

Quantifying physician's bias to terminate resuscitation : the TERMINATOR Study

Table des matières

| | |
|---|---|
| Supplementary appendix 1: detailed methodology | 2 |
| Supplementary appendix 2: Model performance and evaluation | 3 |
| Validity analysis..... | 3 |
| Performance analysis | 4 |
| Supplementary appendix 3: Categorization of continuous variables | 5 |
| Supplementary appendix 4: Analysis without including OHCA treated by a doctor with fewer than three appearances in the database. | 6 |
| Motivation and method..... | 6 |
| Performance and validity..... | 6 |
| Results | 6 |
| Supplementary appendix 5: Non gender specific analysis | 7 |
| Motivation and method..... | 7 |
| Performance and validity..... | 7 |
| Results | 7 |
| Supplementary appendix 6: Source code..... | 8 |
| References..... | 8 |

Supplementary appendix 1: detailed methodology

GLMMs are extensions of GLMs where another source of sample variability in the form of one or more “random effects” is added. The random effect framework consists in treating discrete covariates/grouping variables (physician’s identity in our case) with a pooling technique that assumes that the effects for the different grouping variable realizations are drawn from a known distribution (e.g. normal distribution). This parametrization allowed for better statistical efficiency through the inference of the parameters of the assumed distribution.

In our study, we first used a GLM with a logit link function (also known as a standard logistic regression) to model the binary outcome (termination or continuation of resuscitation). For this GLM model, all parameters chosen (see [Data selection](#)) for the analysis have been included (they constitute fixed factors of both models), apart from the numerical identification of the doctor.

Second, we extended this logistic regression model by adding a term to account for the numerical identification of the physician responsible for OHCA care (“doctor effect”). As stated previously, we modeled this “doctor effect” as a random effect, and more precisely as a random intercept term. The addition of this random effect term to our fixed effect logistic regression model constitutes a mixed model (GLMM). This random intercept represents the physician's individual propensity to carry on or terminate resuscitation compared to the average physician (or, more accurately, to the average decision over all physicians against the exact same patient). Modelling this “doctor effect” as a random effect allowed us to be more statistically efficient to infer grouping variables for each physician, and thus reduce the risk of overfitting compared to a fixed effect. This reduced risk of overfitting was of primary importance as we were trying to infer conclusions on clinicians’ behaviours from a small sample of doctors, especially as many of the physicians in the database only took over 1 or 2 cardiac arrests.

The null hypothesis (H0)—all physicians have the same propensity to terminate or carry-on resuscitation—was that the “doctor effect” was not statistically relevant. We tested this hypothesis by doing a likelihood ratio test between our two models. The explanation about the computation of odds ratio and their confidence intervals is explained in the paper in the [analysis](#) section.

Supplementary appendix 2: Model performance and evaluation

As said in the [analysis section](#), three controls have been carried out assessing the validity and two others assessing the performance of chosen models (GLMM versus GLM). These analysis have been suggested in past articles discussing proper ways of evaluating GLMMs.¹⁻³ Here we show figures that relate to these analysis.

The choice of treating continuous variables as categorical is explained in a distinct appendix. Other validity and performances analysis (comparing the base model to a model without gender or to a model considering only arrest treated by a doctor with more than three arrests in the dataset) are presented in dedicated appendixes.

Validity analysis

Fixed variable independence

The correlation between each variable was analyzed through the covariance matrix given by the “lme4” package. It is to be noted that this analysis involved only linear correlations but given the dataset (mostly binary data) we didn’t expect other correlations to be relevant. The results showed that the most relevant correlations were between rhythm and ROSC, witness and low flow, gender and age, witness and age, medical history and age, cardiac medical history, and other medical issues. All those correlations were moderate and of no surprise, besides, the robustness of GLMM to this type off correlations allows us to expect the model to be interpretable with few reserves concerning this aspect.⁵

Distribution of random effect (“Doctor effect”)

The repartition of the random effects (“doctor effect”) has been verified graphically using sjPlot (version 2.8.9) in the QQ-plot attached (Figure 3). This QQ-plot shows random effects (which are intercept only in our model) represented by blue dots against the blue line representing the expected normal distribution. No major derivation from expected distribution appeared on visual analysis.

Independence of random effect to other variables (exogeneity)

The exogeneity (the independence of the random effect “doctor” to other variables) has been verified by calculating the odds ratio of the random effect with different model omitting one or more variables. The

odds ratio of the random effect changes were minor no matter the model choice. The results are shown in Table 3. 95% confidence intervals were computed with a semi-parametric bootstrap, limited to 100 iterations to preserve computing power.

Performance analysis

Pseudo-R² of the GLMM model

The performance of the model has been evaluated in two ways. First, the pseudo-R² of the model was calculated, giving a “goodness of fit” value. We calculated a pseudo-R² for the whole model (accounting for the “doctor effect”). The result was 62 % (using partR2) or 68.5% (using MuMIn), and this result can be interpreted as the “fraction of variance” explained by the model.

