Application: Center effects

Theodor Balan & Hein Putter

Department of Medical Statistics and Bioinformatics
Leiden University Medical Center

Frailty models: Theory & Practice
November 3, 2017, Prague

Outline

Outline

Comparing centers

- Comparison of centers or benchmarking is increasingly important
- Aim is to monitor centers annually
- To identify underperforming centers
- To learn from the best performing centers
- Ultimate aim is to improve health care
- But there are methodological issues in these center comparisons
- Hard to disentangle random from systematic differences
- Random effect models useful
- Frailty models when dealing with a time-to-event outcomes

Frailty models

- Say we have clusters $j = 1, \ldots, J$
- Individual with covariates x from cluster j

$$h(t \mid x, Z_j) = Z_j h_0(t) \exp(\beta^\top x)$$

- Z_j's are independent random variables with the same distribution (mean 1)
- If $Z_j > 1$, then the individuals in the cluster have higher event rate than expected on the basis of their covariates
- This can be due to insufficient case-mix correction (omitted covariates) or to sub-optimal treatment
- No distinction possible between these two explanations

Outline

CLL data

- Data collected by EBMT (European Society for Blood and Marrow Transplantation)
- CLL Data Quality Initiative
- Outcome after allogeneic stem cell transplantation of chronic lymphocytic leukemia (CLL) patients
- Failures of interest: death, relapse
- Covariates of interest: donor category, remission status, age, year of SCT (stem cell transplantation), combination of sex recipient-donor

