Using linear mixed models to compare a self-assessed frailty score with clinician assessed scores in patients approaching major surgery

Mohammad Sayari[1], James Durrand[2], Christopher Taylor[3], Jochen Einbeck[1], Ehsan Kharatikoopaei[4], Joshua Craig[5], Nathan Griffiths[2]

[1] Durham University, UK
[2] The Newcastle upon Tyne NHS Foundation Trust, UK
[3] The South Tees NHS Foundation Trust, UK
[4] Manchester Metropolitan University, UK
[5] Northumbria Healthcare NHS Foundation Trust

E-mail for correspondence: `mohammad.sayari@durham.ac.uk`

Abstract: Frailty is a syndrome of reduced physiological and cognitive reserve resulting in vulnerability to physiological insult and delayed recovery. It is a recognised predictor of poor perioperative outcomes. The Rockwood clinical frailty score (CFS) is a validated frailty screening tool based on the appearance of the patient in clinics. A study sponsored by South Tees Hospitals NHS Foundation Trust investigated whether patients may be able to self-assess their frailty utilizing a modified Rockwood CFS, by benchmarking the self-assessed scores with a clinician- and a researcher-assessed CFS score. A linear mixed-effects model, involving covariates such as age and ASA scores, was used to compare the CFS frailty scores and to identify any differences in their agreement. Linear mixed-effect model trees were also used for a better understanding of interactions of covariates and scorer effects.

Keywords: Frailty; peri-operative care; Clinical Frailty Scale; Cohen's kappa statistic; linear mixed-effects model

1 Introduction

Frailty is a recognised predictor of poor perioperative outcomes (Lin et al., 2016). Preoperative assessment of frailty is key to allow planning of peri-

This paper was published as a part of the proceedings of the 38th International Workshop on Statistical Modelling (IWSM), Durham, UK, 14–19 July 2024.

operative care, and discussions with patients to manage risk, expectations, and facilitate shared decision-making and informed consent.
The Rockwood clinical frailty score (CFS) is a validated scoring system-based global clinical impression of frailty based on the appearance of the patient in clinic. It is in routine use in patients over 64 in the perioperative and wider clinical settings. The CFS groups patients into 9 classes ranging from very fit to severe frailty, each allocated a numerical value of 1-9, increasing with rising frailty (Rockwood et al., 2005). Typically, a person allocated a score of 1-3 is labelled as 'non-frail'. A person scoring 4 is labelled 'pre-frail', a score of 5-8 is 'frail' and 9 is 'terminally ill'.
Recently, drivers toward a digitalised NHS, along with the COVID-19 pandemic, have encouraged remote clinical working and telemedicine to deliver patient care. This limits the applicability of the CFS without a face-to-face patient contact, removing a key component of comprehensive preoperative assessment.
A surrogate marker for frailty is required. We propose that patients may be able to self-assess their frailty utilizing a modified Rockwood CFS. If patient self-assessment is feasible and agreement with clinician assessed CFS is acceptable, this would be a stepping stone to wider validation and utilisation as a remotely delivered preoperative frailty assessment tool.

2 Methods

Initially, agreement between CFS frailty scores was examined using the quadratic weighted Cohen's Kappa. Values for levels of agreement using the Kappa coefficient are interpreted as follows: < 0 = no agreement; 0.00-0.20 = slight agreement; 0.21-0.40 = fair agreement; 0.41-0.60 = moderate agreement; 0.61-0.80 = substantial agreement; 0.81-1.00 = almost perfect agreement (Landis et al., 1977).
However, such an analysis does not allow for the investigation of covariate effects such as age or ASA score, on the strength of agreement between scores. Hence, a linear mixed-effects model was set up to compare the CFS frailty scores. The linear mixed-effects model allows assessing covariate impacts and interactions when comparing CFS frailty scores, and accounts for intra-patient correlation using a patient-level random effect, hence enabling the computation of robust standard errors to minimise the likelihood of false conclusions. We consider the scores produced by the patient, clinician, and researcher, as pertaining to assessment groups $j = 0, 1$ and 2, respectively. We denote by y_{ij} the measured score for patient i on group j, and by the vector x_i any covariates of interest for patient i. Then a linear mixed-effects model can be formulated as:

$$y_{ij} = \sum_{j=1}^{2} \gamma_j 1_{\{\text{group}=j\}} + x_i^T \beta + u_i + \varepsilon_{ij},$$

