Incorporating unobserved heterogeneity in Weibull survival models: A Bayesian approach

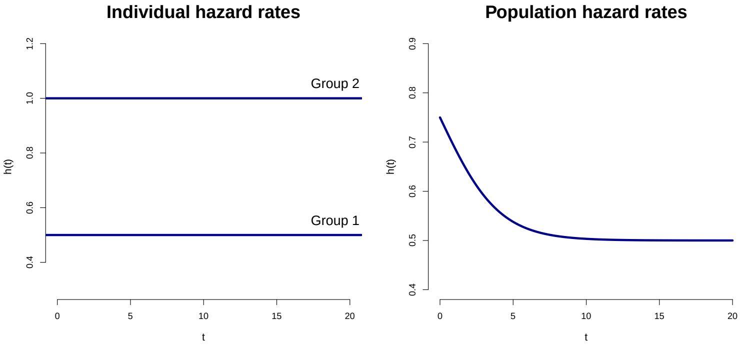
Catalina A. Vallejos[1] Mark F.J. Steel[2]

[1] MRC Biostatistics Unit, EMBL-European Bioinformatics Institute.
[2]Dept. of Statistics, University of Warwick.

Workshop on Flexible Models for Longitudinal and Survival Data with Applications in Biostatistics. 27-29 July, 2015

Motivation

What happens if we ignore **unobserved heterogeneity**?



Mixture families of life distributions

Definition

T_i is distributed as a **mixture of life distributions**, iff its density function is given by

$$f(t_i|\psi,\theta) \equiv \int_{\mathcal{L}} f^*(t_i|\psi,\Lambda_i=\lambda_i)\, dP_{\Lambda_i}(\lambda_i|\theta),$$

where $f^*(\cdot|\psi,\Lambda_i=\lambda_i)$ is a lifetime density and $P_{\Lambda_i}(\cdot|\theta)$ is a cdf on $\mathcal{L}$ possibly depending on a parameter θ, $\theta \in \Theta$.

- **Distinction** between individual and population-level survival
- The **intuition** behind the underlying model is preserved
- The influence of **outlying observations** is attenuated

Rate Mixtures of Weibull distributions

Definition

T_i is distributed as a **Rate Mixtures of Weibull** (RMW) distributions iff

$$T_i|\alpha, \gamma, \Lambda_i = \lambda_i \sim \text{Weibull}\left(\alpha\lambda_i, \gamma\right), \quad \Lambda_i|\theta \sim P_{\Lambda_i}(\cdot|\theta),$$

i.e.

$$f(t_i|\alpha, \gamma, \theta) = \int_{\mathcal{L}} \gamma\alpha\lambda_i \, e^{-\alpha\lambda_i t_i^{\gamma}} t_i^{\gamma-1} \, dP_{\Lambda_i}(\lambda_i|\theta), \quad t_i > 0,$$

where $\alpha, \gamma > 0$ and $P_{\Lambda_i}(\cdot|\theta)$ is a cdf on $\mathcal{L}$ possibly depending on a parameter $\theta \in \Theta$.

Denote $T_i \sim \mathbf{RMW}_P(\alpha, \gamma, \theta)$

Rate Mixtures of Weibull distributions

- Relates to existing literature in **frailty models**
  - Typically $\gamma = 1$ and $\Lambda_i \sim$ gamma (Lomax distribution) $\Rightarrow$ *e.g.* Jewell (1982), Abbring and Van Den Berg (2007)
  - Non-parametric mixtures $\Rightarrow$ *e.g.* Kottas (2006)

- Case $\gamma = 1$: Rate Mixtures of Exponentials $T_i \sim \text{RME}_P(\alpha, \theta)$

