# Bayesian Cox models for assessing *S.* Typhimurium virulence changes under an alternative antimicrobial treatment

**E. Lázaro**[(1)], M. Sanz-Puig[(2)], D. Alvares[(1)], C. Armero[(1)], A. Martínez[(2)], and D. Rodrigo[(2)].

**September 9, 2016**

[(1)] Department of Statistics and Operations Research, Universitat de València, Burjassot.
[(2)] Instituto de Agroquímica y Tecnología de Alimentos (CSIC), Paterna.

# Table of contents

# Introduction

# Background

- ***Salmonella spp.*** is one of the most relevant foodborne pathogens which causes about 155,000 salmonellosis-related deaths each year worldwide.
- ***Salmonella enterica serovar Typhimurium*** is one of the most usual serotypes related to salmonellosis outbreaks.
- Different kinds of **alternative preservation treatments** have been developed to reduce or eliminate this pathogen.
- The addition of **bioactive substances** from nature or agroindustrial by-products or the application of non-thermal treatments such as **High Hydrostatic Pressure** and **Pulsed Electric Fields** are the most usual antimicriobial alternatives.
- Most of the **alternative treatments** can generate **bacterial changes** or **adaptations** as a consequence of **repetitive** applications.

# Experiment characterisation: aim

- **Evaluation** of the possible appearance of **bacterial adaptions** and **changes** caused by the **repetitive use** of an alternative preservation treatment in terms of virulence:
  - The not parasitic and transparent nematode, ***C. elegans***, was used as the host system.
  - Virulence was measured in terms of the lifespan of the worm, egg laying and mobility.
  - We only focus here on **lifespans**.

# Experiment characterisation: description

- **Three *S.* Typhimurium populations**:
  - **Untreated** *S.* Tyhimurium.
  - *S.* Tyhimurium treated **one time**.
  - *S.* Tyhimurium treated **three times**.
- A population of 250 individuals of *C. elegans* **synchronised** by **age** was exposed to each of the three microbial populations.
- Nematodes were distributed in 25 plates of 10 worms.
- **Lifespan** of each nematode:
  - Measured in days.
  - Monitored each two or three days until the individual is dead.
  - Alive nematodes were transfered to a new plate at each monitoring time. This process can cause the accidental death of the nematode.

# Modeling strategy

# General Statistical context

- ***C. elegans*** lifespan is modeled in each of the three populations of *S.* Typhimurium considered.
- **Survival regression model** with:
  - **Right censored** observations accidentally caused in the process of changing the worm from a plate to another one.
  - **Interval censored** observations because the event of interest, death of the worm, occurred between two consecutive monitoring times.
  - **Correlated observations** off the worms within the same plate.

**Two** types of **models**:

- **Cox Proportional Hazards** (CPH) models
- **Accelerated Failure Time** (AFT) models

# First Statistical context

- ***C. elegans*** lifespan is modeled in each of the three populations of *S.* Typhimurium considered.
- **Survival regression model** with:
  - **Right censored** observations accidentally caused in the process of changing the worm from a plate to another one.
  - ~~**Interval censored**~~ observations because the event of interest, death of the worm, occurred between two consecutive monitoring times.
  - ~~**Correlated observations**~~ off the worms within the same plate.

**Two** types of **models**:

- **Cox Proportional Hazards** (CPH) models
- ~~**Accelerated Failure Time**~~ (AFT) models

# Statitistical analysis: characterisation

Work in the context of **CPH models**:

- Allow to make inference about the parameters without assuming any distribution for the survival time (*semiparametric models*).
- Result in a natural estimate for the risk for failure associated with a vector of covariates.
- Widely accepted and utilized model for survival analysis.

Work in the context of the **right censored** data and ignoring interval censored by imputing the exact event time:

- Interval censored data analysis presents some implementation particularities in Bayesian software.
- **Interval censored observations** have been **imputed** into **non censored** through a random function defined in their respective interval time.

Work in the context of the **Bayesian Analysis**:

- Powerful and flexible methodology in relation to the modeling and quantification of uncertainties.
- Good in dealing with censored data.

