



Data-driven health deficit assessment improves a frailty index's prediction of current cognitive status and future conversion to dementia: results from ADNI

Andreas Engvig · Luigi A. Maglanoc · Nhat Trung Doan · Lars T. Westlye · Alzheimer's Disease Neuroimaging Initiative

Received: 17 May 2022 / Accepted: 2 October 2022 / Published online: 19 October 2022
© The Author(s) 2022

Abstract Frailty is a dementia risk factor commonly measured by a frailty index (FI). The standard procedure for creating an FI requires manually selecting health deficit items and lacks criteria for selection optimization. We hypothesized that refining the item selection using data-driven assessment improves sensitivity to cognitive status and future dementia conversion, and compared the predictive value of three FIs: a standard 93-item FI was created after selecting health deficit items according to standard criteria (FI_s) from the ADNI database. A refined FI (FI_r) was calculated by

using a subset of items, identified using factor analysis of mixed data (FAMD)-based cluster analysis. We developed both FIs for the ADNI1 cohort ($n=819$). We also calculated another standard FI (FI_c) developed by Canevelli and coworkers. Results were validated in an external sample by pooling ADNI2 and ADNI-GO cohorts ($n=815$). Cluster analysis yielded two clusters of subjects, which significantly ($p_{FDR} < .05$) differed on 26 health items, which were used to compute FI_r . The data-driven subset of items included in FI_r covered a range of systems and included well-known frailty components, e.g., gait alterations and low energy. In prediction analyses, FI_r outperformed FI_s and FI_c in terms of baseline cognition and future dementia conversion in the training and validation cohorts. In conclusion, the data show that data-driven health deficit assessment improves an FI's prediction of current cognitive status and future dementia, and suggest that the standard FI procedure needs to be refined when used for dementia risk assessment purposes.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11357-022-00669-2>.

A. Engvig (✉) · L. A. Maglanoc · N. Doan · L. Westlye
NORMENT, Division of Mental Health and Addiction,
Oslo University Hospital & Institute of Clinical Medicine,
University of Oslo, Oslo, Norway
e-mail: andreas.engvig@gmail.com

A. Engvig
Department of Nephrology, Oslo University Hospital,
Oslo, Ullevål, Norway

Keywords Alzheimer's disease · Dementia · Mild cognitive impairment · Frailty · Machine learning · Frailty index

L. A. Maglanoc
University Center for Information Technology, University
of Oslo, Oslo, Norway

L. Westlye
Department of Psychology, University of Oslo, Oslo,
Norway

Introduction

Frailty is an age-related state of multisystem physiological decline increasing the risk of adverse outcomes such as hospital complications and death [1–3]. The frailty index (FI) operationalizes frailty along a continuum based on the accumulation of health deficits model [4]. Evidence suggests that FIs may predict conversion to dementia [2, 5, 6] and points to FIs as useful measures for identifying subjects at high dementia risk. As frailty may be reversible [7], it is conceivable that frailty may serve as a target for dementia prevention. FIs are suitable for measuring frailty in this regard, as they are validated for longitudinal assessment [8] and show promise as outcome measures in clinical trials [9].

Standard FIs comprise a selection of several age-related health deficits which fit pre-defined criteria [10], and, when reflecting the accumulated burden of 30 to 40 health deficit variables, are robust for prediction of mortality [11]. In terms of dementia risk prediction, to the best of our knowledge, FI studies with external validation have so far been lacking. Traditionally, the deficit accumulation model has included more cognition-related measures when operationalizing frailty compared with its rival, the phenotype model [20]. Although frailty and dementia are inter-related, they are distinct concepts and by including dementia-related measures (e.g., certain activities of daily living, cognitive test results) into an FI, its use in the prediction of dementia may become circular. While Ward and coworkers found a significant association between FI and future dementia risk when adjusting for global cognition [12], results across studies show, however, that the association between FIs and future dementia risk weakens after removing deficits which might represent early core dementia symptoms [6, 12]. How to construct an optimal FI for dementia risk prediction purposes remains unknown. One caveat is that the standard procedure lacks criteria for discarding deficits with little explained variance, which may reduce FI performance [14].