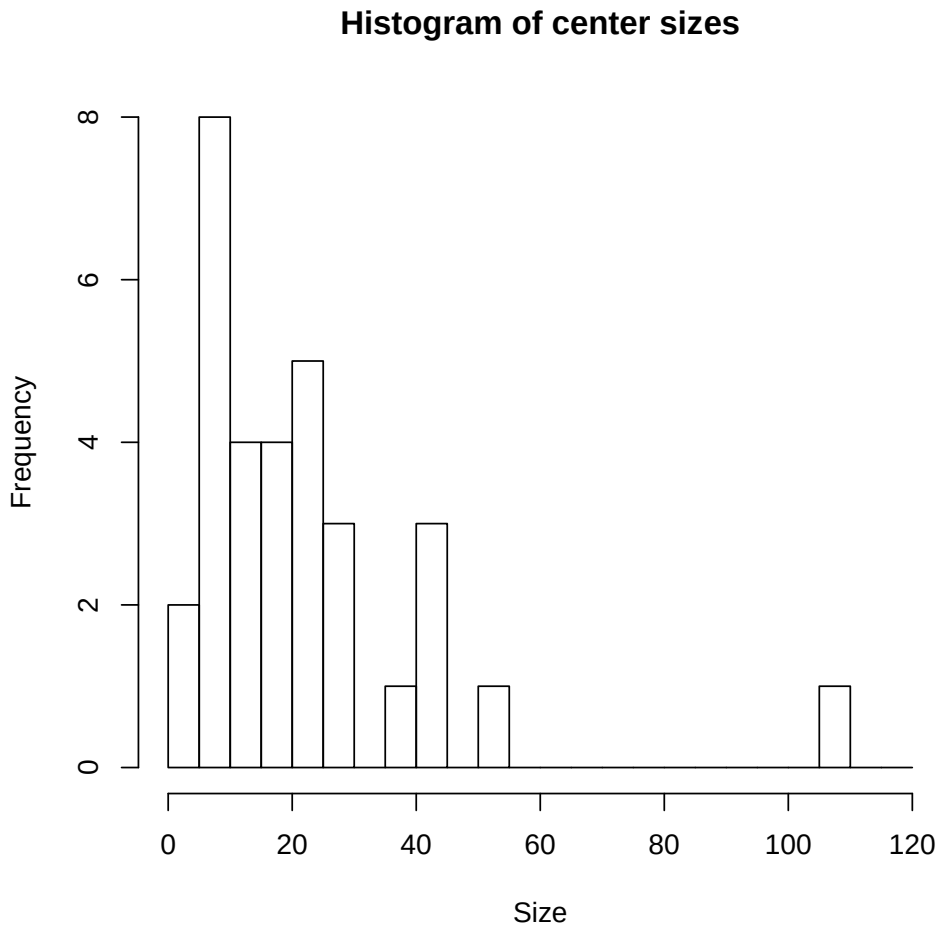
Prognostic factors

$n = 724$

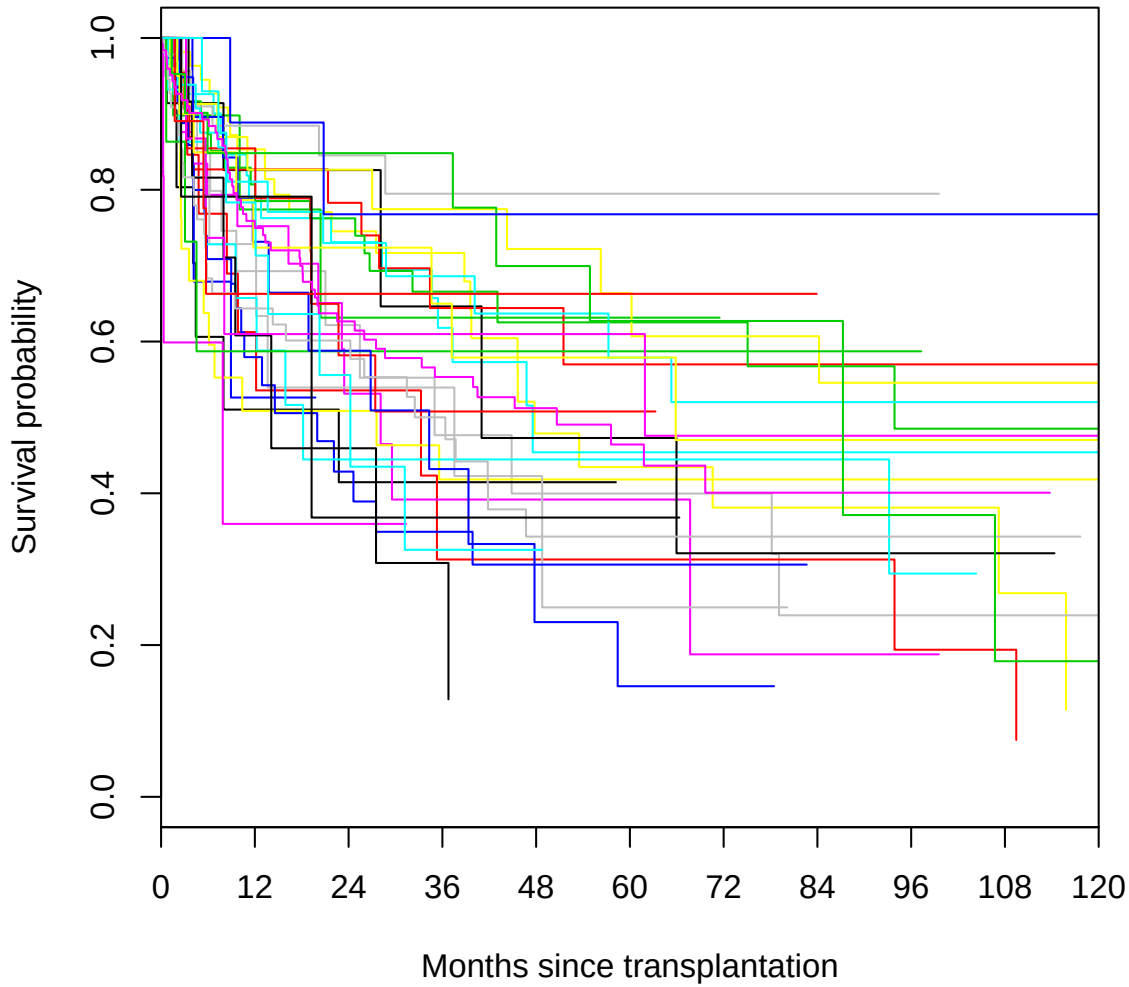
Prognostic factor		n (%)
Donor	HLA-id sib donor	291 (40)
	Other	433 (60)
Remission status	Complete Remission	94 (13)
	Partial Remission	366 (51)
	Stable/progressive disease	264 (36)
Sex match	other	567 (78)
	female to male	157 (22)
Age at SCT		55 (19–74)
Year of SCT		2007 (2000–2011)

Centers

- The CLL data has 32 centers, sizes shown below



Stratified survival curves



Ordinary Cox model

```
> coxOS <- coxph(Surv(srv_mo, srv_s) ~ age + remstat + sibdonor + sexm
+                  data=cll, method="breslow")
> coxOS

Call:
coxph(formula = Surv(srv_mo, srv_s) ~ age + remstat + sibdonor +
    sexmatch, data = cll, method = "breslow")


                          coef exp(coef) se(coef)    z       p
age                      0.025      1.03  0.00699 3.58 0.00035
remstatPR                0.214      1.24  0.18699 1.14 0.25000
remstatSD/PD             0.584      1.79  0.18755 3.11 0.00180
sibdonorother donor      0.291      1.34  0.11256 2.58 0.00980
sexmatchfemale to male 0.380      1.46  0.12523 3.03 0.00240

Likelihood ratio test=42.9  on 5 df, p=3.92e-08  n= 724, number of eve
```

