BAYESIAN MODELING IN THE SURVIVAL ANALYSIS OF PATIENTS WITH MULTIPLE MYELOMA WITH EMPHASIS ON MISSING DATA ANALYSIS.

Biana Andrea González Artunduaga
https://orcid.org/0000-0001-8250-519X, Medica 1, Especialización en Salud Pública 1, Especialización en Epidemiología Clínica 2., Especialización en Medicina Nuclear 3., Estudiante PhD doctorado en Oncología.
eagonzaleza@unal.edu.co
Dauris Lineth Mejía Pérez
Mujer Guajira: Médica Cirujana, Ginecóloga y Obstetra Profesora del Departamento de Ginecología y Obstetricia Universidad Nacional de Colombia, Magíster en Salud Pública, Magíster en Política Pública, Universidad de los Andes https://orcid.org/0000-0002-5087-7885
Nelson Alirio Cruz Gutiérrez
https://orcid.org/0000-0002-7370-5111, Estadístico1, Maestría en estadística1, PhD en ciencias en estadística 1. Visiting Professor 2. Nelson-alirio.cruz@uib.cat, https://orcid.org/0000-0002-7370-5111
Kenny Mauricio Gálvez Cárdenas, MD
Hematologo Hospital Pablo Tobon Uribe, Medellín, Colombia Especialista en Hematología, Fundación Universitaria de Ciencias de la Salud, Especialista en Medicina Interna, Universidad Caldas, MD, Universidad Caldas, https://orcid.org/0000-0003-1192-9053
1. Universidad Nacional de Colombia.
2. Fundación Universitaria de Ciencias de la Salud.
3. Fundación Universitaria Sanitas
4. Universitat de les Illes Balears Palma de Mallorca.Departamento de Matematicas e Informatica. Spain.

BACKGROUND

In the survival prognosis of patients with multiple myeloma, there are staging models based on Cox survival models. Current staging models do not include patients with missing data, although there are multiple investigations where the problem of missing information is very common. This situation causes results that do not represent the reality of the populations, generating biases in the true results of the research and excluding patients both in training and in the application of the model to new patients.

METHODS

To solve this problem, a survival model based on Bayesian statistics was built that not only allowed us to impute the missing data of patients with multiple Myeloma using statistical methods but also allowed us to predict new patients with missing data. This model was built and tuned using data from multiple myeloma patients from a public database from a study conducted in Arkansas.

MAIN OBJECTIVE

Compare the accuracy of Bayesian models and Cox regression in estimating survival in multiple myeloma patients using missing data from the MAQC-II study (2002-2006).

Database:
Data from the MAQC-II project (2000-2006), which includes clinical, genomic, and MRI data on multiple myeloma, providing a robust foundation for survival analysis.

STATISTICAL METHODS

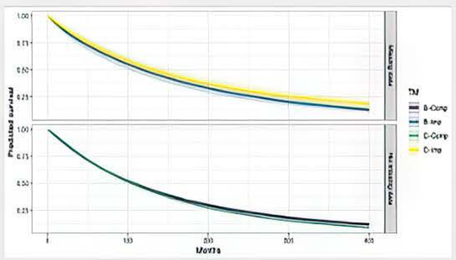
- Survival package for Cox proportional hazards models, facilitating the estimation and comparison of survival curves.
- Bayesian Model: bayesSurv package for Bayesian regressions, handling censored data and employing sampling techniques for variable selection.

ADDITIONAL ANALYSIS

- Evaluation of model fit with the GH function and concordance index (discSurv).
- Calculation of time-dependent ROC curves to assess predictive accuracy..

