Supplemental Material 2

Quality of Life of Patients with Dementia in Acute Hospitals: A non-randomised case-control-study comparing a regular ward to a special care ward with dementia care concept.

Methodological comments

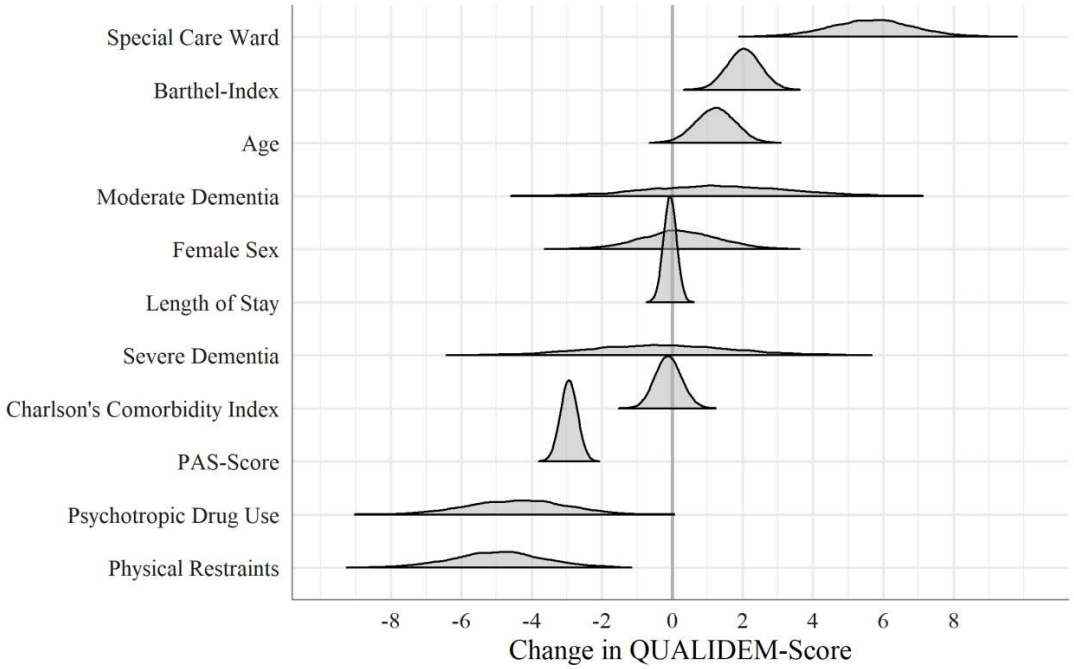
1. Short note on Bayesian Methods

Bayesian methods are probably not very common, and readers may ask "Why use Bayesian regression models?" Gelman et al.[1] give a well summarized answer to this question: "A pragmatic rationale for the use of Bayesian methods is the inherent flexibility introduced by their incorporation of multiple levels of randomness and the resultant ability to combine information from different sources, while incorporating all reasonable sources of uncertainty in inferential summaries. Such methods naturally lead to smoothed estimates in complicated data structures and consequently have the ability to obtain better real-world answers."[1]. Shortly speaking, advantages of Bayesian methods are, for instance, the ability to derive probability statements for every quantity of interest or explicitly incorporate prior knowledge about parameters into the model.

2. Distribution of Posterior Samples from the Regression Model

Contrary to a frequentist approach, there is no unique point estimate in Bayesian analysis. The posterior median minimises the expected absolute error (i.e. the difference of estimates from true values over posterior samples), but there are many other plausible values to describe the association between predictors and outcome, which is called the *distribution of posterior samples*. Credible or Highest Density Intervals are based on quantiles of this distribution and indicate the probability of an estimate being inside this interval. In terms of inference, this allows us to say that with, e.g. 50% probability, the true effect of a *Special Care Ward* on our outcome lies between 4.9 and 6.5. Figure S1 shows the distribution of posterior samples from our regression model.

