

# Predicting Mortality using Frailty Index and Latent Class Approaches: AUC-Based Evaluation in Simulated Data

Martella Donato<sup>(1,2)</sup>, Zambon Antonella<sup>(1,3)</sup>

(1) Laboratory of Quantitative Methods for Life, Health and Society, Department of Statistics and Quantitative

(2) Methods, University of Milano-Bicocca, Milan, Italy

(3) Unit of Epidemiological Psychiatry and Digital Mental Health, IRCCS Istituto Centro San Giovanni di Dio

(4) Fatebenefratelli, Brescia, Italy

(5) Biostatistics Unit, IRCCS Istituto Auxologico Italiano, Milan, Italy

CORRESPONDING AUTHOR: Martella Donato, [d.martella2@campus.unimib.it](mailto:d.martella2@campus.unimib.it)

## BACKGROUND

The global population is aging rapidly, creating significant challenges for healthcare systems [1]. Traditional disease-centered models often fail to meet the complex needs of older adults, who frequently have multiple chronic conditions. In geriatric medicine, frailty has become a key concept [2-3], representing increased vulnerability due to a decline in physiological reserves and functional capacity. Frailty is a multi-dimensional syndrome that includes physical, cognitive, psychological, and social impairments, making its identification vital for improving patient outcomes and healthcare resource allocation. The Frailty Index (FI) is a widely used tool that quantifies frailty by measuring the ratio of health deficits to the total number of health variables [4]. This approach allows for practical and scalable assessments across various settings.

However, while the FI offers a comprehensive assessment, it remains an observed composite measure that may not fully capture the underlying latent nature of frailty. Frailty can be conceptualized as a latent [5-6] construct. Studying frailty as a latent variable enables a deeper understanding of its structure and heterogeneity. It allows researchers to explore whether distinct frailty phenotypes [7] exist and whether they differ in their relationship to key outcomes such as mortality or functional decline. Moreover, latent variable models can uncover hidden patterns that are not evident from single observed measures like the FI, potentially offering more nuanced tools for risk stratification and clinical decision-making.

## OBJECTIVES

To compare the predictive accuracy of mortality models based on a continuous FI and latent class approaches

through simulation, using the area under the curve (AUC) as the evaluation metric.

## METHODS

The simulation uses real-world data on 50 binary and ordinal items related to HyperFrail, focusing on marginal probabilities and empirical correlations. Data simulation began by generating multivariate normal data using the empirical correlation matrix for a sample of 1,000 individuals with 50 variables each. These continuous values were transformed into uniform probabilities and mapped to discrete response levels based on specified marginal distributions. A dataset was created with one row per individual and the calculated FI as the sum of the 50-item responses, normalized by the number of items. Five domain-specific subscores were derived: activity (items 1-16), health (items 17-20), psychological (items 21-25), comorbidity (items 26-43), and cognitive (items 44-50). Age was simulated based on the frailty index (FI) levels. Individuals with an FI of less than 0.12 were assigned a random age between 20 and 50 years. For those with an FI of 0.12 or greater, age was sampled between 50 and 80 years, following an exponentially increasing distribution to reflect the greater frailty commonly observed in older adults. Mortality was simulated conditionally based on age: individuals under 50 were assigned a death status of 0 (indicating no death), while those aged 50 and above were assigned a death status with a probability ranging from 30% to 80% (in 5% increments) to represent various levels of risk. Three models were developed to predict binary death outcomes. The first model utilized logistic regression, employing the continuous Frailty Index (FI) on a scale from 0 to 100. The second model applied Latent Class Analysis (LCA) using the polCA [8] package on 50 item variables to identify various frailty classes,

followed by logistic regression where frailty class served as a categorical predictor. The third model used Gaussian Mixture Modeling (GMM) with the Mclust [9] package, analyzing five domain-specific summary scores to identify latent clusters. The Area Under the Curve (AUC) was calculated for each model to evaluate discrimination performance. Finally, a simulation study was conducted to assess model performance across different mortality scenarios.