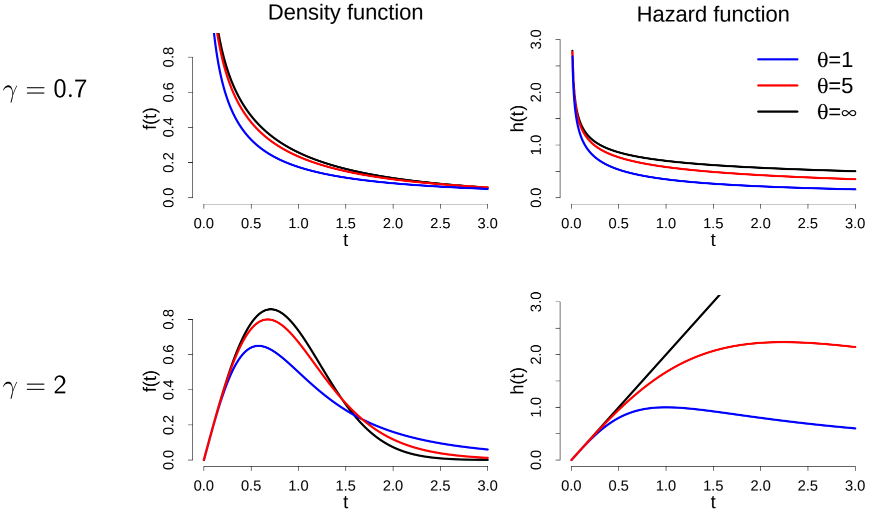
- If $T_i \sim \text{RME}_P(\alpha, \theta)$ then $T_i^{1/\gamma} \sim \text{RMW}_P(\alpha, \gamma, \theta)$.

- For $\gamma \leq 1$: decreasing hazard rate (Marshall and Olkin, 2007)

- Identifiability precludes unknown scale parameters in P
  $\Rightarrow$ Fix scale parameters in P or set $\text{E}(\Lambda_i | \theta) = 1$

Rate Mixtures of Weibull distributions

Example: RMW model with Gamma(θ, θ) mixing and $\alpha = 1$



Theorem

If all the required moments exist, the **coefficient of variation** *(cv) of distributions in the RMW family is*

$$cv(\gamma,\theta) = \sqrt{\frac{\Gamma(1+2/\gamma)}{\Gamma^2(1+1/\gamma)} \underbrace{\frac{var_{\Lambda_i}(\Lambda_i^{-1/\gamma}|\theta)}{E^2_{\Lambda_i}(\Lambda_i^{-1/\gamma}|\theta)}}_{(cv^*(\gamma,\theta))^2} + \underbrace{\frac{\left[\Gamma(1+2/\gamma) - \Gamma^2(1+1/\gamma)\right]}{\Gamma^2(1+1/\gamma)}}_{(cv^W(\gamma))^2}}.$$

It simplifies to $\sqrt{2\frac{var_{\Lambda_i}(\Lambda_i^{-1}|\theta)}{E^2_{\Lambda_i}(\Lambda_i^{-1}|\theta)} + 1}$ *when* $\gamma = 1$.

Theorem

If all the required moments exist, the **coefficient of variation** *(cv) of distributions in the RMW family is*

$$cv(\gamma,\theta) = \sqrt{\frac{\Gamma(1+2/\gamma)}{\Gamma^2(1+1/\gamma)} \underbrace{\frac{var_{\Lambda_i}(\Lambda_i^{-1/\gamma}|\theta)}{E^2_{\Lambda_i}(\Lambda_i^{-1/\gamma}|\theta)}}_{(cv^*(\gamma,\theta))^2} + \underbrace{\frac{\left[\Gamma(1+2/\gamma) - \Gamma^2(1+1/\gamma)\right]}{\Gamma^2(1+1/\gamma)}}_{(cv^W(\gamma))^2}}.$$

It simplifies to $\sqrt{2\frac{var_{\Lambda_i}(\Lambda_i^{-1}|\theta)}{E^2_{\Lambda_i}(\Lambda_i^{-1}|\theta)} + 1}$ *when* $\gamma = 1$.

If θ is unknown, we restrict the range of (γ, θ) such that cv is finite

A regression model based on RMW distributions

Proportional Hazards (PH) models are popular in this context

$$h_{T_i}(t_i|x_i, \beta, \Lambda_i = \lambda_i) = \lambda_i \gamma t_i^{\gamma-1} \, e^{x_i' \beta^*}, \qquad \Lambda_i \sim P_{\Lambda_i}(\theta)$$

A regression model based on RMW distributions

Proportional Hazards (PH) models are popular in this context

$$h_{T_i}(t_i|x_i,\beta,\Lambda_i=\lambda_i)=\lambda_i\gamma t_i^{\gamma-1}\,e^{x_i'\beta^*},\qquad \Lambda_i\sim P_{\Lambda_i}(\theta)$$

But the PH property is not preserved after mixture!

A regression model based on RMW distributions

Proportional Hazards (PH) models are popular in this context

$$h_{T_i}(t_i|x_i, \beta, \Lambda_i = \lambda_i) = \lambda_i \gamma t_i^{\gamma-1} e^{x_i' \beta^*}, \quad \Lambda_i \sim P_{\Lambda_i}(\theta)$$

But the PH property is not preserved after mixture!

