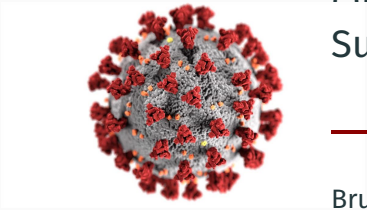


Mixture Cure Rate Models for the Analysis of Survival Times in the COVID-19 Scenario

Bruna Wundervald, Rafael De Andrade Moral, Andrew Parnell, and the Hamilton Institute COVID-19 research group

Maynooth University
National University of Ireland Maynooth

Summary

- Introduction
- Methods
  - Mixture Cure Rate Models
  - GAMLSS Models
- Data
- Results
- Conclusions and Final Remarks

Introduction

- The pandemic of the new coronavirus has become an enormous health challenge worldwide
- The COVID-19 infection can be asymptomatic, but a significant proportion of the cases require hospitalization, including intensive care unit admission
- A few factors have been linked to the risk of more severe disease and death, such as age and sex

Survival modeling : helps us understand the correlation between human and environmental factors (age, sex) and the survival probabilities of each person

Methods

- Choice: motivated by the imbalance present in the data (deaths versus censored observations)
- Let T be the time to some specific event (death, for example), treated as a non-negative random variable
- In survival analysis, we commonly the survival function $S(t) = P(T > t)$, that is nothing more than $S(t) = 1 - F(t)$, where $F(t)$ is the cumulative distribution function of T

Mixture Cure Rate Models

- Two sub-populations: one with a standard failure rate (**cured fraction**) and one with the same rate **added** with a subject-specific parametric failure rate
- The mixture survival function has the form

$$S_{pop}(t|\mathbf{x}, \mathbf{z}, \mathbf{n}) = (1 - \pi(\mathbf{z}))S(t|\mathbf{x}, \mathbf{n} = 1) + \pi(\mathbf{z}),$$

where $\pi(\mathbf{z})$ is the cured proportion, $S(t|\mathbf{x}, \mathbf{n} = 1)$ is the survival function of susceptible (uncured) individuals, and $\mathbf{x}$ and $\mathbf{z}$ are the covariates associated with π and S, respectively.

GAMLSS Models

- GAMLSS: generalized additive models for location, scale, and shape [2]
- Semi-parametric *Weibullcr* [1] distribution, for which the survival function will be

$$S_{pop}(t|\mu,\sigma,\nu) = \nu + (1-\nu)\exp\left\{-\left[\frac{t\Gamma(1/\sigma+1)}{\mu}\right]^{\sigma}\right\}.$$

- The parameters $\theta = (\mu,\sigma,\nu) =$ (location, scale, shape) are estimated through a regression model that allows for complex regression structures

Data

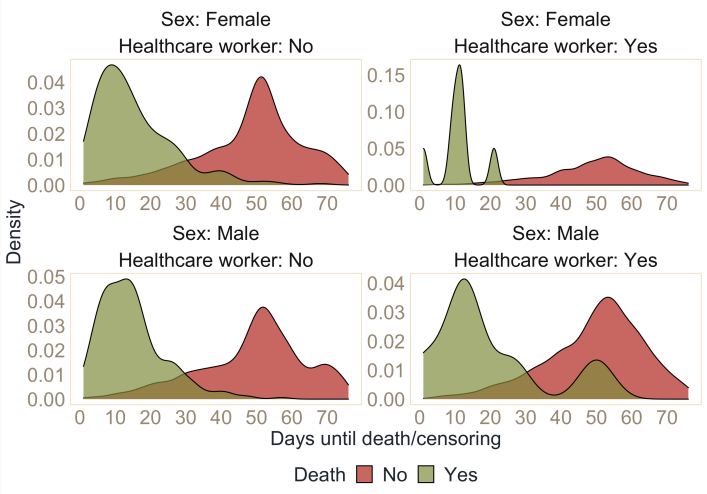
- Time until death or the end of the study (censoring)

Explanatory variables:

- Age
- Sex
- ICU status (yes or no)
- Healthcare worker status (yes or no)

Note: the results presented here are from 'faked data', because the true Irish cases data is confidential

Data


Sex: Female
Healthcare worker: No
Sex: Female
Healthcare worker: Yes
Sex: Male
Healthcare worker: No
Sex: Male
Healthcare worker: Yes
Density
0.04
0.03
0.02
0.01
0.00
0.15
0.10
0.05
0.05
0 10 20 30 40 50 60 70
Days until death/censoring
Death No Yes