In general, one approach to maximize explained variance is by employment of data-driven techniques which decompose and weigh correlated input variables. While the accuracy of a neural-network based machine learning approach outperformed an unweighted (i.e., standard) FI in the prediction of mortality [15], the application of machine learning weights to individual patients has been intangible. Using machine learning to guide selection of health deficit items at the development

stage of an FI may represent a more feasible approach. The resulting FI may in turn be applied to any subject for which all or a subset of the relevant health deficit data exists. In general, discarding non-informative variables is an efficient step to remove noise and improve model performance. As an example, a data-driven refined FI based on a selection of 35-items identified using factor analysis among a larger set of items nearly outperformed a 139-item standard FI in terms of mortality prediction [16]. In contrast to factor analysis, principal component analysis (PCA) aims to maximize explained variance in variable set and may be used for weighting health deficits [17]. As most FIs include mixed data, factor analysis of mixed data (FAMD)—a combination of PCA for continuous variables and multiple correspondence analysis (MCA) for categorical—is needed [18, 19]. FAMD has thus far, however, not been applied to the field of frailty assessment.

To this end, the purpose of the present study was to test adding FAMD-based cluster analysis to the standard FI procedure as a way to empirically guide health deficit selection (Fig. 1). Cluster analysis has previously been shown to identify subjects living with frailty in an unsupervised manner [20]. We hypothesized that data-driven health deficit selection would (1) improve the stability and replicability of an FI compared with standard procedure only, and (2) enhance its prediction of cognitive impairment, even when adjusting for cognitive and functional performance using a validated Clinical Dementia Rating (CDR) scale. To test our hypotheses, we created a data-driven FI (FI_r) by applying FAMD-based clustering to a set of health variables selected according to standard criteria [10] from the ADNI database. The main objective was to develop and externally validate FI_r and compare it to two standard FIs using the ADNI1 (development) and ADNI2 and ADNI-GO (validation) cohorts against cognitive status and future dementia risk.

Methods

Data source

For model training, we used data from ADNI1, an observational prospective case–control cohort study (hereafter denoted “development sample”) of subjects between ages 55 to 90 living with normal cognition,

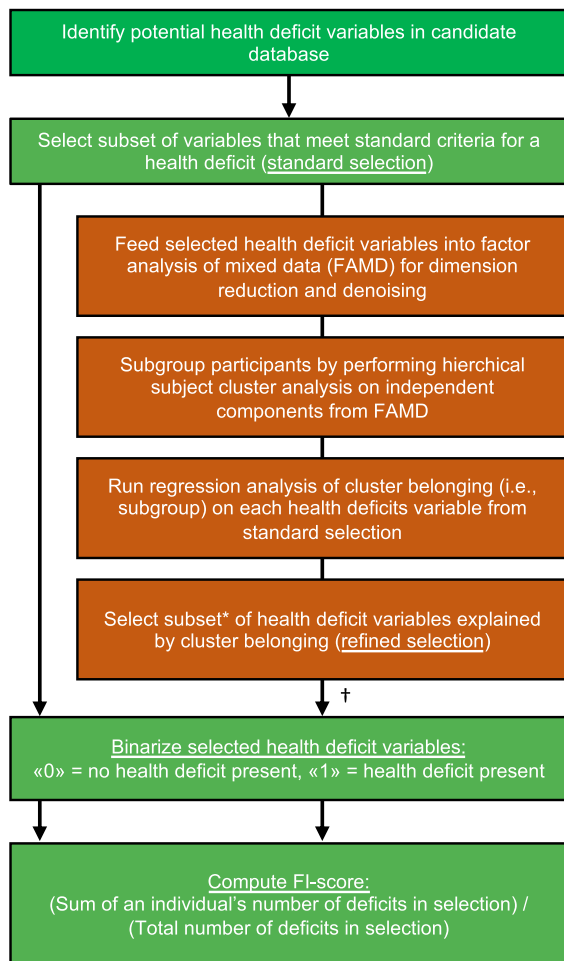


Fig. 1 Data-driven supplement (red boxes) to the standard procedure (green boxes) for creating a frailty index (FI). *For the refined selection of health deficit variables, a false discovery rate (FDR)-adjusted p value $< .05$ from regression analyses of each variable on cluster belonging is used as selection threshold. †Assess face validity of refined selection against standard criteria and core frailty construct. FI_r (see “Methods” section) was developed using the data-driven supplement to the standard procedure as shown, whereas FI_s and FI_c were developed using standard procedure only

mild cognitive impairment (MCI) or Alzheimer’s disease dementia (AD). For external validation, we pooled participants from two other prospective cohorts, ADNIGO and ADNI2. All data are publicly available (<https://ida.loni.usc.edu/>). The ADNI dataset used herein was downloaded on October 21st, 2021. ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been

to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and AD. For up-to-date information, see www.adni-info.org.