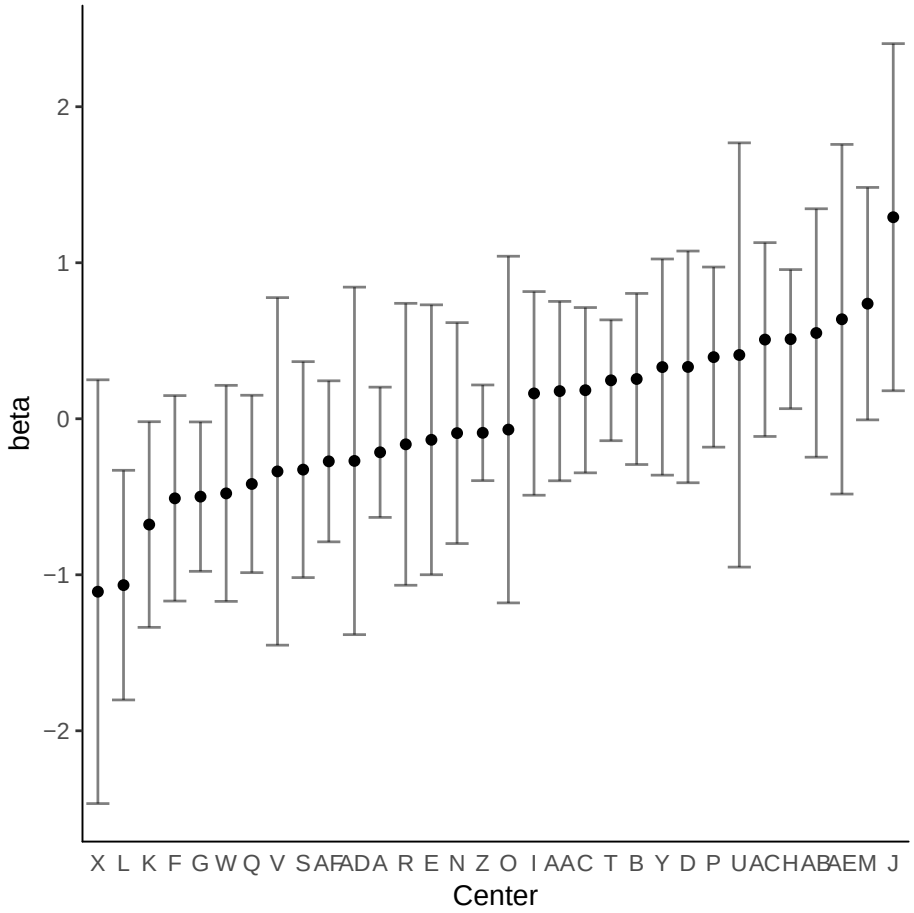
Cox model, with cluster as fixed effects

```
> coxOSfe <- coxph(Surv(srv_mo, srv_s) ~ age + remstat + sibdonor + se
+                  cic, data=cll, method="breslow")
> coxOSfe

Call:
coxph(formula = Surv(srv_mo, srv_s) ~ age + remstat + sibdonor +
    sexmatch + cic, data = cll, method = "breslow")


                              coef exp(coef) se(coef)    z       p
age                        0.02034   1.02055  0.00753 2.70 0.00693
remstatPR                  0.26224   1.29984  0.19609 1.34 0.18112
remstatSD/PD               0.64981   1.91517  0.19953 3.26 0.00113
sibdonorother donor        0.25310   1.28801  0.12454 2.03 0.04213
sexmatchfemale to male     0.49413   1.63907  0.13034 3.79 0.00015
cicAA                      0.39225   1.48031  0.35748 1.10 0.27253
cicAB                      0.76467   2.14829  0.46002 1.66 0.09646
.....
Likelihood ratio test=98.2 on 36 df, p=1.12e-07
n= 724, number of events= 351
```

Cox model, with cluster as fixed effects

- Only difference with stratified Cox model is that proportional hazards is now assumed for the centers
- Yields $J-1$ $\hat{\beta}$'s for J centers
- First center is reference
- J effects can be obtained by subtracting $\sum_{j=2}^{J} \hat{\beta}_j/J$ from each $\hat{\beta}_j$, and from 0 (reference center, nr 1)
- This yields, for each center, estimate and standard error, with the average β being 0
- So can be interpreted as deviation from average
  - Negative beta: better than average
  - Positive beta: worse than average
- Often visualized in a *caterpillar plot*

Caterpillar plot



Caterpillar plot

Pros and cons

- Pro
  - Quick and easy overview of relative quality of centers
- Cons
  - Inclined to pinpoint the worst center(s), even though variability may be large
  - Ranking is in general not a good idea
  - Ranking seen in this analysis may just be randomness

Fixed versus random effects

- Fixed effects assumes that interest is in centers A, B, C
- More natural to think of centers A, B, C etc as randomly drawn from a population of centers
- That implies that their effect on survival can be modeled by random effects
- Frailty model
- One (half) degree of freedom, instead of 31