## RESULTS

Latent class analysis using Bayesian Information Criterion (BIC) identified nine distinct frailty classes across five domains. Class 1 represented “low frailty” with minimal deficits, while Class 4 showed “high activity limitation” mainly affecting the activity domain. Class 6 had a “cognitive-predominant” phenotype, marked by significant cognitive impairment. Class 8 displayed a “multi-domain severe” pattern with high frailty scores in activity, health, and comorbidity. Finally, Class 9 exhibited extreme frailty with the highest burden, especially relating to comorbidity and health status. The comparison of the AUC showed that the continuous FI consistently outperformed both latent variable approaches in all mortality probability scenarios. The FI displayed superior discriminative ability, particularly at higher mortality probabilities. It was followed closely by the GMM, while the LCA demonstrated the lowest predictive performance.

## CONCLUSION

The continuous FI showed better predictive accuracy for mortality outcomes, while the latent class approach identified significant frailty phenotypes with distinct patterns in specific domains that may have important clinical implications. These findings indicate that, although the comprehensive nature of the FI provides more effective discrimination of mortality risk, understanding frailty as a latent variable offers valuable insights into the diverse characteristics of this syndrome.

## REFERENCES

1. Braithwaite J, Ludlow K, Testa L, et al. Built to last? The sustainability of healthcare system improvements, programmes and interventions: a systematic integrative review. *BMJ Open*. 2020;10(6):e036453. doi:10.1136/bmjopen-2019-036453
2. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ*. 2005;173(5):489-495. doi:10.1503/cmaj.050051
3. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet*. 2013 Mar 2;381(9868):752-62. doi: 10.1016/S0140-6736(12)62167-9. Epub 2013 Feb 8. Erratum in: *Lancet*. 2013 Oct 19;382(9901):1328. PMID: 23395245; PMCID: PMC4098658.
4. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. *J Gerontol A Biol Sci Med Sci*. 2007 Jul;62(7):722-7. doi: 10.1093/gerona/62.7.722. PMID: 17634318.
5. Feng Y, Hancock GR. A structural equation modeling approach for modeling variability as a latent variable. *Psychol Methods*. Published online April 11, 2022. doi:10.1037/met0000477
6. Salem BE, Nyamathi A, Brecht ML, et al. Constructing and identifying predictors of frailty among homeless adults—a latent variable structural equations model approach. *Arch Gerontol Geriatr*. 2014;58(2):248-256. doi:10.1016/j.archger.2013.09.005
7. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146-156. doi:10.1093/gerona/56.3.m146
8. Drew A, Linzer, Jeffrey B. Lewis (2011). *polCA: An R Package for Polytomous Variable Latent Class Analysis*. *Journal of Statistical Software*, 42(10), 1-29. URL <https://www.jstatsoft.org/v42/i10>
9. Scrucca, L., Fraley, C., Murphy, T.B., & Raftery, A.E. (2023). *Model-Based Clustering, Classification, and Density Estimation Using mclust in R* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781003277965>

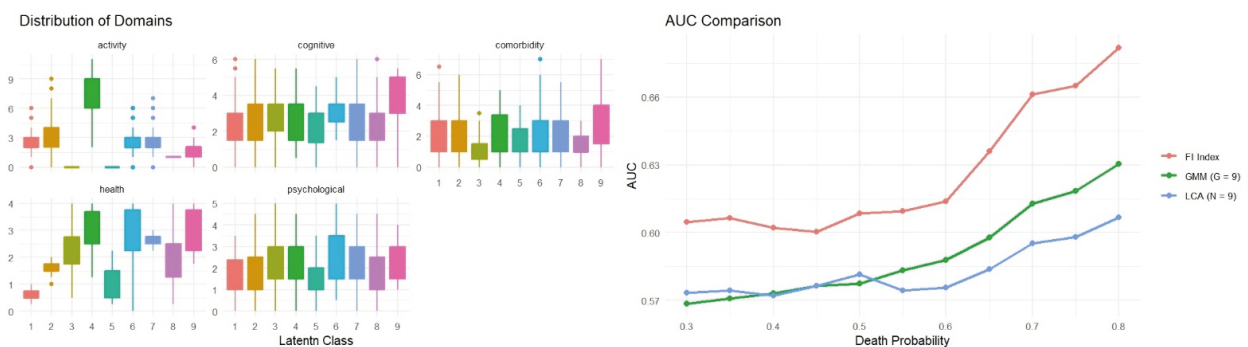


Figure 1. Distribution of Domains across latent classes and comparison of AUC for three models in different death scenarios