# Statistical analysis: Cox Proportional Hazard Model

- The CPH model assumes the following modeling for the hazard function

$$h(t \mid h_0, \mathbf{x}, \beta) = h_0(t)\, \mathrm{e}^{\mathbf{x}'\beta}$$

  where:

  - $h_0(\cdot)$: **baseline hazard function**. It is a completely arbitrary hazard function that determines a baseline distribution with density $f_0(\cdot)$, cdf $F_0(\cdot)$, and survival function $S_0(\cdot)$.
  - $\beta$ is a vector of unknown parameters associated to covariates $\mathbf{x}$. The element $\mathbf{x}'\beta$ is known as the linear predictor.
- Much of the popularity of the CPH model is that Cox proposed methods for estimating the coefficients $\beta$ without knowing $h_0(t)$.
- **Bayesian** inference **always** needs a model for $h_0(t)$.

# Statistical analysis: Baseline Hazard Function proposals

- **Our proposals** for modeling $h_0(t)$ are based on:

  1. **Weibull distribution**
     - $h_0(t \mid \alpha, \lambda) = \lambda \alpha t^{\alpha-1}$.

  2. **Piecewise function**,
     - Time is divided into $g$ prespecified intervals $I_k = (t_{k-1}, t_k]$ for $k = 1, 2, \ldots, g$ where $0 = t_0 < t_1 < \ldots < t_g < \infty$, being $t_g$ the last survival or censored time.
     - Baseline hazard is constant within intervals, $h_0(t_i) = \lambda_k \quad \text{for} \quad t_i \in I_k$

# Statistical analysis: Bayesian process

- Elicit a prior distribution for $(h_0, \beta)$ but to do this we first need to model the baseline hazard function, $h_0(\cdot)$:
  1. Weibull.
  2. Piecewise:
     - 2.1 **Independent** $\lambda$'s
     - 2.2 $\lambda$'s correlated through a discrete-time **martingale process**.
     - 2.3 **Logarithmic transformation** of $\lambda$'s correlated through a discrete-time **martingale process**.
     - 2.4 $\lambda$'s correlated through a **Gamma process** prior.
- Construct the likelihood function of $(\beta, h_0)$ for data $\mathcal{D} = \{(y_i, \delta_i, \mathbf{x}_i),\ i = 1, \ldots, n\}$

$$L(h_0, \beta) = \prod_{i=1}^{n} L_i(h_0, \beta) = \prod_{i=1}^{n} [h_i(y_i \mid h_0, \beta]^{\delta_i} [S_i(y_i \mid h_0, \beta]$$

- Compute (approximate) the posterior distribution for $(h_0, \beta)$
- Compute posterior or predictive distributions for the relevant elements in the problem.

# Statistical analysis: Model specification, I

- CPH model:
  - $h_i(t \mid \mathbf{x}_i, \beta, h_0) = h_0(t) \exp\{\beta_{ST1}\, I_{ST1} + \beta_{ST3}\, I_{ST3}\}$
    - Baseline group: worms fed with untreated *S.*Typhimurium.
    - $I_{ST1}(i) = 1$ if the worm $i$ were fed with *S.* Typhimurium which survived one treatment application.
    - $I_{ST3}(i) = 1$ if the worm $i$ were fed with *S.* Typhimurium which survived three treatment applications.
  - | Group | $h_i(t \mid \mathbf{x}_i, \beta, h_0)$ |
    |---|---|
    | Control | $h_0(t)$ |
    | *S.*T treated one time | $h_0(t)\, \exp\{\beta_{ST1}\}$ |
    | *S.*T treated three times | $h_0(t)\, \exp\{\beta_{ST3}\}$ |
- Parameters of the model: $\beta_{ST1}$, $\beta_{ST3}$ and the ones of the baseline hazard function.
  - $\pi(h_0, \beta) = \pi(h_0)\, \pi(\beta)$,
  - $\pi(\beta) = \prod_{j=1}^{3} \mathrm{N}(0,\, 0.01)$,
  - $\pi(h_0)$ in terms of a **Weibull** distribution or as a **piecewise function**.