Frailty Cox model

Using *coxph* from survival

```
> coxOSf <- coxph(Surv(srv_mo, srv_s) ~ age + remstat + sibdonor + sex
+                 frailty(cic), data=cll, method="breslow")
> coxOSf

Call:
coxph(formula = Surv(srv_mo, srv_s) ~ age + remstat + sibdonor +
    sexmatch + frailty(cic), data = cll, method = "breslow")

                          coef   se(coef) se2     Chisq DF   p
age                       0.0233 0.00725  0.00712 10.34  1.0 0.00130
remstatPR                 0.2360 0.19075  0.18888  1.53  1.0 0.22000
remstatSD/PD              0.6035 0.19235  0.18980  9.84  1.0 0.00170
sibdonorother donor       0.2895 0.11691  0.11445  6.13  1.0 0.01300
sexmatchfemale to male    0.4197 0.12716  0.12614 10.89  1.0 0.00097
frailty(cic)                                      21.75 11.6 0.03400

Iterations: 8 outer, 34 Newton-Raphson
     Variance of random effect= 0.0692   I-likelihood = -2087.4
Degrees of freedom for terms=  1.0  1.9  1.0  1.0 11.6
Likelihood ratio test=77.2  on 16.4 df, p=7.76e-10  n= 724
```

Frailty Cox model

Using `frailtyEM`

```
> mod_emfrail <- emfrail(formula = Surv(srv_mo, srv_s) ~ age + remstat
+                        sexmatch + cluster(cic), data = cll)
> mod_emfrail

Call:
emfrail(formula = Surv(srv_mo, srv_s) ~ age + remstat + sibdonor +
    sexmatch + cluster(cic), data = cll)

log-likelihood: -2087.437
theta: 14.50802

                              coef exp(coef) se(coef) adjusted se      z
age                         0.0233    1.0236   0.0073       0.0073 3.2180 0.
remstatPR                   0.2358    1.2660   0.1906       0.1908 1.2372 0.
remstatSD/PD                0.6033    1.8281   0.1923       0.1925 3.1370 0.
sibdonorother donor         0.2894    1.3357   0.1169       0.1170 2.4751 0.
sexmatchfemale to male      0.4195    1.5213   0.1271       0.1280 3.3011 0.

Score test for heterogeneity: p-val 0.0436
```

Summary function gives more information

```
Call:
emfrail(formula = Surv(srv_mo, srv_s) ~ age + remstat + sibdonor +
    sexmatch + cluster(cic), data = cll)

Regression coefficients:
                          coef exp(coef) se(coef) adjusted se      z
age                     0.0233    1.0236   0.0073       0.0073 3.2180 0.
remstatPR               0.2358    1.2660   0.1906       0.1908 1.2372 0.
remstatSD/PD            0.6033    1.8281   0.1923       0.1925 3.1370 0.
sibdonorother donor     0.2894    1.3357   0.1169       0.1170 2.4751 0.
sexmatchfemale to male 0.4195    1.5213   0.1271       0.1280 3.3001 0.
Estimated distribution: gamma / left truncation: FALSE

Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.0436
(marginal) no-frailty Log-likelihood: -2090.276
(marginal) Log-likelihood: -2087.437
LRT: 1/2 * pchisq(5.68), p-val 0.0086

Frailty summary:
theta = 14.508 (9.05) / 95% CI: [5.168, 118.131]
variance = 0.069 / 95% CI: [0.008, 0.193]
Kendall's tau: 0.033 / 95% CI: [0.004, 0.088]
```

Comparing ordinary and frailty Cox model

- Gamma frailty used (by default)
- Frailty variance estimated to be quite small, 0.069, but significant
- The estimated regression coefficients of the frailty Cox model are mostly larger (in absolute value) than the ordinary one, with higher standard error
- This difference will be larger with larger frailty variance
- This is a well-known general phenomenon, in the presence of unobserved heterogeneity, ignoring this heterogeneity leads to estimates that are biased (attenuated)

