Stability in Frailty Assessment: Can Automated Measures Reduce Visit-to-Visit Variability in Emergency Department Assessment?

Liam Dunnell[1*], Hugh Logan Ellis[2,3*], Ruth Eyres[4], Dan Wilson[5], Cara Jennings[5], Jane Tippett[5], Julie Whitney[5,7], James T Teo[2,5,6], Zina Ibrahim[2], Kenneth Rockwood[3]

[1]University Hospital Lewisham • [2]Department of Biostatistics & Health Informatics, KCL, London • [3]Department of Medicine, Dalhousie University • [4]Princess Royal University Hospital • [5]King's College Hospital NHS Foundation Trust • [6]Guy's and St Thomas' NHS Foundation Trust • [7] School of Life Course & Population Sciences, Faculty of Life Sciences and Medicine, KCL, London • [*] Joint first authors

Background and Objectives

Frailty assessment in Emergency Departments (ED) is crucial but challenging. Nurse-assessed Clinical Frailty Scale (CFS) scores vary significantly between visits - as seen in Figure 1, where each line represents a frequent ED visitor

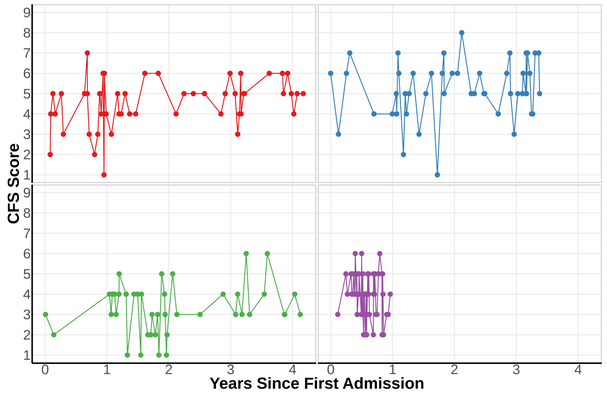

Figure 1. Visit-to-visit variability in Clinical Frailty Scale scores. Each line represents a single patient with frequent ED visits, demonstrating how greatly scores can fluctuate between visits. These changes may reflect both assessment variability and genuine changes in the patient's condition.

This variation reflects clinical context rather than just random error. Patients presenting to the ED with confusion tend to receive higher CFS scores compared to those with chest pain, scores also correlate with illness severity (NEWS score), vary by hospital site, and differ between individual assessors.

Could laboratory-based frailty indices provide more consistent assessments while maintaining clinical usefulness?

Our study compares the stability of automated measures against nurse assessments for older adults in emergency care

Key focus: Comparing stability of automated measures with nurse assessments

Methods

We analyzed ED visits from patients over 70 who attended two London hospitals between 2017-2021:

- King's College Hospital (KCH) – a teaching hospital
- Princess Royal University Hospital (PRUH) – a district general hospital

Our study included 23,956 patients with multiple visits (totaling 60,381 visits)

We developed different automated frailty indices using blood test results as shown in Table 1

We then used statistical models (mixed effects) to measure variability compared to nurse-led CFS scoring

What is an FI-Lab?

FI-Labs are automated frailty measures based on the principle of deficit accumulation. They can be configured as:

- **Chronic FI-Labs** (used in this study):
  - Examine data over extended periods (e.g., 36 months)
  - Can include blood tests, medications, diagnoses, vital signs, and other clinical data
  - Create a stable baseline picture of health status
- **Acute FI-Labs**:
  - Use recent clinical data only
  - Can incorporate the same types of information as chronic FI-Labs
  - Better reflect current health state
  - More sensitive to acute changes

Both follow frailty index principles: more abnormal results = greater frailty. The total number of items measured (the denominator) can vary based on available data, typically ranging from 10-40 variables.

FI-Lab Configurations

Configuration	Approach	Requirements	Strengths	Limitations
Nurse CFS Assessment	Clinical evaluation based on patient presentation	Single assessment at visit	High outcome discrimination, Clinical judgment	Lower stability (ICC=0.35)
Base FI-Lab	Laboratory values over 36-month window	Measurements in ≥3 months	Good stability (ICC=0.55), Objective	Requires historical data
Short-period	Same as base but 12-month window	Measurements in ≥3 months	Reduced data needs	Less historical context
Mean-type	Mean value across months vs reference range	Measurements in ≥3 months	Simple calculation	May be swayed by one-off very high readings
High-features	Base with stricter data requirements	Measurements in ≥10 months	Very stable (ICC=0.62)	Limited population coverage
Low-features	Maximum population coverage	Measurements in ≥1 month	Wider applicability	Less reliable measures
Drug-adjusted	Base FI-Lab plus medications	Base plus medication data	Highest stability (ICC=0.74)	Complex data requirements

Table 1. Comparison of different frailty assessment methods and their characteristics

Results

Nurse-assessed CFS scores showed considerable variability:

- Only 35% of score differences between visits reflected actual patient differences (ICC=0.35)
- Scores were significantly associated with presenting complaints and illness severity

Automated measures demonstrated much higher consistency:

- Stability ranged from 48% to 74% across different configurations
- The drug-adjusted version showed the highest consistency (74%)

Importantly, automated measures remained stable despite acute illness factors while still relating meaningfully to clinical outcomes

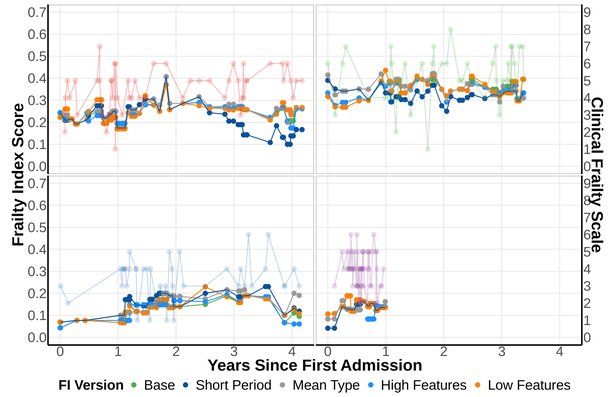

Figure 2. Comparison of visit-to-visit variability between nurse-assessed CFS and automated FI-Lab measures. FI-Lab configurations consistently show higher stability across visits, with much less fluctuation between assessments.

Key Findings

Scale of Study:
23,956 patients with multiple ED visits (60,381 total visits) across two London hospitals

Stability Comparison:
Nurse-assessed CFS scores varied considerably between visits (only 35% consistency), while automated measures showed much higher stability (48-74% consistency)

Combined Approach:
While nurse assessments better predicted outcomes, automated measures provided more stable baseline assessments

Practical Implications:
Automated screening could complement clinical judgment, especially for patients with limited history

Conclusions

Automated measures offer valuable stability that complements clinical assessment

Neither approach fully captures frailty on its own - combining them likely provides the best picture

Automated measures are particularly helpful for:

- Establishing a consistent baseline
- Supporting assessment when patient history is limited
- Providing consistent tracking over time

Future research should focus on:

- Integrating these measures into clinical workflows
- Finding the optimal combination of laboratory features
- Testing these approaches in different healthcare settings

For more information on outcome prediction performance, please see our other poster: "Towards Universal Frailty Screening: Defining Minimum Requirements for Automated Assessment."