where the terms involving the γ_j and β are fixed effects, and u_i is a patient-level random intercept. In the summation term, the patient self-assessment ($j = 0$) serves as the reference category. The fixed effect parameters γ_j capture the agreement differences of interest. Additionally, and not displayed here notationally, we considered models using interaction terms between the grouping variables and the covariates age and ASA score. The linear mixed-effects models were fitted using function `lmer` in R package **lme4**.
In addition, for a more comprehensive understanding of the interaction between covariates and scorers, we used linear mixed-effects model trees. The GLMM tree algorithm is an extension of the model-based recursive partitioning (MOB) method. The MOB method uses a parameter instability test to select partitioning variables. However, MOB is not suitable for multilevel data. To address this limitation, the GLMM tree algorithm was developed to incorporate random effects into the analysis (Fokkema et al., 2018). While random effects are estimated globally using all observations, the fixed effects are estimated locally. The dataset is partitioned based on additional covariates or partitioning variables, and fixed effects are estimated for each partition cell. The GLMM tree model was estimated using the function `lmertree` from the R package **glmertree**.

3 Results

All patients aged 65 or over who were listed for major surgery were included in the study ($n = 80$). Table 1 presents the inter-rater reliability of the CFS frailty scores using Cohen's Kappa. The results demonstrate a moderate agreement between patient-allocated self-score and pre-assessment score on the 9-level scale ($\kappa = 0.43$). There was also a moderate agreement between the pre-assessment score and the research team score on the 9-level scale ($\kappa = 0.59$). There was a substantial agreement on the 9-level scale CFS between the patient-allocated self-score and the research team score ($\kappa = 0.62$). On the 3-level scale, the results indicate a fair agreement between the patient-allocated self-score and pre-assessment score ($\kappa = 0.32$). There was a substantial agreement on the 3-level scale CFS between the patient-allocated self-score and the research team score ($\kappa = 0.68$). There was a moderate agreement between the pre-assessment score and the research team score on the 3-level scale ($\kappa = 0.55$).
Table 2 represents the results of the linear mixed-effect model. The results show that the patient-allocated self-scores were higher than pre-assessment scores (model 1, $p = 0.015$). There were no significant differences between the patient-allocated self-score and the research team score (model 1, $p = 0.588$). In model 2, patient-assessed scores tend to be higher than the other ones, but older patients (age $> 74y$) behave differently than younger patients in the sense that older patients do not assess themselves frailer than the other scores would indicate. The results for interaction between

ASA (American society of anesthesiology) and groups (model 3) indicate that there were no significant differences between patients with $ASA \geqslant 3$ and $ASA < 3$ when comparing the pre-assessment and research team with patient-allocated self-scores.

To gain a deeper understanding of how covariates and scorers interact in models 2 and 3, we used the GLMM tree model. The GLMM trees for models 2 and 3 are shown in Figures 1 and 2, respectively. In each inner node of the plotted trees, the splitting variable and corresponding p-value from the parameter stability test are reported. The diagram in each figure shows two terminal nodes for CFS. In Figure 1, Node 2 shows that patients 74 years of age or younger had higher self-assessment scores compared to pre-assessment and research team scores. In node 3 (patients over 74), there were no substantial differences between scores. In Figure 2, patients with ASA scores under 3 had slightly higher self-assessment scores than pre-assessment and research team scores, and in node 4, patients with ASA scores of 3 or higher had slightly lower pre-assessment compared to the other scores. Please note that the p-values displayed in the top node indicate the significance of the split; that is the existence of subgroups with differing behavior. They make no statement on significant differences in scorer types within those subgroups.

TABLE 1: Inter-rater reliability on 9-point and 3-level clinical frailty score.

Agreement measured (Kappa statistics)	9-point scale	3-level scale
Patient allocated self-score vs. pre-assessment score	0.433	0.319
Patient allocated self-score vs. research team score	0.622	0.683
Pre-assessment score vs. research team score	0.591	0.554

All P-values < 0.01

TABLE 2: Results of linear mixed-effects models.