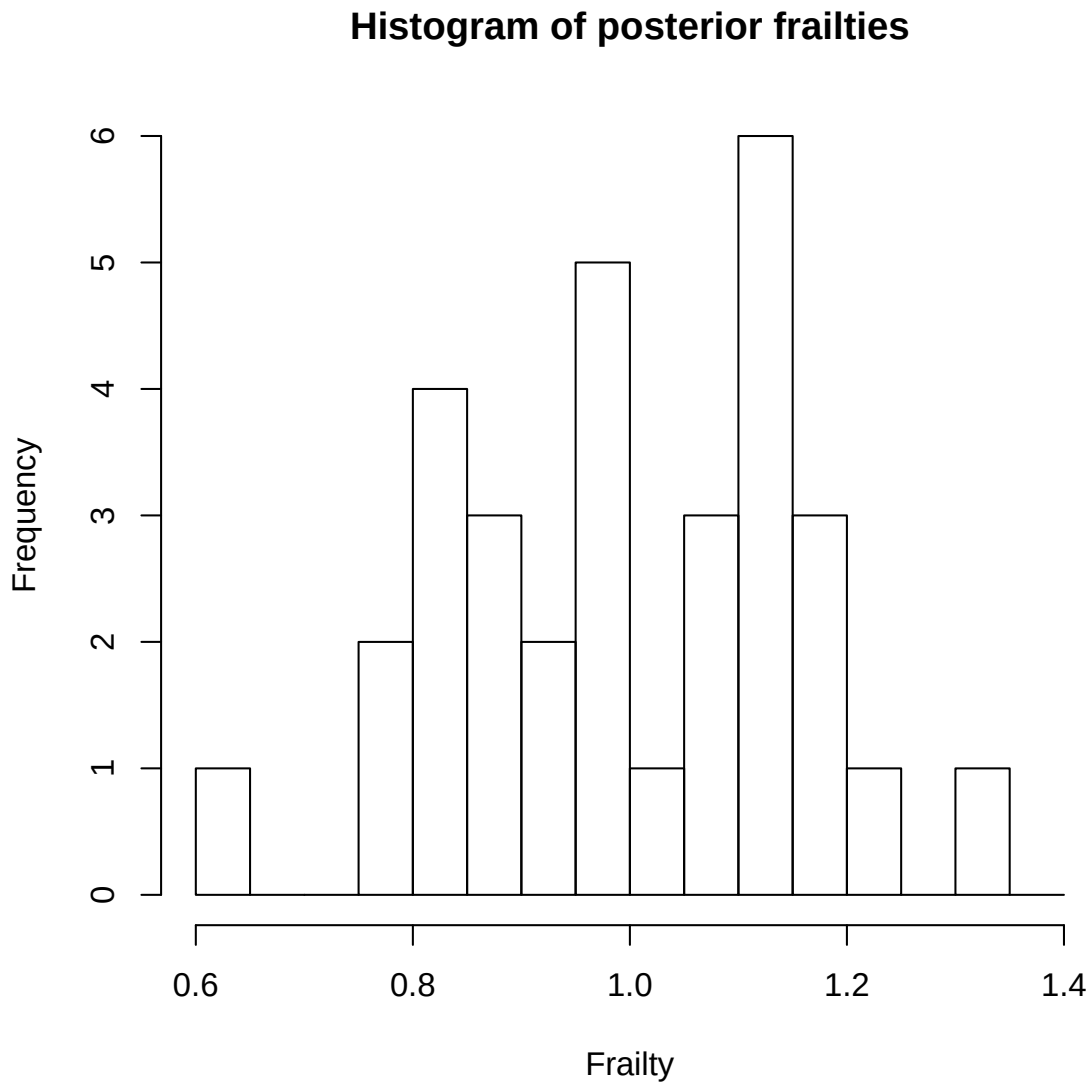
Posterior frailties

- If the frailty distribution is gamma(θ, θ) with variance θ^{-1}, the posterior distribution of Z_j, given the data of center j is gamma$(\hat{\theta} + D_j, \hat{\theta} + \hat{H}_j)$, with
  - $D_j = \sum_k d_{jk}$: total number of events in cluster j (observed)
  - $\hat{H}_j = \sum_k \hat{H}_0(t_{jk}) \exp(\hat{\beta}^\top x_{jk})$ (expected)
- The posterior expectation (also called *empirical Bayes estimate*) is given by

$$\hat{Z}_j = \frac{\hat{\theta} + D_j}{\hat{\theta} + \hat{H}_j}$$

- In fact, the whole posterior distribution is known (also variance, quantiles etc)