Comparative performance against concurrent models

The second way to analyze model performance was to compare its performance to competing GLMM and GLM models, using pseudo-R² (computed through MuMIn version 1.43.17 since part-R² cannot compute pseudo-R² for GLM models) and Akaike Information Criterion (AIC), computed through lme4) to assess for the best model. The results are shown in Table 4 for pseudo-R² (conditional R²) and AIC (lower is better), both in favor of selecting the GLMM when compared to its concurrent models.

Supplementary appendix 3: Categorization of continuous variables

Our study involves 3 continuous explanatory covariates as fixed effects. In this appendix we detail results obtained assuming a linear effect, as well as results obtained adding a quadratic effect term for each of them in Table 5. The latter exhibits coefficients significantly different from 0 for the three quadratic terms and demonstrates the need to account for nonlinear effects of those three variables, in turn justifying discretization of the variables to account for higher order nonlinear effects. We note that, while estimates for the linear terms for the three variables vary greatly between the linear and quadratic models, the estimated coefficients for the other covariates (including the doctor effect) remain very consistent between the linear, the quadratic and the discretized model discussed in the main analysis. We chose the different cut points for the three continuous variables based on both clinical relevance and available sample sizes. We chose the following:

- Age : 65, 75 and 85 years ;
- No flow : 1, 5, 10 and 20 minutes ;
- Low-flow : 10, 20, 40 minutes.

Supplementary appendix 4: Analysis without including OHCA treated by a doctor with fewer than three appearances in the database.

Motivation and method

In our primary analysis, we decided to keep doctors who oversaw few OHCA in the database. This choice is debatable and could lead to an overestimation of variance of random effects.

We reran the study including only OHCA where the doctor had more than three OHCA in the database. All analysis can be reproduced by activating the “ExcludeDrWithFewOHCA” option of the code given in appendix.

This resulted in the exclusion of 1446 others OHCA (from the 5144 included for the main analysis of this paper) for a remaining total of 3698 OHCA used for the subsequent analysis. In total, 42 doctors oversaw all those arrests (down from 173 in the paper), for a median of 84.5 (IQ 53-118) OHCA per doctor.

Performance and validity

Performance and validity analysis led to comparable results to the main analysis.

Results

Results were comparable with the main analysis, with a significant ($p < 0.001$) “doctor effect”, with an odds ratio at 2.24 [1.97-2.53] for a doctor one SD above the mean (2.48 [2.13-2.94] in the main analysis).

Supplementary appendix 5: Non gender specific analysis

Motivation and method

As we conducted this study, we had to make choices about the different factors that should be considered in the model (see [data selection](#)). We chose to include gender among decisive factors that explain the termination of resuscitation. It is one of the few directly accessible variables known by the doctor while tacking his decision, that may reflect other not so straightforwardly available factors such as comorbidities (since at the same age, women tend to have fewer comorbidities). The inclusion of this parameter is questionable, since some study indicates that it is a prognostic factor for reasons mentioned above, but it is also counter-intuitive since it is not usually part of medical thinking to consider gender to decide for TOR.

While evaluating the performance of the model, the best model according to AIC was the one that included gender, but without any impact on pseudo-R², at 68.5% (using MuMIn version 1.43.17) or 62% (using partR2) for both models.

We decided to conduct the same analysis, using the same models but excluding gender as a factor, which can be achieved by using the code given in related appendix and setting the variable “ExcludeSex” to “true”.

Performance and validity

Same validity and performance analyses were carried on with similar results for this model and the GLMM from the main analysis.

Results

The existence of the “doctor effect” remained statistically significant ($p < 0.001$). The “doctor effect” odds ratio was 2.26 [1.94;2.69] for a doctor one SD above the mean, close to the value found in main analysis (2.48 [2.13-2.94]).

Supplementary appendix 6: Source code.

This appendix is given as a separate file “Appendix – Source Code”.

References

- 1 Bolker BM, Brooks ME, Clark CJ, *et al.* Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* 2009; **24**: 127–35.
- 2 Thiele J, Markussen B. Potential of GLMM in modelling invasive spread. *CAB Reviews Perspectives in Agriculture Veterinary Science Nutrition and Natural Resources* 2012; **7**: 1–10.
- 3 Harrison XA, Donaldson L, Correa-Cano ME, *et al.* A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 2018; **6**: e4794.
- 4 Hartig F. DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. <https://cran.r-project.org/web/packages/DHARMA/vignettes/DHARMA.html> (accessed Jan 20, 2022).
- 5 Schielzeth H, Dingemanse NJ, Nakagawa S, *et al.* Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution* 2020; **11**: 1141–52.