RESULTS

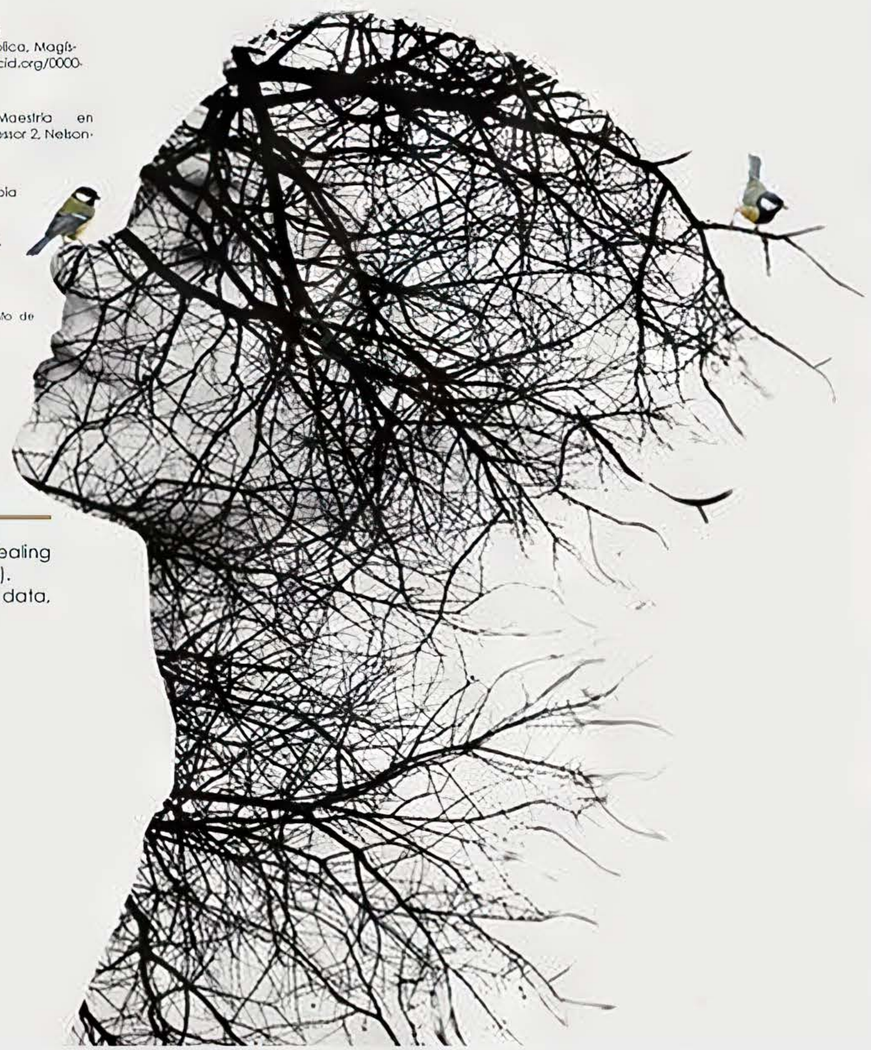
Figure A: Overall Survival (OS) Curves



DATA IMPUTATION

Bayesian: MCMC and SMC methods for dealing with censored data (right and left censoring).
Cox: k-NN algorithm for imputing mixed data, optimized through cross-validation.

Primary Outcome:
Assessment of overall survival (OS), classifying patients into groups based on survival of more or less than two years to understand the overall impact of the disease.



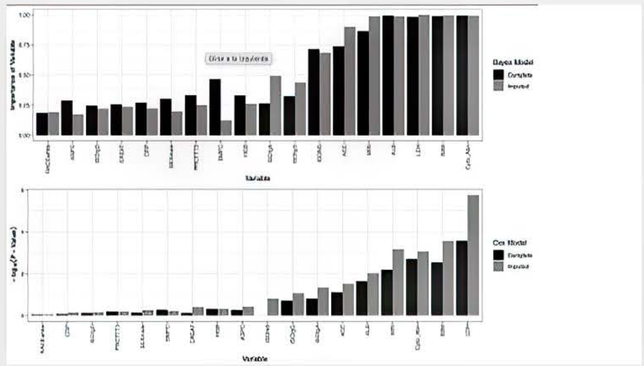
CURVES WITHOUT DATA IMPUTATION

- Cox Model (C-Comp): OS curve for patients with no missing data.
- Bayesian Model (B-Comp): OS curve for patients with no missing data.
- Comparison: No significant differences observed between the OS curves of the two models.

CURVES WITH DATA IMPUTATION

- Cox Model (C-Imp): OS curve for patients with missing data imputed.
- Bayesian Model (B-Imp): OS curve for patients with missing data imputed.
- Comparison: No significant differences observed between the OS curves of the two models with imputed data.

Statistical Significance of Variables for OS Function of Cox Models and Importance of Variables for OS Bayesian Models with Complete and Imputed Data.



This illustrates the absence of difference in the SG function between the Cox-Comp and Cox-Impu models, and similarly, the lower bars show no difference in the SG function between the B-Comp and B-Impu models.

MODEL EVALUATION

Findings: Concordance indices, GH, D, and AUC curves demonstrate notable similarity among the four analyzed models, both with complete and imputed data.

Modelo	Indice de concordancia	Indice GH	AUC
Cox-Comp	0,713	0,6579	71.69%
Bayes-Comp	0,7324	0,6753	76.20%
Cox-Imp	0,7355	0,6711	76.83
Bayes-Imp	0,7326	0,6813	75.7%

CONCLUSIONS

A model that performs as well as Cox models in patients with complete data and maintains this performance in patients with missing data is achieved. It also allows prediction of new patients who are diagnosed with multiple myeloma without the restriction that their complete clinical or genetic data is needed

BIBLIOGRAFIA

Yu B. A Bayesian MCMC approach to survival analysis with doubly-censored data. Comput Stat Data Anal [Internet]. 2010 Aug 8 [cited 2024 Jul 20];54(8):1921. Available from: /pmc/articles/PMC2877214/
Survival Analysis with R · R Views [Internet]. [cited 2024 Jul 29]. Available from: https://rR Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.