Instead, we use an Accelerated Failure Times (AFT) specification

$$T_i \sim RMW_P(\alpha_i, \gamma, \theta), \quad \alpha_i = e^{-\gamma x_i' \beta},$$

which is equivalent to

$$\log(T_i) = x_i' \beta + \log(\Lambda_i^{-1/\gamma} T_0), \quad \Lambda_i \sim P_{\Lambda_i}(\theta), \quad T_0 \sim \text{Weibull}(1, \gamma)$$

A regression model based on RMW distributions

Proportional Hazards (PH) models are popular in this context

$$h_{T_i}(t_i|x_i, \beta, \Lambda_i = \lambda_i) = \lambda_i \gamma t_i^{\gamma-1} \, e^{x_i'\beta^*}, \quad \Lambda_i \sim P_{\Lambda_i}(\theta)$$

But the PH property is not preserved after mixture!

Instead, we use an Accelerated Failure Times (AFT) specification

$$T_i \sim RMW_P(\alpha_i, \gamma, \theta), \quad \alpha_i = e^{-\gamma x_i'\beta},$$

which is equivalent to

$$\log(T_i) = x_i'\beta + \log(\Lambda_i^{-1/\gamma} T_0), \quad \Lambda_i \sim P_{\Lambda_i}(\theta), \quad T_0 \sim \text{Weibull}(1, \gamma)$$

These regressions are equivalent setting $\beta = -\beta^*/\gamma$

First consider the RME case ($\gamma = 1$)

Jeffreys and independence Jeffreys priors have structure

$$\pi(\beta, \theta) \propto \pi(\theta),$$

but they are complicated to derive and $\pi(\theta)$ might not be proper.

First consider the RME case ($\gamma = 1$)

Jeffreys and independence Jeffreys priors have structure

$$\pi(\beta, \theta) \propto \pi(\theta),$$

but they are complicated to derive and $\pi(\theta)$ might not be proper.

Approach:

- Keep Jeffreys structure but use a proper $\pi(\theta)$
- Match priors through common proper prior for *cv*, say $\pi^*(cv)$
- Exploting the functional relationship between *cv* and θ

Table : Relationship between *cv* and θ for some RME models.

Mixing density	Range of *cv*	$cv(\theta)$	$\left\lvert \frac{dcv(\theta)}{d\theta} \right\rvert$
Gamma(θ, θ)	$(1, \infty)$	$\sqrt{\frac{\theta}{\theta-2}}$	$\theta^{-1/2}(\theta-2)^{-3/2}$
Inverse-Gamma$(\theta, 1)$	$(1, \sqrt{3})$	$\sqrt{\frac{\theta+2}{\theta}}$	$\theta^{-3/2}(\theta+2)^{-1/2}$
Inverse-Gaussian$(\theta, 1)$	$(1, \sqrt{5})$	$\sqrt{\frac{5\theta^2+4\theta+1}{\theta^2+2\theta+1}}$	$\frac{3\theta+1}{(5\theta^2+4\theta+1)^{1/2}(\theta+1)^2}$
Log-Normal$(0, \theta)$	$(1, \infty)$	$\sqrt{2\,e^{\theta}-1}$	$e^{\theta}(2\,e^{\theta}-1)^{-1/2}$

Bayesian inference for the RMW-AFT model

A weakly informative prior

For the general RMW case

We choose

$$\pi(\beta, \gamma, \theta) \propto \pi(\gamma, \theta) \equiv \pi(\theta|\gamma)\pi(\gamma),$$

where $\pi(\theta|\gamma)$ and $\pi(\gamma)$ are proper.

For the general RMW case

We choose

$$\pi(\beta, \gamma, \theta) \propto \pi(\gamma, \theta) \equiv \pi(\theta|\gamma)\pi(\gamma),$$

where $\pi(\theta|\gamma)$ and $\pi(\gamma)$ are proper.

Approach:

- Define $\pi(\theta|\gamma)$ as before through $\pi^*(cv)$, given γ
- Choose a proper $\pi(\gamma)$

Bayesian inference for the RMW-AFT model

A weakly informative prior

For the general RMW case

We choose

$$\pi(\beta, \gamma, \theta) \propto \pi(\gamma, \theta) \equiv \pi(\theta|\gamma)\pi(\gamma),$$

where $\pi(\theta|\gamma)$ and $\pi(\gamma)$ are proper.

