Intensive care of the very old: What do they wish for themselves?

Statistical analyses

We report basic descriptive statistics for the survey participants – counts, percentages, means and standard deviations – both for the overall cohort and stratified by ICU admission preference.

To estimate the participants' preferences regarding life-sustaining treatment, we report the marginal proportion of responses in each category ('wants ICU admission', 'does not want ICU admission', 'does not want to engage in the decision'), along with 95% confidence intervals (CIs). To take into account the dependence between the responses for the three scenarios (e.g., the answer to scenario 1 and 2 for a single patient is more likely to be the same than if the two answers came from two *different* patients), all proportions were calculated using Generalized Estimating Equation (GEE) models with an 'exchangeable' correlation structure.

Sample size calculations done before data collection showed that with responses from 200 respondents, we could estimate the preference proportions with an absolute margin of error less than 7%, based on 95% confidence intervals, which we deemed sufficient. For example, if 50% of the respondents wanted ICU admission, the corresponding 95% confidence interval would be approx. 43%–57%.

To investigate factors associated with the participants' ICU admission preference, we fitted mixed-effects logistic models with 'wanting ICU admission' as the outcome. Explanatory variables included the experimental framing effects (vignette gender and mortality/survival) and demographic and health-related characteristics of the very old respondents. Vignette (there were twenty possible vignettes) was included as a random intercept. All variables included in the model were pre-specified before data collection, except CPS (Comorbidity Polypharmacy Score), as a measure of comorbidity, and health care professional background. These two variables were included post-hoc because we thought they might *potentially* be important predictors of the patient preferences.

Sample size calculations for the logistic models showed that we would only have statistical power to detect large effects. For example, the effect of a binary predictor with a prevalence of 50% (e.g., the survival vs. mortality framing predictor) had to correspond to an odds ratio (OR) of at least 2.33 (outcome 50% vs. 70%) for the power to be 80% or greater (with 200 participants). For continuous predictors, an OR of 1.5 for the event rate at one standard deviation above the mean compared to the event rate at the mean, would be sufficient.

The above calculations were based on simple logistic regression, i.e., assuming that only *one* preference (from one patient scenario) was expressed for each respondent. Initially, we were hoping to use mixed-effects models to include all *three* scenarios for each respondent in the same model, thereby increasing the statistical power, i.e., making it possible to detect lower odds ratios. We would take into account the dependence between the three scenarios responses from the same respondent by having a random intercept for 'respondent' (in addition to a random intercept for 'scenarios') in the model.

However, this model gave confusing coefficient estimates. (Basically, the OR estimates were unrealistically large, e.g., an estimated OR > 26 for the effect of male respondent gender. The reason was that all male respondents were predicted to have a very small 'respondent' random intercept, which would be compensated by the large OR.) We believe this problem would be reduced or eliminated if we would had a few more scenarios for each respondent, as this would be make it easier to estimate the 'respondent' effects.

While we do think the model that includes three scenarios for each respondent is technically *valid* (e.g., low *P*-values do indicate real effects), we believe the coefficient estimates would be confusing for the reader, so in our main analysis, we only report the results for the *first* scenario. (This was scenario chosen because the response to the first scenario would in theory not be influenced by the responses to the subsequent scenarios, while the response on the second and third scenarios might influenced by the respondent's initial response.) However, we also report the results for similar models for the second and third scenario, and for all three scenarios.

There were some missing data. To increase precision and reduce bias in estimates caused by missing data, we used multiple imputation for the main analysis (scenario 1). In the imputation model, the outcome variable and explanatory variables from the regression model were included along with the variables 'education' (four levels) and 'admitted to hospital within the previous 12 months' ('yes' vs. 'no'), which we thought would be able to predict the values of these variables or the probability of a variable having missing values.

The data were imputed using the mice function from the R 'mice' package, with 50 imputations, 30 iterations and otherwise default options. Depending on the variables imputed, this uses predictive mean matching (for continuous variables), logistic regression ('yes'/'no' variables), or polytomous regression (categorical variables). We report the results both from a complete-case analysis and from the multiple imputation analysis.

To assess the next of kins' ability to predict their family member's preference, we report the overall percentage agreement between the next of kins' proxy preference and the old patient's actual preference (*proxy accuracy*). We also report this stratified by the next of kins' proxy response. Similarly, we also report percentage agreement between the next of kins' proxy preference and their *own* preference (*assumed similarity*), and between the next of kins' own preference and the old patient's actual preference (*true similarity*). Estimates and CIs are as before based on GEE models.

To examine variables associated with 'correct prediction', we looked at the first scenrio and only at responses where the old respondent *had* a preference ('wants ICU admission' or 'does not want ICU admission'), the reasoning being that it is only reasonable to try to predict the preference if the respondent actually *has* a preference. The model fitted was a mixed-effects logistic regression model with vignette as random intercept.

All statistical models were fitted using R version 4.2.3. We used the R packages 'mice' version 3.16.0 (for multiple imputation), 'lme4' version 1.1-34 and 'glmmTMB' version 1.1.8 (for mixed-effects logistic models), and 'geepack' version 1.3.9 (for GEE models). Other analyses were performed using IBM SPSS Statistics for Windows version 29.0 or Microsoft Excel.