Figure S1: *Distribution of Posterior Samples from Regression Model*



3. Tabular Summary from the Regression Model

Table S1 provides a summary of the regression model, including the data for two different Highest Density Intervals as well as the ratio of effective number of samples. The ratio should be 1.0 or close to 1.0 and indicates whether the samples were efficient, which means that the results for the related coefficient are reliable. Furthermore, all Rhat values of the models were approximately 1. The Rhat statistic[2] measures the ratio of the average variance of the draws within each chain to the variance of the pooled draws across chains. If the chains have not converged to a common distribution, the Rhat statistic will tend to be greater than one, indicating that more iterations may be required for reliable results.

Table S1: *Predictors of Quality of Life, Regression Coefficients, Bayesian Linear Mixed Model, Posterior Median (+50% and 89% Highest Density Intervals)*

Term	*Estimate*	*SE*	*50% HDI*	*89% HDI*	*Ratio of effective Number of Samples*
(Intercept)	36.4	5.1	33.4 – 40.2	28.1 – 43.9	1.00
Length of Stay	-0.1	0.1	-0.1 – 0.0	-0.2 – 0.1	1.00
Age	1.2	0.5	0.8 – 1.5	0.4 – 2.1	1.00
Moderate Dementia	1.2	2.0	-0.1 – 2.7	-1.8 – 4.6	1.00
Severe Dementia	-0.4	2.0	-1.6 – 1.1	-3.6 – 2.7	1.00
Female	0.2	1.1	-0.5 – 1.0	-1.6 – 1.9	1.00
Barthel Score	2.0	0.5	1.8 – 2.4	1.3 – 2.8	1.00
Physical Restraints (yes)	-4.9	1.2	-5.8 – -4.1	-7.0 – -2.8	1.00
Special Care Ward (Intervention)	5.7	1.2	4.9 – 6.5	3.8 – 7.6	1.00
PAS-Score	-2.9	0.2	-3.1 – -2.8	-3.2 – -2.7	1.00
Charlson's Comorbidity Index	-0.1	0.3	-0.4 – 0.1	-0.6 – 0.5	1.00
Psychotropic Drug Use (yes, as-needed)	-4.4	1.4	-5.3 – -3.5	-6.5 – -2.1	1.00
sigma	11.9	0.4	11.5 – 12.0	11.3 – 12.5	1.00

All Rhat values ~ 1, all mcse < 0.05. Results based on 4 chains each 1000 iterations for each chain. LOO-adjusted R^2: 0.500

4. Test for Practical Equivalence of Parameters

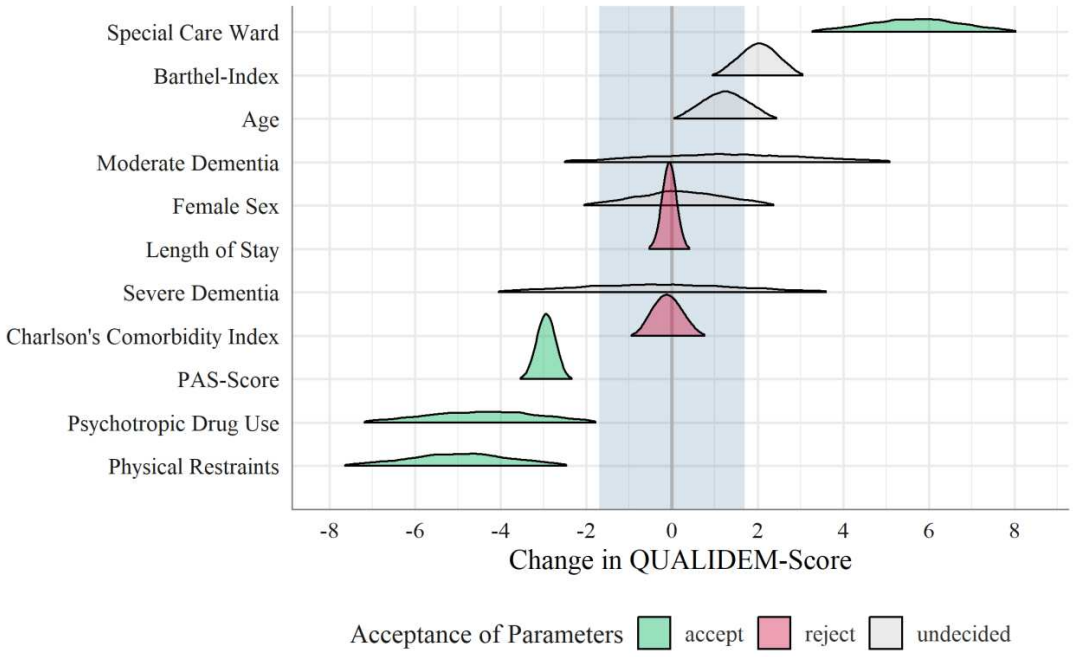
Bayesian methods do not perform classical "null hypothesis significance tests". Rather, to account for the uncertainty in estimation and the magnitude of parameter values, Kruschke[3] suggests checking whether parameter values lie inside a certain range that is considered as "practically no effect".