Approach:

- Define $\pi(\theta|\gamma)$ as before through $\pi^*(cv)$, given γ
- Choose a proper $\pi(\gamma)$

These priors are improper but the posterior distribution is well defined under mild conditions

Bayesian inference for the RMW-AFT model

Outlier detection

No heterogeneity

$\Downarrow$

$\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$

Extreme value of λ_i

$\Downarrow$

Potential outlier

Bayesian inference for the RMW-AFT model

Outlier detection

No heterogeneity

$\Downarrow$

$\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$

Extreme value of λ_i

$\Downarrow$

Potential outlier

Formally, we contrast the models

$$M_0 \quad : \quad \Lambda_i = \lambda_{ref}$$
$$M_1 \quad : \quad \Lambda_i \neq \lambda_{ref} \text{ (with all other } \Lambda_j, j \neq i \text{ free)}$$

$$BF_{01}^{(i)} = \pi(\lambda_i | t, c) E\left(\frac{1}{dP(\lambda_i|\theta)}\right)\Bigg|_{\lambda_i = \lambda_{ref}}$$

Bayesian inference for the RMW-AFT model

Outlier detection

No heterogeneity

$\Downarrow$

$\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$

Extreme value of λ_i

$\Downarrow$

Potential outlier

Formally, we contrast the models

$$
\begin{aligned}
M_0 &: \quad \Lambda_i = \lambda_{ref} \\
M_1 &: \quad \Lambda_i \neq \lambda_{ref} \text{ (with all other } \Lambda_j, j \neq i \text{ free)}
\end{aligned}
$$

$$BF_{01}^{(i)} = \pi(\lambda_i | t, c) E\left(\frac{1}{dP(\lambda_i|\theta)}\right)\Bigg|_{\lambda_i = \lambda_{ref}}$$

Choice of λ_{ref}?

- In Vallejos and Steel (2014) we recommended $\lambda_{ref} = \mathrm{E}(\Lambda_i|\theta)$

Bayesian inference for the RMW-AFT model

Outlier detection

- In Vallejos and Steel (2014) we recommended $\lambda_{ref} = \mathrm{E}(\Lambda_i|\theta)$
- This is not appropriate for RMW models where censoring is very informative for λ_i's.

For censored observations we use **correction factor**

$$\lambda_{ref}^{c} = R_i(\beta, \gamma, \theta)\lambda_{ref}^{o}, \text{ with } R_i(\beta, \gamma, \theta) = \frac{E(\Lambda_i|t_i, c_i = 0, \beta, \gamma, \theta)}{E(\Lambda_i|t_i, c_i = 1, \beta, \gamma, \theta)}.$$

Applications

To illustrate, we analyse 2 real datasets:

Dataset	n	Censoring	# covariates
Veteran's administration lung cancer (VA)	137	7%	5
Cerebral palsy (CP)	1,549	84%	2

We use RMW-AFT models as well as a Weibull model.

We compare models defined by different mixing distributions using

- Bayes Factors
- Conditional Predictive Ordinate (CPO): for observation i,

$$\text{CPO}_i = f(t_i | t_{-i}),\ t_{-i} = (t_1, ..., t_{i-1}, t_{i+1}, ..., t_n),$$

where $f(\cdot | t_{-i})$ is the predictive density given t_{-i}.

- $\text{PsML} = \prod_{i=1}^{n} \text{CPO}_i$ (Geisser and Eddy, 1979)

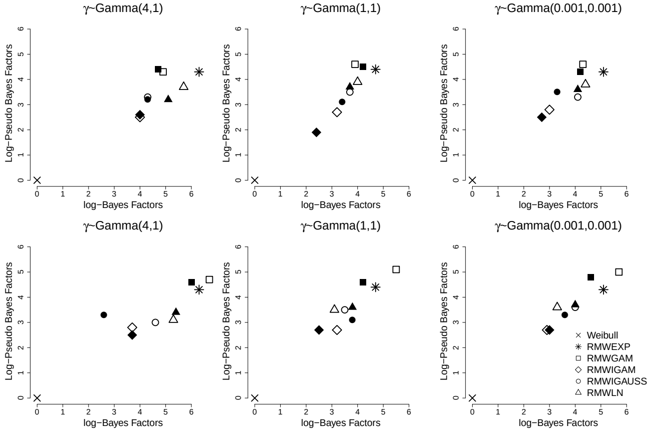
  $\Rightarrow$ Ratios of PsML's defining pseudo Bayes factors (PsBF)

