frailty Log-likelihood: -184.657
(marginal) Log-likelihood: -182.053
LRT: 1/2 * pchisq(5.21), p-val 0.0112

Frailty summary:
theta = 2.517 (1.49) / 95% CI: [0.97, 21.802]
variance = 0.397 / 95% CI: [0.046, 1.031]
Kendall's tau: 0.166 / 95% CI: [0.022, 0.34]
Median concordance: 0.162 / 95% CI: [0.022, 0.341]
E[log Z]: -0.212 / 95% CI: [-0.597, -0.023]
```

Is it really frailty?

```
Call:
emfrail(formula = Surv(time, status) ~ age + sex + cluster(id),
    data = kdn2, distribution = emfrail_dist(dist = "stable"))

Regression coefficients:
             coef exp(coef)  se(coef) adjusted se         z      p
age     0.0021816 1.0021839 0.0092248   0.0092248 0.2364892 0.8131
sexmale 0.8209988 2.2727687 0.2987240   0.2987245 2.7483521 0.0060
Estimated distribution: stable / left truncation: FALSE

Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.0245
(marginal) no-frailty Log-likelihood: -184.657
(marginal) Log-likelihood: -184.657
LRT: 1/2 * pchisq(-1.96e-05), p-val 0.5

Frailty summary:
theta = 105683.7 (33775246) / 95% CI: [2.879, Inf]
Kendall's tau: 0 / 95% CI: [0, 0.258]
Median concordance: 0 / 95% CI: [0, 0.255]
E[log Z]: 0 / 95% CI: [0, 0.2]
Var[log Z]: 0 / 95% CI: [0, 1.341]
```

zph test

`cox.zph()` test shows non-proportionality:

```
          rho   chisq        p
age    0.0214  0.0231 8.79e-01
sex    0.4390 29.2598 6.33e-08
GLOBAL     NA 29.3325 4.27e-07
```

zph test

`cox.zph()` test shows non-proportionality:

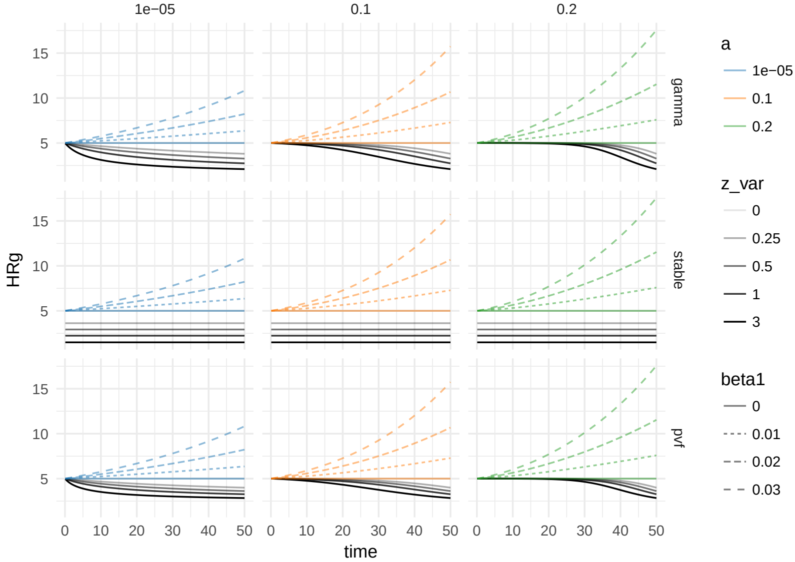
```
          rho   chisq        p
age    0.0214  0.0231 8.79e-01
sex    0.4390 29.2598 6.33e-08
GLOBAL     NA 29.3325 4.27e-07
```

`cox.zph()` test conditional on the frailty does not show the non-proportionality:

```
           rho   chisq     p
age    -0.0145 0.00427 0.948
sex     0.2170 1.39043 0.238
GLOBAL      NA 1.41146 0.494
```

Conclusion

Further questions



Conclusion

- Using frailty models for small clusters (or few recurrent events) might pick up marginal non-proportional hazards instead of heterogeneity
- Larger cluster size helps with distinguishing non-proportionality from heterogeneity
- Results sensitive to the actual shape of the hazard
- More complicated models (e.g. joint models) that use shared random effects to model recurrent events might simply pick up non-proportionality instead of heterogeneity

Subtle advertising

- `frailtyEM`: an `R` package for estimating semiparametric shared frailty models (Balan & Putter 2017, submitted, `CRAN` & `GitHub`)
- positive stable, inverse Gaussian, compound Poisson, left truncation, Commenges-Andersen test, nice plots, etc.