# Statistical analysis: Model specification, II

1. **Weibull distribution**
   $\alpha \sim \text{Ga}(1, 0.001)$, $\lambda \sim \text{Ga}(0.001, 0.001)$
2. **Piecewise function**
   2.1 **Independent** $\lambda$'s

   $\lambda_k \sim \text{Ga}(0.01, 0.01)\,, k = 1, \ldots, g.$

   2.2 $\lambda$'s correlated through a discrete-time **martingale process**.

   $\lambda_1 \sim \text{Ga}(0.01, 0.01)$
   $\lambda_k \mid \lambda_1, \ldots, \lambda_{k-1} \sim \text{Ga}(0.01, 0.01/\lambda_{k-1})\,, k = 2, \ldots, g.$

   2.3 **Logarithmic transformation** of $\lambda$'s correlated through a discrete-time **martingale process**.

   $\log(\lambda_k) = c_k$
   $c_1 \sim \text{N}(0, \tau)$
   $c_k \mid c_1, \ldots, c_{k-1} \sim \text{N}(c_{k-1}, \tau)\,, k = 2, \ldots, g.$
   $\tau \sim \text{Ga}(0.01, 0.01)$

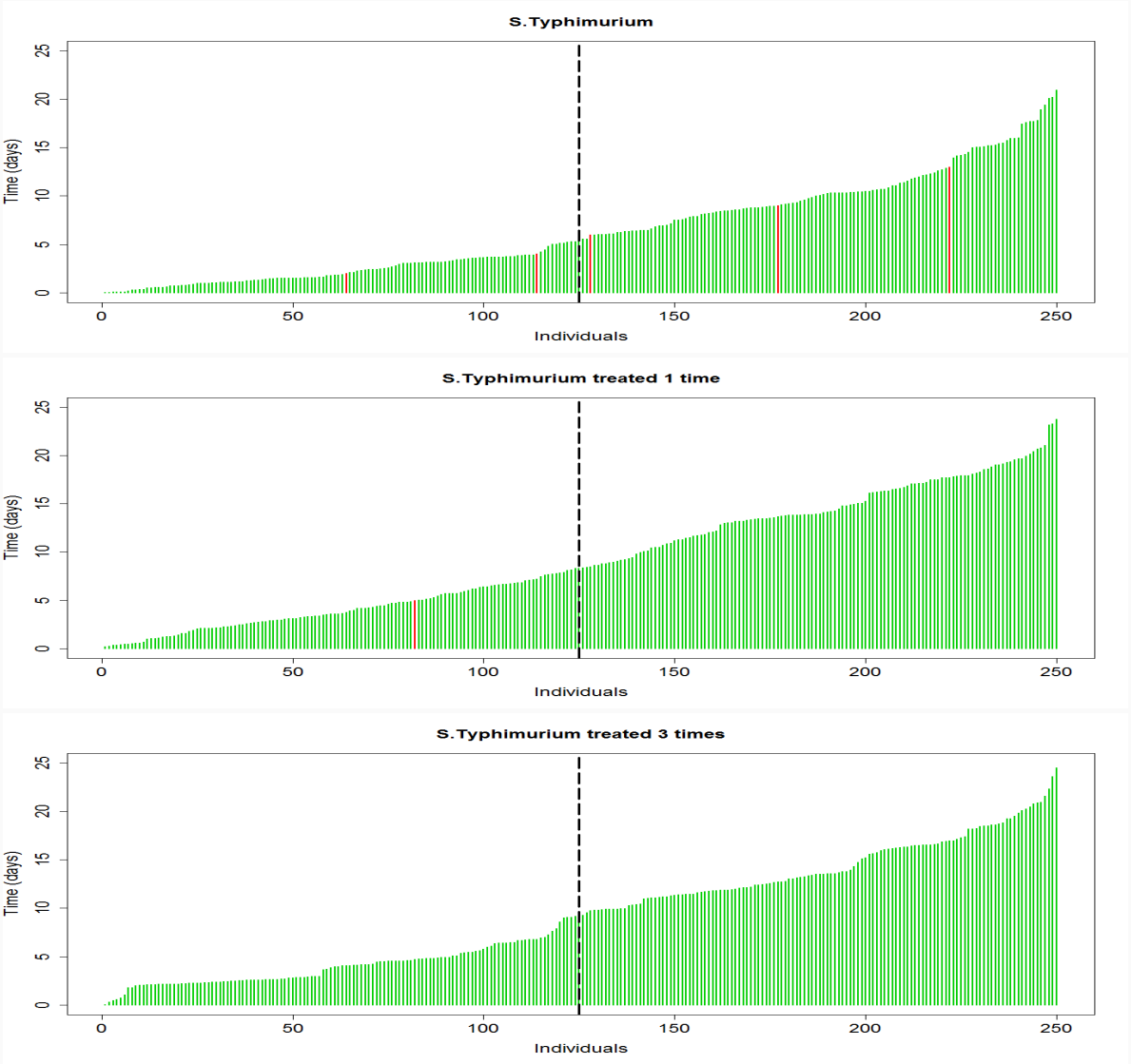
   2.4 $\lambda$'s correlated through a **Gamma process** prior.