Posterior frailties



Posterior frailties

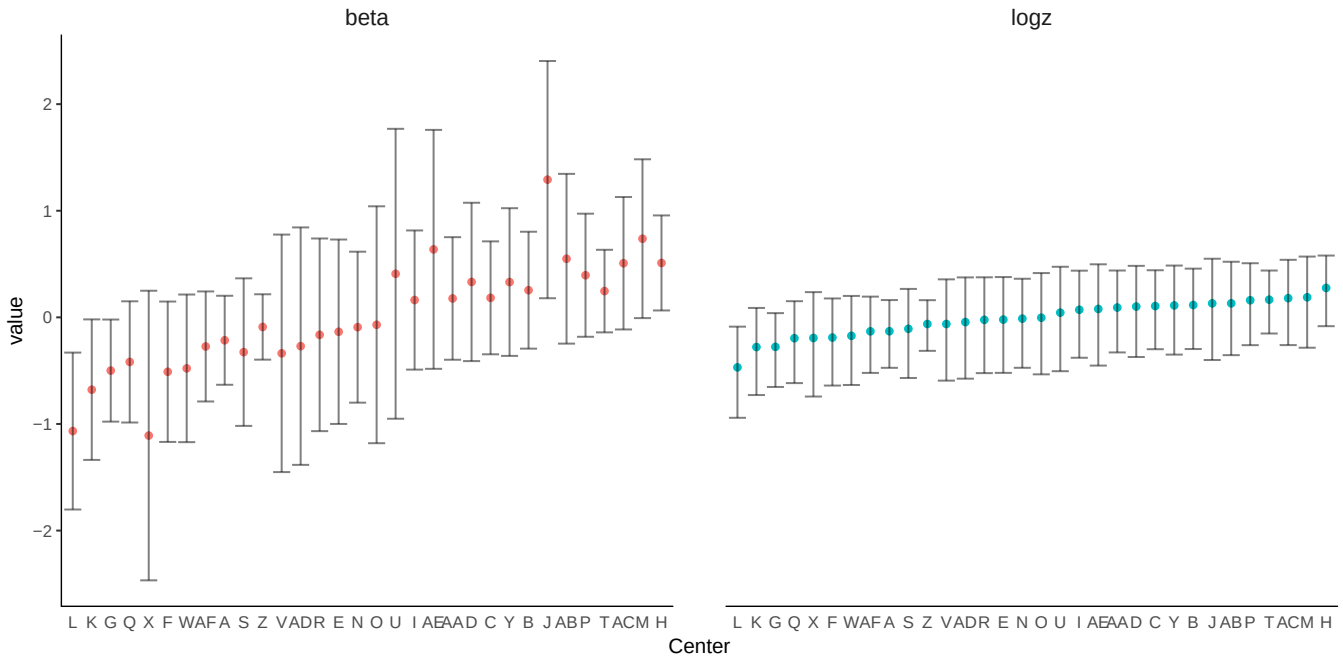
- The six lowest frailties

	cic	frail
18	L	0.6248699
17	K	0.7563921
13	G	0.7578368
23	Q	0.8218523
30	X	0.8237069
12	F	0.8268701

- The six highest frailties

	cic	frail
3	AB	1.141098
22	P	1.174341
26	T	1.180178
4	AC	1.197522
19	M	1.207932
14	H	1.318338

Caterpillar plots



Shrinkage

- The posterior expectation (empirical Bayes estimate) is given by

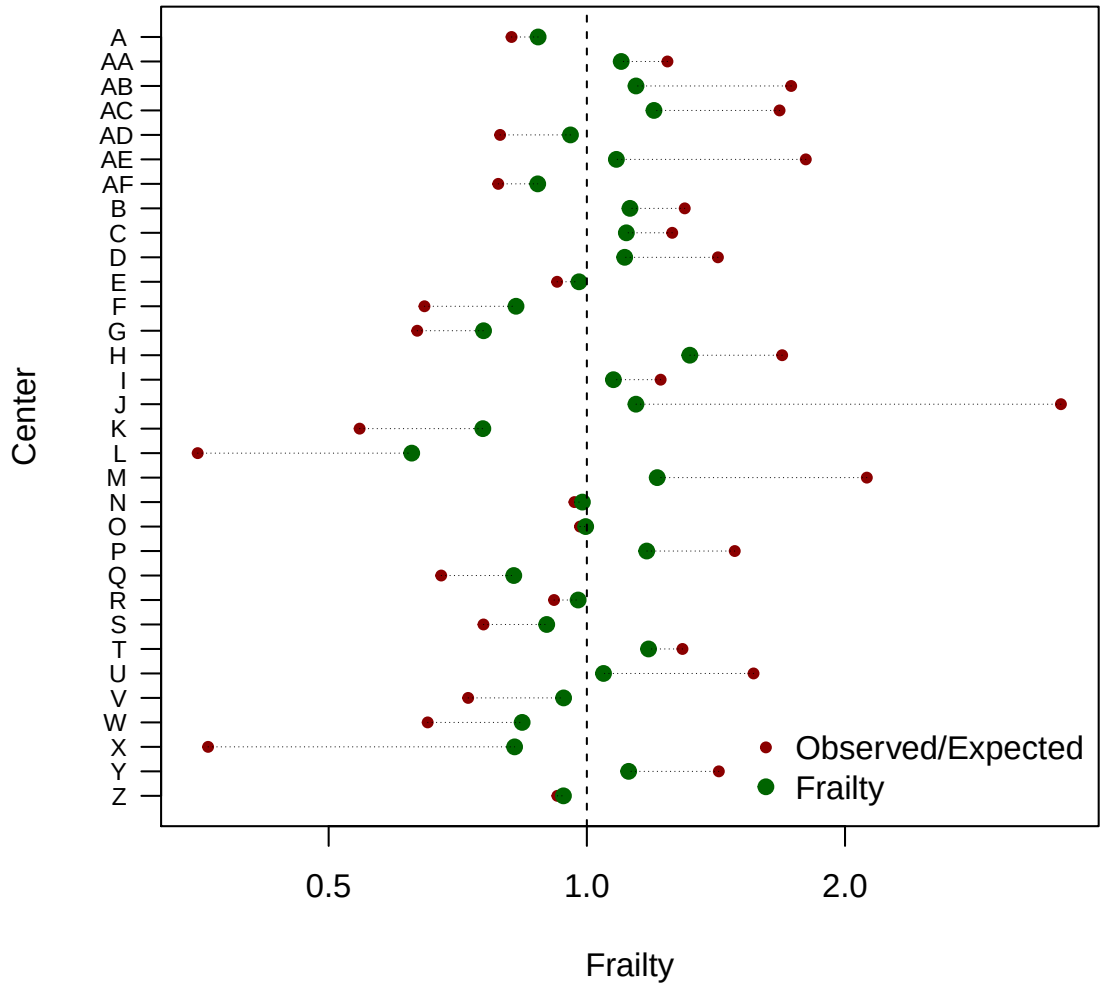
$$\hat{Z}_j = \frac{\hat{\theta} + D_j}{\hat{\theta} + \hat{H}_j}$$