Figure S2 shows the distribution of the 95% Highest Density Intervals of posterior samples and indicates whether these lie completely outside, completely inside or partially inside the region of practical equivalence (ROPE). If the HDI is completely outside the ROPE, the "null hypothesis" for this parameter is "rejected". If the ROPE completely covers the HDI, i.e. all most credible values of a parameter are inside the region of practical equivalence, the null hypothesis is accepted. Else, it is undecided whether to accept or reject the null hypothesis.

The ROPE limits are specified following Kruschke's suggestion, which is defined as *0 +/- SD(dependent variable) * 0.1* for linear models.

Physical restraint, psychotropic drug use, agitation (PAS-score) and the intervention (special care ward) are those predictors that can be considered as relevant for our outcome.

Figure S2: *95%-Range of the Distribution of Posterior Samples from Regression Model; Region of Practical Equivalence emphasized in light-blue.*



5. Summary of the Distribution of Prior and Posterior Samples

Bayesian regression models can incorporate prior information about the parameters. Informative or weakly informative priors have the advantage of regularization of parameters to rule out large effects. This leads to less biased results and reductions of outliers in situations where the sample size in general, or for certain groups in the data, is small. Where we found information about our parameters, we included it by specifying *informative priors*. For instance, studies suggest that physical restraints are, on average, associated with an about 5-point reduction in quality of life, so the *location* of the prior distribution of our parameter physical restraint is at -5.0 (see Table S2). *Weakly informative priors* simply have a location of zero.