   $w_0 = 0.01; \eta_0 = 0.05$
   $\lambda_1 \mid w_0 \sim \text{Ga}(w_0\, 10000, 10000)$
   $\lambda_k \mid w_0, \eta_0 \sim \text{Ga}(w_0\, \eta_0([t_{k+1} - t_k]), w_0([t_{k+1} - t_k]))\,, k = 2, \ldots, g.$

# Results

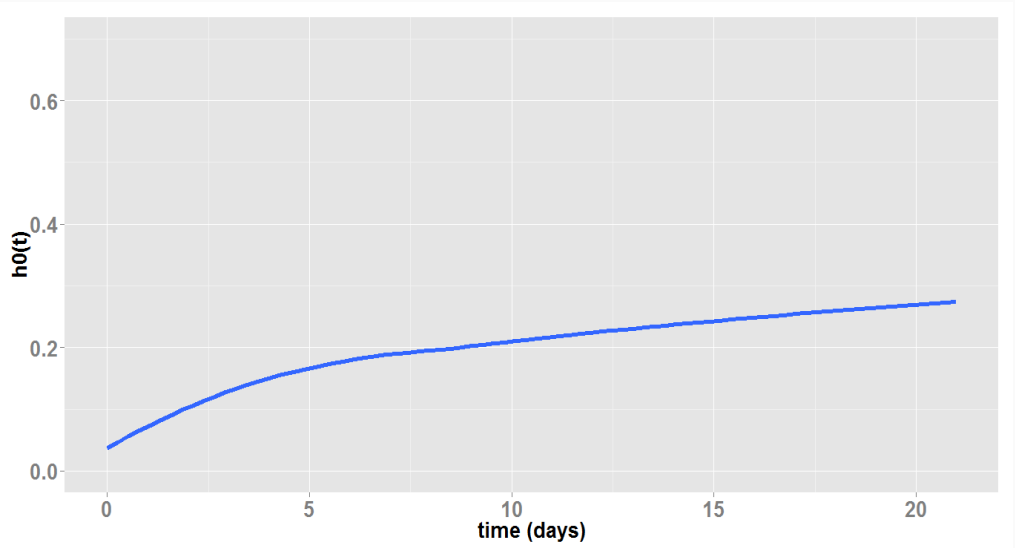
# Data description

## Survival time description

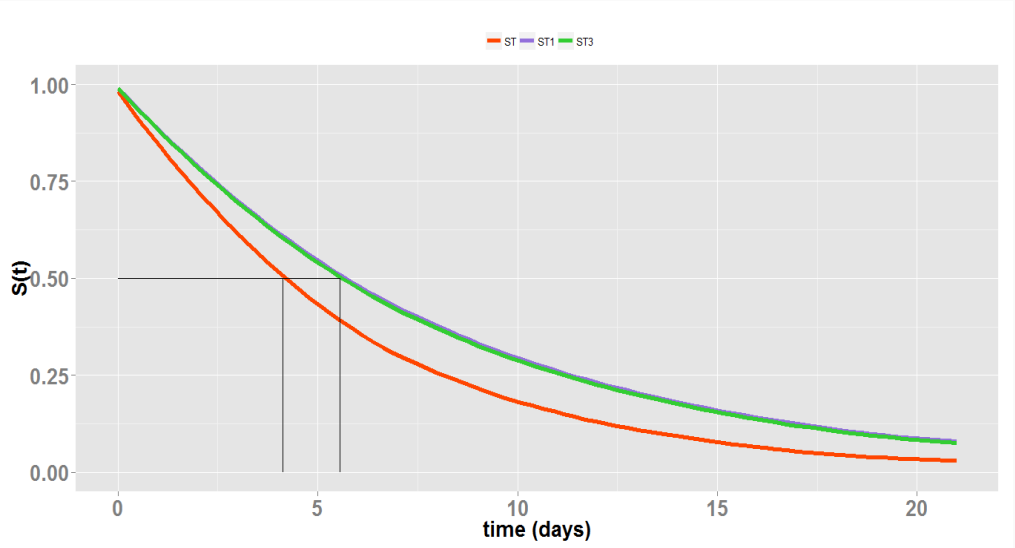


# Weibull model

- Median of the posterior distribution for the hazard baseline risk function distribution $\pi(h_0(t) \mid \mathcal{D})$.