Results

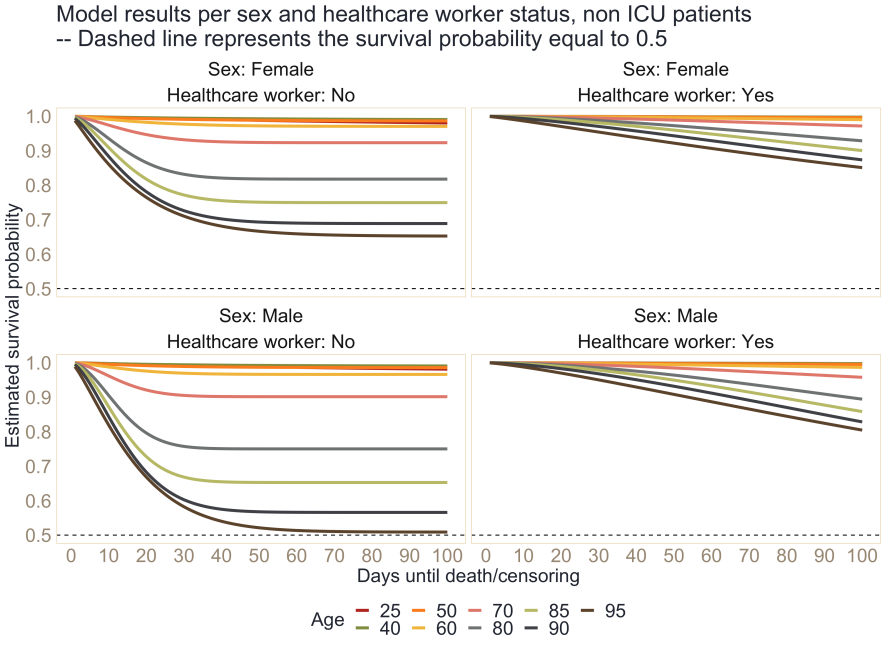
- Best model: selected by minimising the AIC
- Final model: does not use the same covariates in all parameters, and includes polynomial and interaction terms

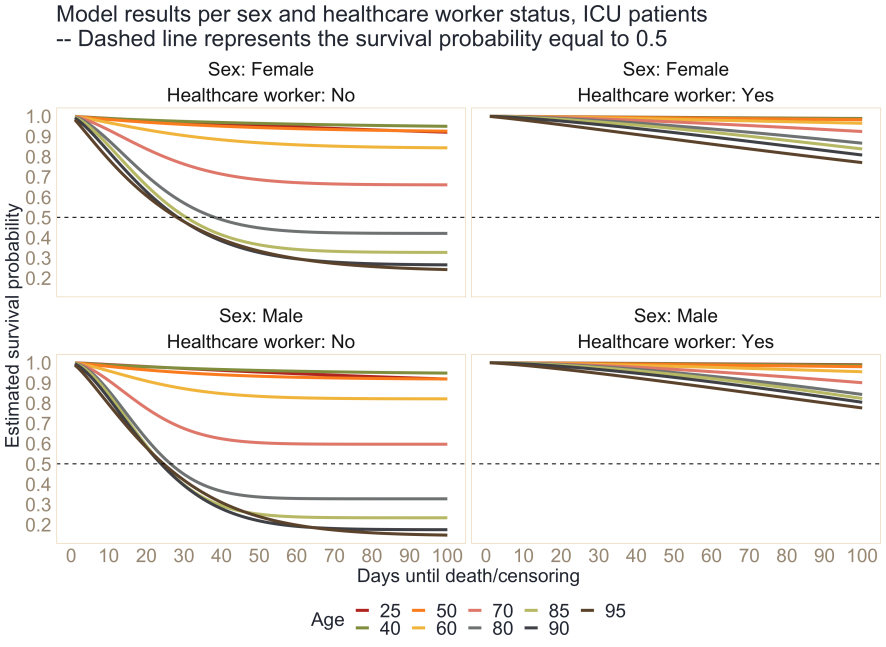
$$\log(\mu_i) = \beta_{01} + \beta_{11}\text{age} + \beta_{21}\text{age}^2 + \beta_{31}\text{sex} + \beta_{41}\text{HW} + \beta_{51}\text{ICU} + \beta_{61}(\text{age} \times \text{sex}),$$