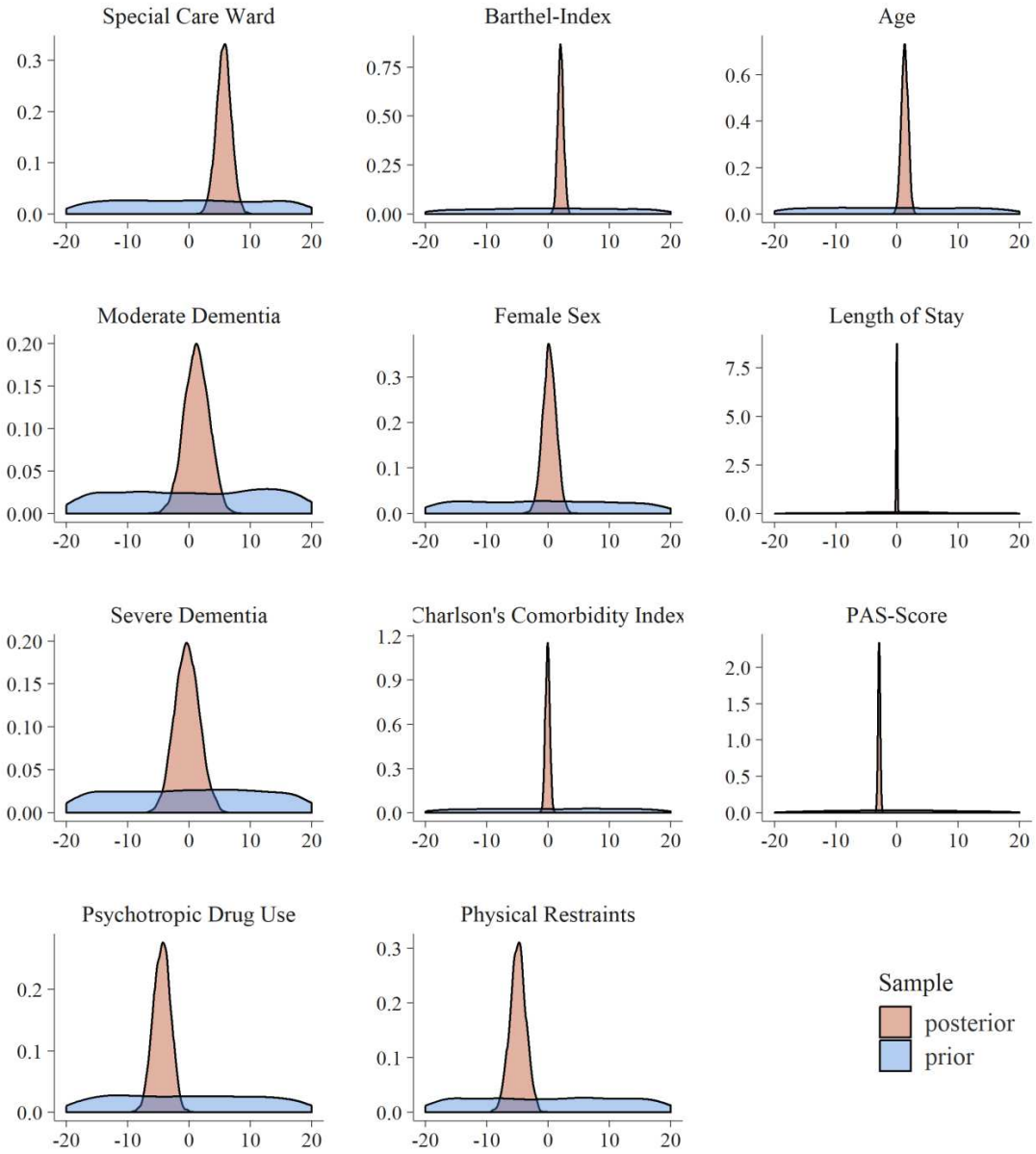
Table S2: *Prior Summary*

Predictor	Location	Scale
Length of Stay	0.00	6
Age	0.16	40
Moderate dementia	0.00	42
Severe dementia	-0.44	42
Sex, female	-3.22	42
Barthel-Score	0.00	29
Physical restraints	-5.00	42
Special Care Ward	0.00	42
PAS-Score	0.00	13
Charlson's Comorbidity Index	0.00	27
Psychotic Drug Use	0.29	42

To prevent a too strong shrinkage of uncertainty, the *scale* of the prior distribution should be large enough. If no assumptions about the scale of the effect are made, a recommended way (for normally distributed priors) is to use a scale of 2.5 and adjust it by dividing the scale by the standard deviation of the related predictor and multiply it with the standard deviation of the outcome (*2.5 * SD(y) / SD(x)*, see https://cran.r-project.org/web/packages/rstanarm/vignettes/priors.html).

When the evidence in the data is *strong*, the impact of the prior information is rather low. On the other hand, if the evidence of the data is *weak*, for instance due to small sample size, the impact of the prior information increases. Figure S3 shows the summary of the prior and posterior sample distributions and suggests that the evidence in the data was rather strong and results were not biased due to unrealistic prior information.