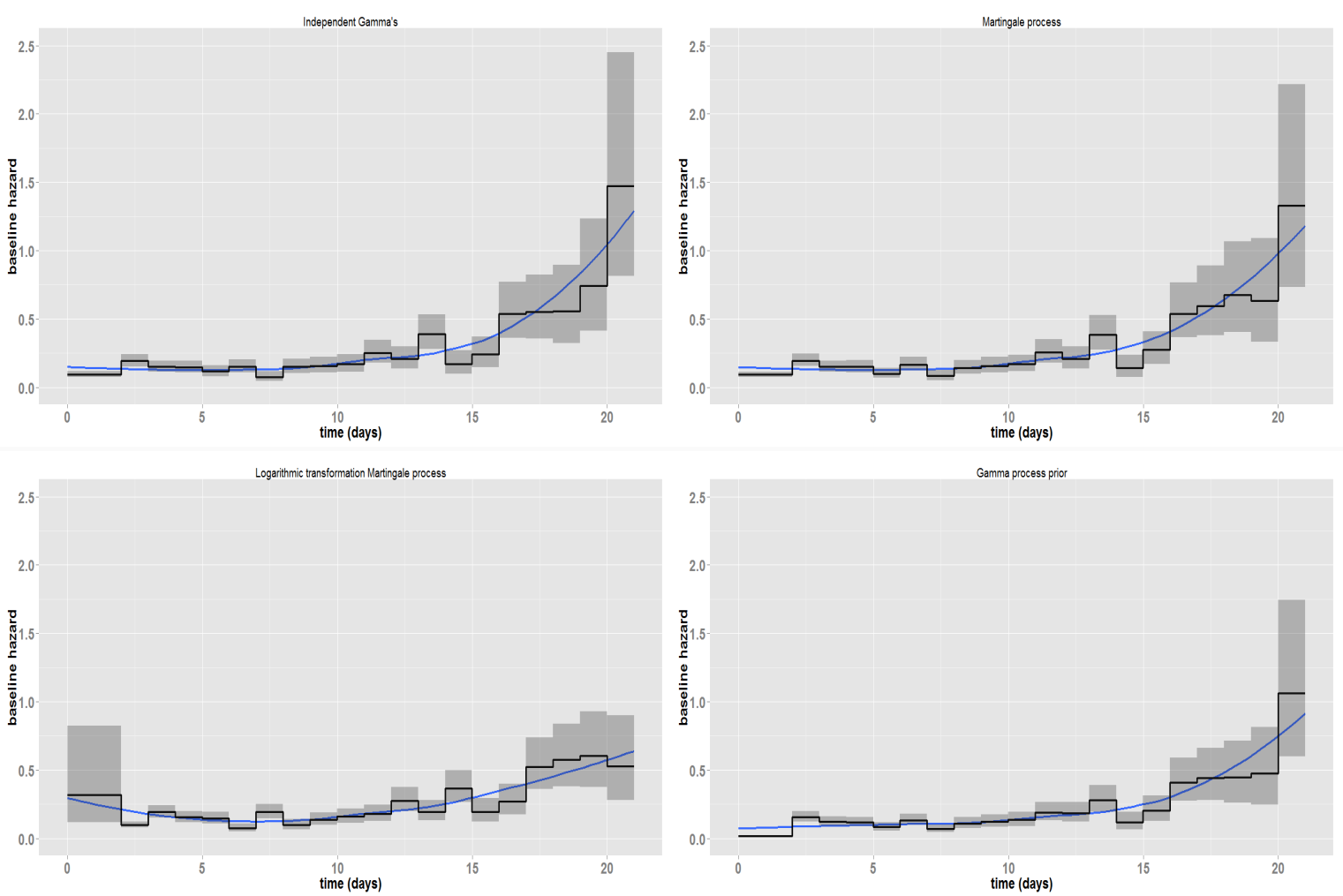



- Posterior median of the survival function distribution, $\pi(S(t \mid h_0, \beta) \mid \mathcal{D})$, for the three treatments.



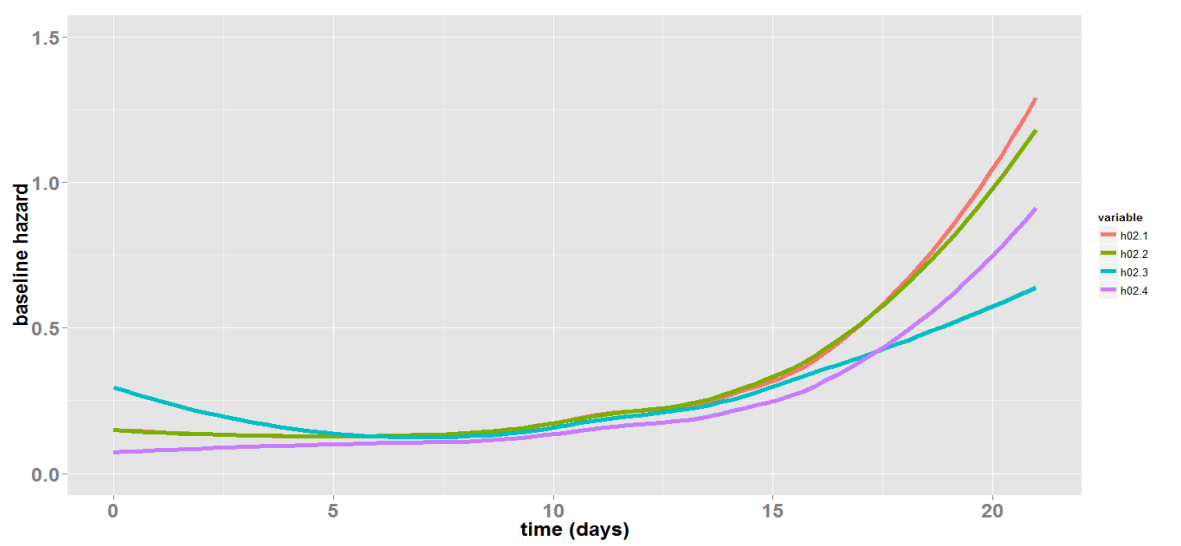
# Piecewise function models, I.

- Median, smoothing median and 95% credible interval for the posterior distribution for the hazard baseline risk function $\pi(h_0(t) \mid \mathcal{D})$.