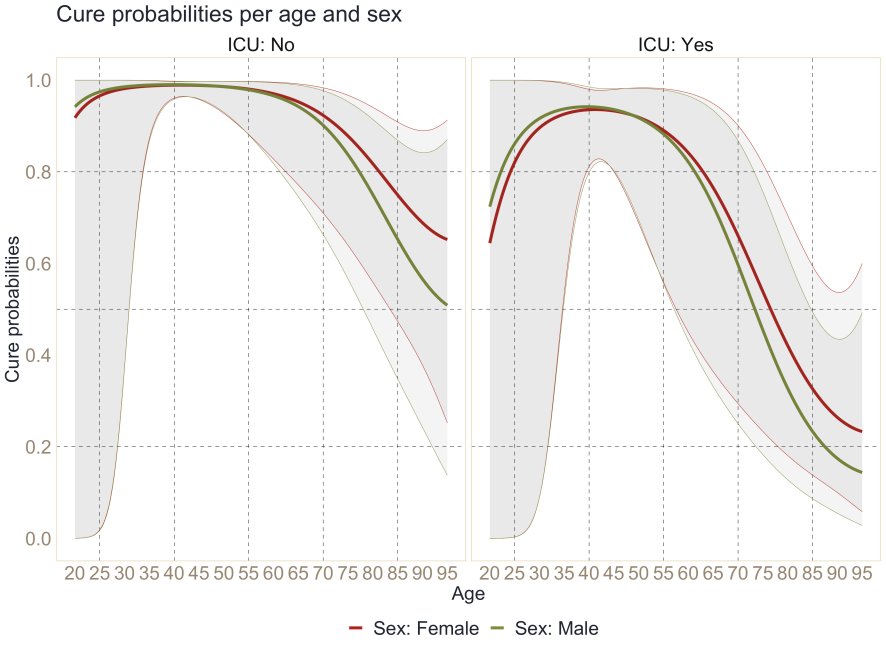
$$\log(\sigma_i) = \beta_{02} + \beta_{12}\text{sex} + \beta_{22}\text{age} + \beta_{32}\text{age}^2 + \beta_{42}\text{age}^3,$$

$$\text{logit}(\nu_i) = \beta_{03} + \beta_{13}\text{age} + \beta_{23}\text{age}^2 + \beta_{33}\text{age}^3 + \beta_{43}\text{ICU} + \beta_{53}\text{sex} + \beta_{63}(\text{age} \times \text{sex}),$$

where $\text{logit}(x) = log(x/(1+x))$.


Model results per sex and healthcare worker status, non ICU patients
-- Dashed line represents the survival probability equal to 0.5
Sex: Female
Healthcare worker: No
Sex: Female
Healthcare worker: Yes
Sex: Male
Healthcare worker: No
Sex: Male
Healthcare worker: Yes
Estimated survival probability
1.0
0.9
0.8
0.7
0.6
0.5
0 10 20 30 40 50 60 70 80 90 100
Days until death/censoring
Age
25 40 50 60 70 80 85 90 95


Model results per sex and healthcare worker status, ICU patients
-- Dashed line represents the survival probability equal to 0.5
Sex: Female
Healthcare worker: No
Sex: Female
Healthcare worker: Yes
Sex: Male
Healthcare worker: No
Sex: Male
Healthcare worker: Yes
Estimated survival probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0 10 20 30 40 50 60 70 80 90 100
Days until death/censoring
Age
25 40 50 60 70 80 85 90 95


Cure probabilities per age and sex
ICU: No
ICU: Yes
Cure probabilities
1.0
0.8
0.6
0.4
0.2
0.0
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Age
Sex: Female
Sex: Male

Conclusions and Final Remarks

- ➔ Age and sex seem to play the most important roles in the decrease of survival chances
- ➔ ICU patients, particularly males, have the lowest survival probabilities
- ➔ Healthcare workers appear to have a low mortality for COVID-19

- ➔ Many factors were not captured here, such as time and spatially variant dynamics and other influential covariates

Acknowledgements

This work was supported by Science Foundation Ireland and their emergency COVID-19 funding portal via grant number 20/COV/0081

Science Foundation Ireland sfi For what's next

https://www.hamilton.ie/covid19/

References

- Thiago G Ramires et al. “A new semiparametric Weibull cure rate model: fitting different behaviors within GAMLSS”. In: *Journal of Applied Statistics* 46.15 (2019), pp. 2744–2760.
- D Mikis Stasinopoulos, Robert A Rigby, et al. “Generalized additive models for location scale and shape (GAMLSS) in R”. In: *Journal of Statistical Software* 23.7 (2007), pp. 1–46.

Thanks!