Model	Formula	Fixed effect	Effect Estimate (β coefficient)	Standard error	P-value
1	CFS score $\sim$ group + $(1\|ID)$	(Intercept)	3.539	0.126	< 0.001
		group (pre-assessment)	-0.289	0.117	0.015
		group (research team)	-0.064	0.117	0.588
2	CFS score $\sim$ group + age>74 + group:age>74 + $(1\|ID)$	(Intercept)	3.498	0.164	< 0.001
		group (pre-assessment)	-0.498	0.154	0.001
		group (research team)	-0.253	0.154	0.101
		age>74	0.084	0.25	0.737
		group (pre-assessment):age>74	0.487	0.235	0.039
		group (research team):age>74	0.443	0.235	0.06
3	CFS score $\sim$ group + ASA $\geqslant$ 3 + group:ASA $\geqslant$ 3 + $(1\|ID)$	(Intercept)	3.171	0.176	< 0.001
		group (pre-assessment)	-0.229	0.171	0.184
		group (research team)	-0.200	0.171	0.245
		ASA $\geqslant$ 3	0.660	0.24	0.006
		group (pre-assessment):ASA $\geqslant$ 3	-0.114	0.235	0.629
		group (research team):ASA $\geqslant$ 3	0.236	0.235	0.316

Reference category: patient self-assessment score

CFS clinical frailty score, ASA American society of anesthesiology

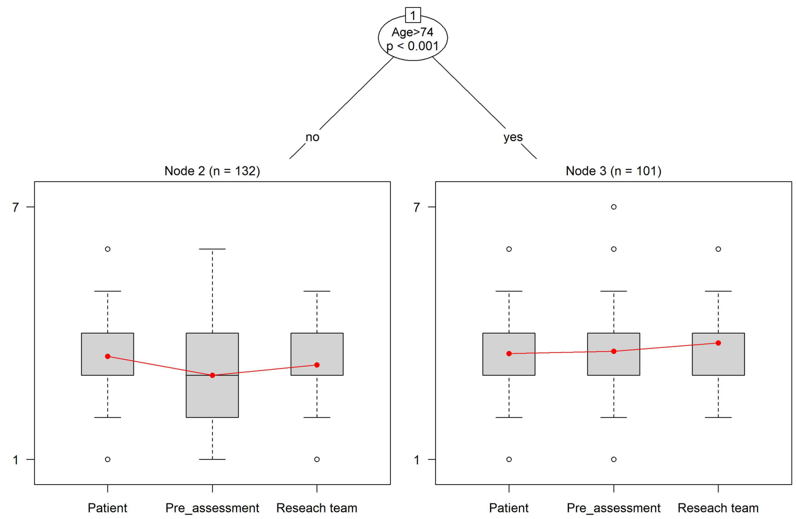

FIGURE 1: Fitted linear mixed-effects model tree for model 2

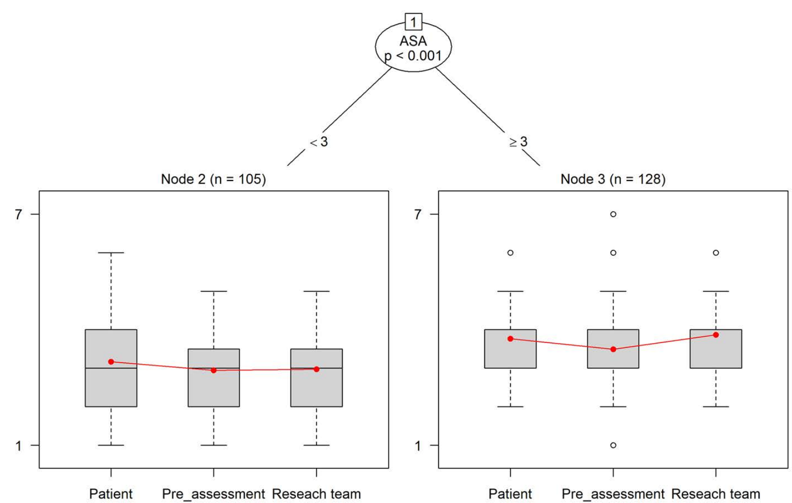

FIGURE 2: Fitted linear mixed-effects model tree for model 3

4 Conclusion

In this study, we evaluated the use of patient self-assessment as a surrogate marker for clinician-assessed frailty. Our findings suggest that patients can evaluate their frailty by using a modified Rockwood CFS. In an additional analysis, we also assessed the agreement between CFS frailty scores using the intraclass correlation coefficient and Bland–Altman plots. All of the results confirmed that there was an acceptable agreement between the self-scores allocated by the patients and the research team scores, with some tendency for relatively younger patients to assign themselves larger frailty scores.

References

Fokkema, M., Smits, N., Zeileis, A., et al. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior research methods*, **50**, 2016–2034.

Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.

Lin, H.S., Watts, J.N., Peel, N.M. and Hubbard, R.E. (2016). Frailty and post-operative outcomes in older surgical patients: a systematic review. *BMC Geriatrics*, **16**, 1–12.

Rockwood, K., Song, X., MacKnight, C., Bergman, H., et al. (2005). A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal*, **173**, 489 – 495.