# Piecewise function models, II.

- Smoothing median posterior distribution for the hazard baseline risk function $\pi(h_0(t) \mid \mathcal{D})$

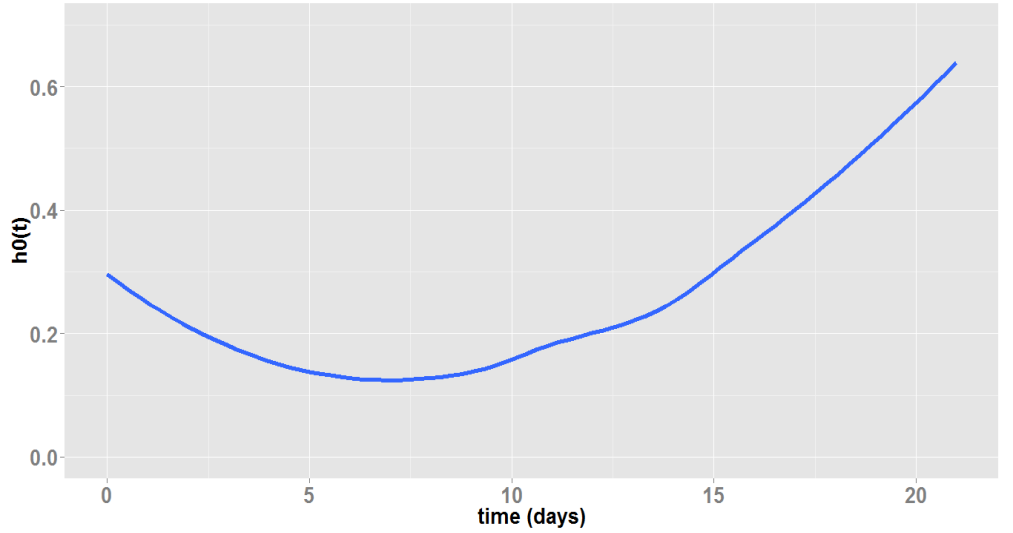



- Model selection:

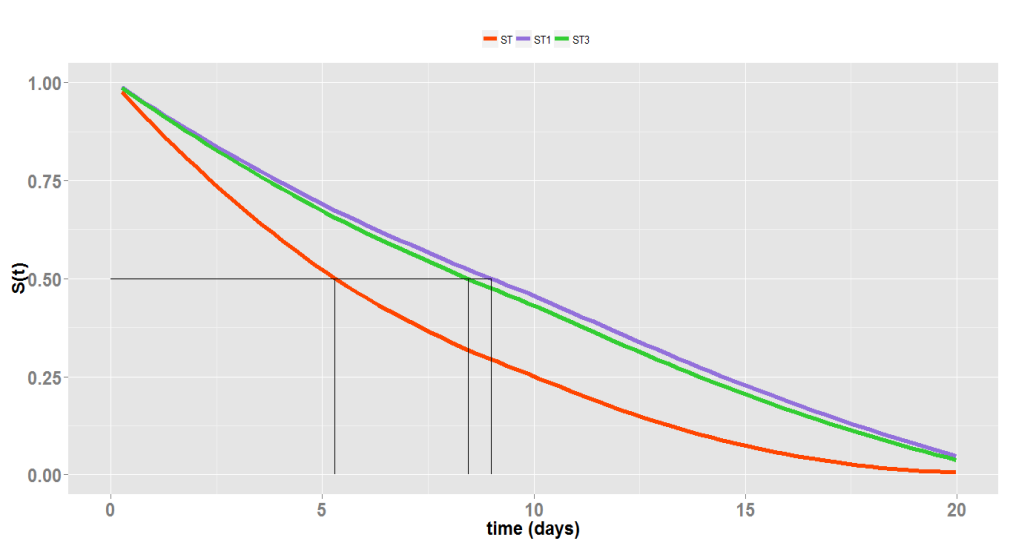
| DIC | $h_{02.1}$ | $h_{02.2}$ | $h_{02.3}$ | $h_{02.4}$ |
|---|---|---|---|---|
| | 5797.36 | 5746.11 | **5681.83** | 5924.94 |

# Piecewise model

- Posterior median for the hazard baseline risk function distribution $\pi(h_0(t) \mid \mathcal{D})$