- $\hat{Z}_j$ is a weighted average of observed/expected in cluster j and the overall average (1) in the complete data

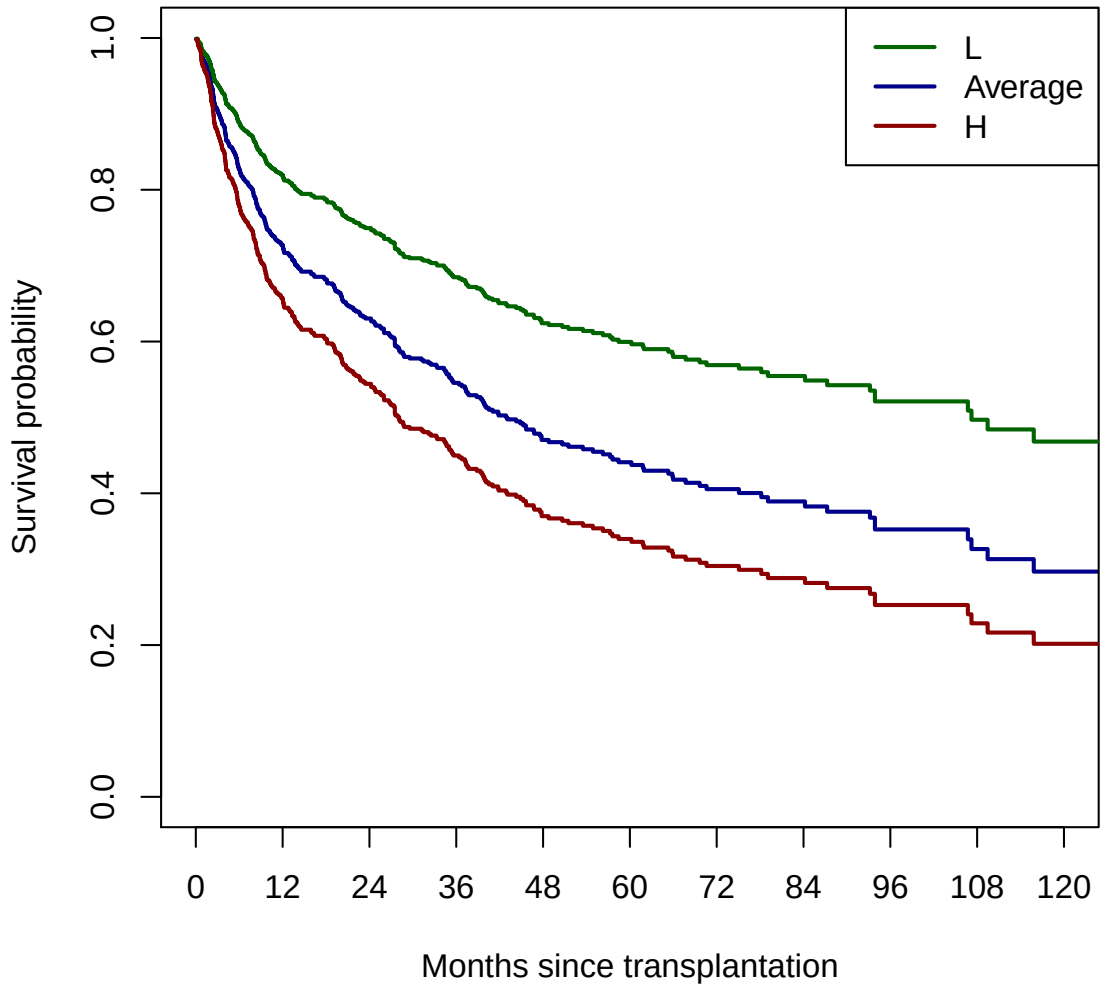
$$\hat{Z}_j = \frac{\hat{\theta} + D_j}{\hat{\theta} + \hat{H}_j} = \lambda \frac{D_j}{\hat{H}_j} + (1 - \lambda), \lambda = \frac{\hat{H}_j}{\hat{\theta} + \hat{H}_j}$$