Application: VA data

Model comparison in terms of BF and PsBF

E$(cv) = 1.5$

E$(cv) = 5.0$

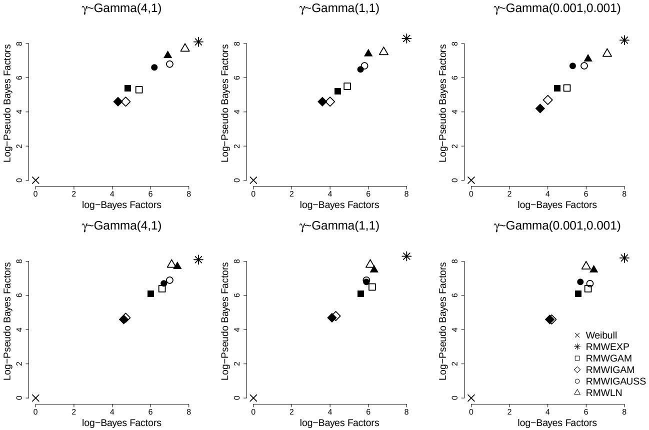


Application: CP data

Model comparison in terms of BF and PsBF (Geisser and Eddy, 1979)

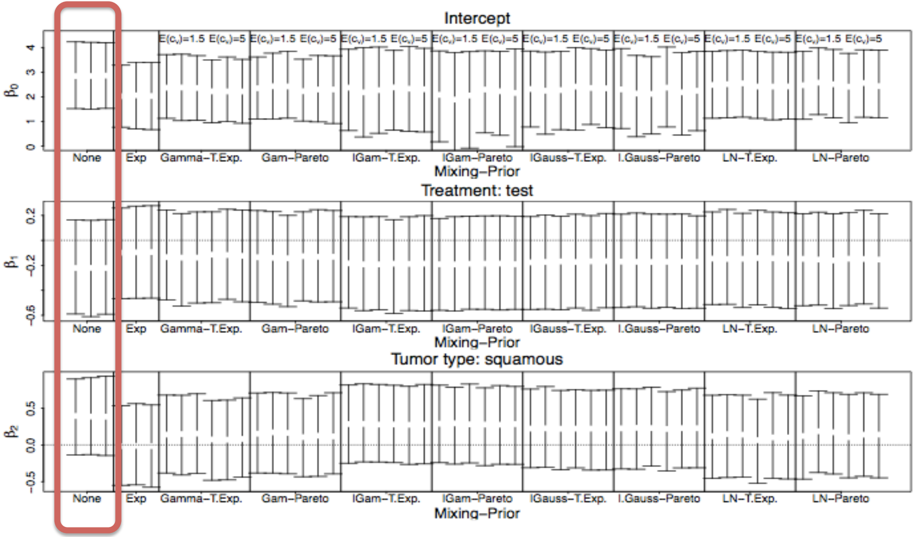
$E(cv) = 1.5$

$E(cv) = 5.0$



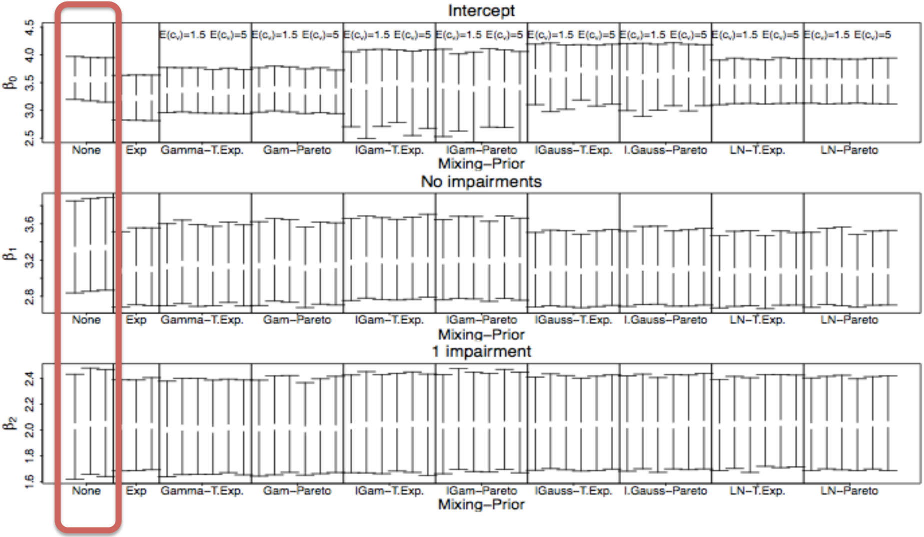
Applications: VA dataset

Posterior medians and HPD 95% interval for some regression coefficients



Applications: CP dataset

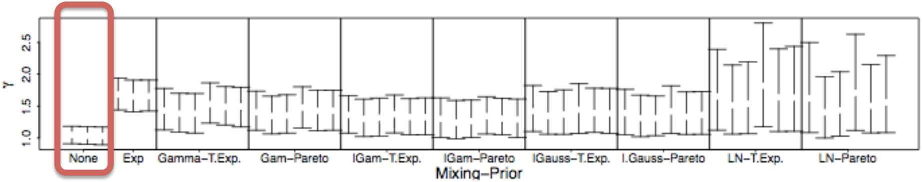
Posterior medians and HPD 95% interval for some regression coefficients



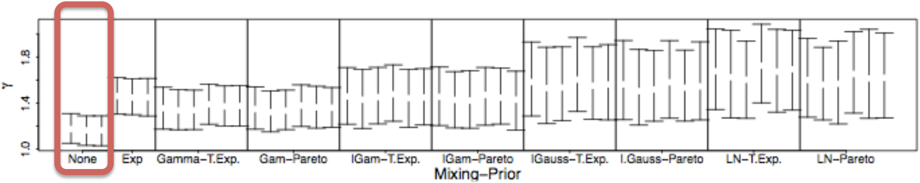
Applications

Posterior medians and HPD 95% interval for γ

VA dataset



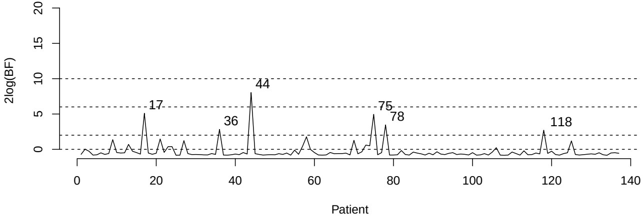
CP dataset



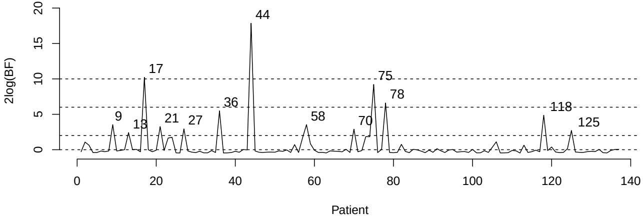
Application: VA data

Outlier detection (Gamma(θ,θ) mixing)

$\mathrm{E}(cv) = 1.5$

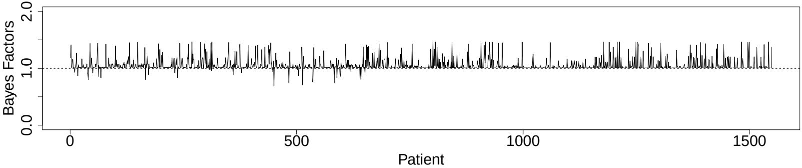


$\mathrm{E}(cv) = 5.0$



Application: CP data

Outlier detection (Exponential(1) mixing)



Conclusions

1. We explored mixtures of life distributions (e.g RMW family) to deal with unobserved heterogeneity and outliers
2. Covariates through AFT specification: retains AFT structure and the interpretation of β
3. Prior based on structure of Jeffreys prior, but allows meaningful BFs
4. Proposal of outlier detection method based on mixing parameters
5. Data support mixing; critical for estimation of β and γ

Acknowledgements

This research project was funded by

- University of Warwick
- Pontificia Universidad Católica de Chile

Many thanks to P.O.D. Pharoah and Prof. Jane Hutton for access to the cerebral palsy dataset.

Full references list and more details in

- C.A. Vallejos and M.F.J. Steel (2014), Incorporating unobserved heterogeneity in Weibull survival models: A Bayesian approach. *CRiSM-WP 14-20*.
- C.A. Vallejos and M.F.J. Steel (2015), Objective Bayesian survival analysis using scale mixtures of log-normal distributions. *JASA*