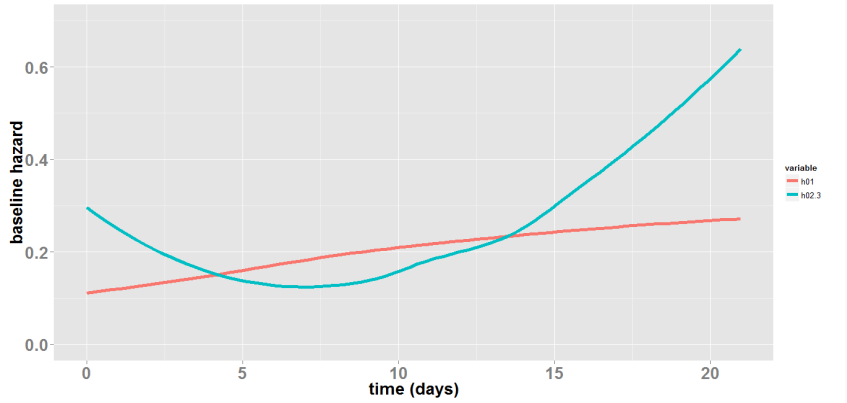



- Posterior median of the survival function distribution, $\pi(S(t \mid h_0, \beta) \mid \mathcal{D})$, for the three treatments.

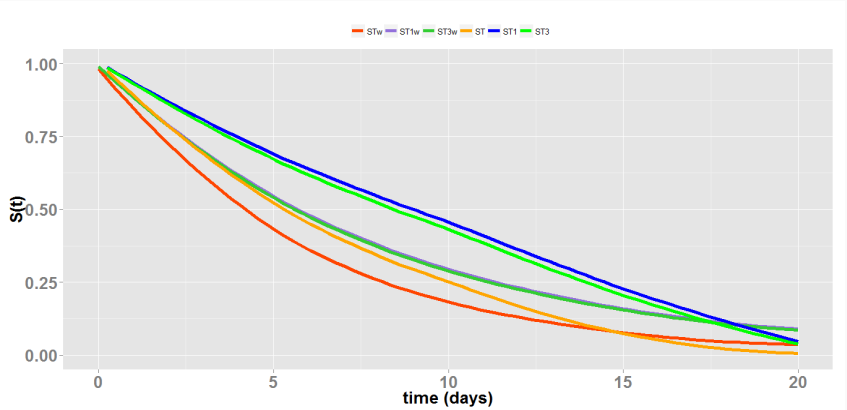


# Model comparison

- Posterior median for the hazard baseline risk function distribution $\pi(h_0(t) \mid \mathcal{D})$




- Posterior median of the survival function distribution, $\pi(S(t \mid h_0, \beta) \mid \mathcal{D})$, for the three treatments.




- DIC:

| Weibull | Piecewise |
|---|---|
| 4553.4 | 5681.83 |

# Conclusions and Future Research

# Conclusions and Future Research

- **Conclusions:**
  - Treated *S.* Typhimurium was less virulent than non-treated S. Typhimurium.
  - Weak effect of frequency treatment.
  - Piecewise models are heavily sensitive to prior specification for the baseline hazard model.
  - Evident differences between Weibull and Piecewise model.
  - Piecewise model based on Martingala logarithmic transformation seems to perform better.
- **Future Research:**
  - Implement *semiparametric* models for interval censored data.
  - Study different model selection criteria.
  - Study sensitivy to number of intervals in the case of piecewise models.
  - Include hierarchical structure in the models to correlate observations.
  - And much more....

# Questions?

# References I



Christensen, R., Johnson, W.O., Branscum, A. J., Hanson, T.E.
***Bayesian Ideas and Data Analysis: An Introduction to Scientists and Statisticians.***
Chapman and Hall/CRC, 2011.


Gómez, G., Calle, M. L., and Oller, R.
**Frequentist and Bayesian approaches for interval-censored data.**
*Statistical Papers*, 45(2), 139–173, 2004.


Ibrahim, J.G., Chen, M.H., and Sinha, D.
***Bayesian Survival Analysis.***
Springer, 2005.

# References II



Klein, J.P. and Moeschberger, M.L.
***Survival Analysis: Techniques for Censored and Truncated Data.*** **New York:**
Springer, 2003.


Sahu, S. K., Dey, D. K., Aslanidou, H., and Sinha, D.
**A Weibull regression model with gamma frailties for multivariate survival data.**
*Lifetime data analysis* , 3(2), 123–137, 1997.