- Shrinkage towards 1 is stronger for smaller centers (smaller $\hat{H}_j$)

Shrinkage


Center
A
AA
AB
AC
AD
AE
AF
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
0.5
1.0
2.0
Frailty
Observed/Expected
Frailty

Frailty-model based survival curves



Outline

Discussion

- Previous plot shows most extreme (L and H) and average ($Z = 1$) survival plots
- Much less variable than the model-based survival curves based on the stratified Cox model
- The most extreme centers in that plot were smallest centers, which have been shrunken to the average survival
- You could view the frailty-based curves as noise-free versions of the stratified Cox model-based curves
- Again: excess mortality in center H may be due to “bad treatment” but also to in adequate patient-mix correction!
- This is also true for other hospital performance comparisons

Summary

- Frailties are random effects models for survival data
- Used for two purposes
  1. Explaining lack of fit of univariate survival models, like deviation from the proportional hazards assumption
  2. The modeling of dependence in clustered data
- Useful for center comparisons
  - More realistic (shrunken) center effects provided by posterior means of frailties

Acknowledgements

- Johannes Schetelig (clinical background, CLL dataset)
- EBMT Data Office Leiden, DKMS Clinical Trials Unit Dresden, EBMT centers (CLL dataset)

Outline

Extensions

- So far emphasis on
  - Univariate frailties (selection effects, distortion of PH)
  - Shared frailty models
    - Clustered survival data
    - Recurrent events
- Shared frailty models assume
  - Single frailty, shared by all subjects in the cluster
  - Frailty does not change over time (*information* might!)

Limitations of the shared frailty model

- Frailty term has a double role
  - It quantifies dependence
  - It quantifies population heterogeneity
  - (With a conditional proportional hazards model it also quantifies non-PH)
- Shared frailty will only induce positive association within the group
- Looking back at where a frailty term might come from: if either x_{omit} or β_{omit} are time-varying, then the frailty $Z = \exp(\beta_{\text{omit}}^{\top} x_{\text{omit}})$ would also be time-varying

Two extensions considered

- Correlated frailties
- Time-varying frailties

Use of correlated frailties

Examples

- Twin pairs: each of the twins has its own frailty, they are correlated, possibly negatively correlated
  - Model provides not only variances of the frailties, but also separately the correlation between frailties in each group
- Competing risks in clustered data, one frailty for each competing risk
- Recurrent events: the frailty in first and later recurrent events might be different

Construction of correlated frailties

- For log-normal frailties it is simple
  - Specify $U = (U_1, U_2)$ bivariate normal with mean 0, covariance matrix $\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$
  - Set $X_1 = \exp(U_1)$ and $X_2 = \exp(U_2)$
  - Note that σ_1^2 and σ_2^2 and ρ are not the variances and correlation of $X = (X_1, X_2)$!
  - Both positive and negative correlation possible
- Gamma frailties: next slides
- Other frailties: similar construction as gamma

Construction of correlated frailties

How to make two correlated gamma random variables

- Construction by Yashin et al. (1995)
- Three independent gamma variables X_0, X_1, X_2
- $X_0 \sim \mathrm{Ga}(\rho\alpha, \beta)$, $X_1, X_2 \sim \mathrm{Ga}((1-\rho)\alpha, \beta)$
- $Z_1 = X_0 + X_1$, $Z_2 = X_0 + X_2$
- Z_1 and Z_2 both $\mathrm{Ga}(\alpha, \beta)$
- Correlation between Z_1 and Z_2 equals ρ (relative contribution of the *common* part X_0)
- Note: only positive correlation possible in this construction!

Estimation

- Estimation of these models is similar to, but somewhat more involved than, estimation of shared frailty models
- EM-type algorithms, where the E-step is more complicated than before
- No standard software available (but try `coxme` for log-normal frailties)

Time-varying frailties

- Model for hazard is

$$h(t \mid Z(t)) = Z(t)h(t)$$

- The conditional hazard $h(t)$ itself may follow a proportional hazards model
- Frailty varies over time
- Possibilities for $Z(t)$
  - Piecewise constant (Paik et al. 1994, Wintrebert et al. 2004)
  - Log-normal Gaussian process $Z(t) = \exp(U(t))$ (Yau & McGilchrist 1998)
  - Auto-correlated gamma processes (Putter & van Houwelingen 2015)
    - Implemented in package `dynfrail` (Balan & Putter, forthcoming)

Outline

Conclusion

- Discussed
  - Unobserved heterogeneity
  - Univariate frailty
  - Shared frailty models
  - Extensions
  - Software
  - Applications
- We hope this introduction to frailty models has been useful
- For questions, do not hesitate to contact us