Figure S3: *Posterior versus Prior Summary*

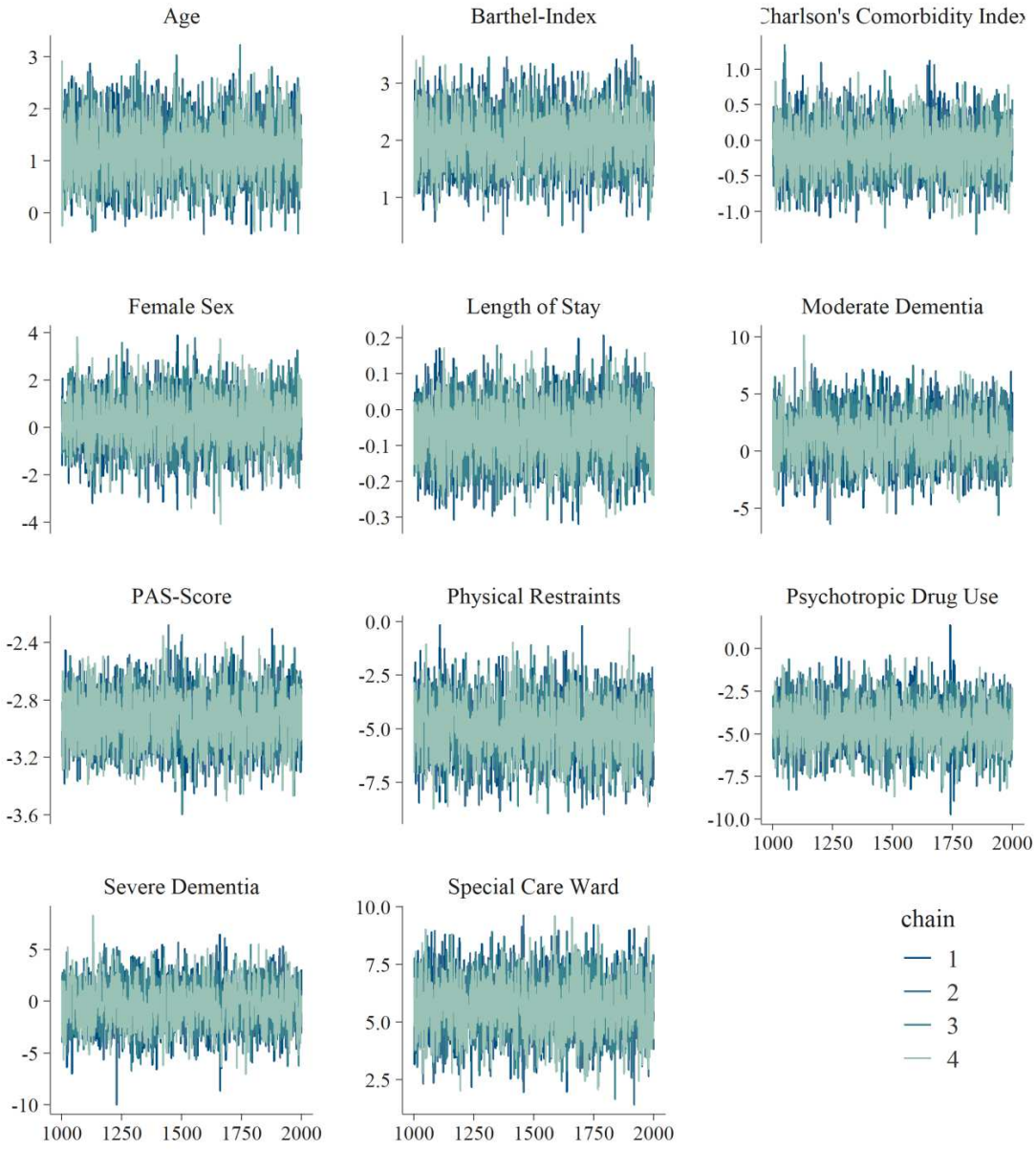


6. Trace plots

A trace plot provides a visual way to inspect the sampling behaviour and can be used, besides the Rhat-statistics, to assess convergence. The desired pattern is a “fat, hairy caterpillar”, which shows no suspicious bends. If the chains converged to the posterior distribution, we can be confident in our model-based inferences.

The trace plot in Figure S4 indicates that all chains have converged well.

Figure S4: *Markov Chain trace plot, 4 chains with 1000 iterations each (excluding warmup-samples)*



7. References

1. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Third edition. Boca Raton: CRC Press; 2014: p.24

2. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science. 1992;7(4):457-472. doi:10.1214/ss/1177011136

3. Kruschke JK. Rejecting or Accepting Parameter Values in Bayesian Estimation. Advances in Methods and Practices in Psychological Science. 2018; doi: 10.1177/2515245918771304

4. Sorensen T, Hohenstein S, Vasishth S. Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. The Quantitative Methods for Psychology. 2016;12:175–200.

5. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. The BUGS book: a practical introduction to Bayesian analysis. Boca Raton, FL: CRC Press, Taylor & Francis Group; 2013.