Sample

Inclusion criteria for all participants included age 55 to 90 years; study partner to provide evaluation of function; speaks English; ability to undergo all testing, blood samples for genotyping and biomarkers, and neuroimaging procedures; completed six grades of education or work history; for women postmenopausal or surgically sterile, not depressed, and a modified Hachinski score less than five in order to rule out vascular dementia. Individuals with AD satisfied criteria for NINCDS/ADRDA for probable AD. Subjects enrolled as MCI had memory complaints verified by a study partner, Mini Mental Status Examination (MMSE) score of 24 to 30, Clinical Dementia Rating (CDR) score = 0.5 with sum of boxes (CDRSB) score of at least 0.5, and general cognition and functional performance sufficiently preserved such that a diagnosis of dementia could not be made. In later phases of ADNI, subjects with MCI were further sub-divided by their Wechsler Memory Scale Logical Memory II score into early (EMCI) and late (LMCI): in the present study, EMCI and LMCI-subjects were pooled and merely labeled MCI. Healthy controls (HC) had no memory complaints aside from those common to other normal subjects of that age range, MMSE score of 24 to 30, CDR = 0 (with CDRSB score = 0), and were deemed cognitively normal, based on an absence of significant impairment in cognitive functions or activities of daily living. Description of the enrollment is found online (http://www.adni-info.org/Scientists/doc/ADNI_GeneralProceduresManual.pdf).

Statistical analyses

FI computation and statistical analyses were performed in R version 3.6.2. A github-repo is freely available online with the code used to generate the FIs and to perform the analyses detailed below (<https://github.com/LAMaglan/ADNI-FI-clustering>). An FI is based on a cumulative deficits model grading

heterogeneity in health status on a continuous scale from 0 to 1, where greater scores indicate higher degree of frailty [10]. We selected candidate health deficits by assessing screening and baseline variables in the ADNI database. The selection process adhered to standard criteria described by Searle and coworker [10]. Specifically, we included variables reflecting health deficits, i.e., symptoms, clinical signs, diseases, laboratory abnormalities, or other measures associated with adverse health outcomes. Jointly, the deficits should cover many organ systems and consist of more than co-morbidities or function. The selected deficits should generally be considered age-related at the population level, but a variable representing a deficit does not, individually, need to be significantly related to age [11]. We selected health deficits which were present in at least 1% of the sample, but not more than 80%, and discarded variables which were missing in more than 5% of the patients. In accordance with previous research on frailty and cognitive outcomes by the developers of the FI method, we excluded certain health deficits related to the diagnosis of dementia [6, 21]. Specifically, we excluded neuropsychological test results, the remembering item in the Functional Activities Questionnaire (FAQ) and specific neuropsychiatric inventory (NPI) items such as disinhibition and delusions. A total of 93 health deficit variables fit the criteria and were used to compute a standard FI, denoted FI_s . Next, we optimized the initial selection of health deficits using a data-driven refinement process (Fig. 1):

- 1) Factor analysis of mixed data (FAMD): we used FAMD (Pagès, 2004) in the R package *FactoMineR* [19] to reduce the 93 variables into principal components (PC). FAMD is a combination of principal component analysis (PCA) for continuous variables and multiple correspondence analysis (MCA) for categorical variables. All variables were normalized prior to the dimensionality reduction. Missing data for the continuous variables were imputed for dimensionality reduction using the k-Nearest Neighbour algorithm from the R package *bnstruct* [22].
- 2) Subject cluster analysis: the number of PCs from FAMD explaining 80% cumulative variance was used as input for cluster analysis, which is in line with previous clinical studies employing FAMD-based clustering [23, 24]. Above this

level, cumulative explained variance as function of PC number gradually plateaus (Supplementary Figure S1). For clustering, we used Hierarchical Clustering on Principal Components (HCPC) in the R-package *FactoMineR* [19]. HCPC entails agglomerative hierarchical clustering in the first step and k-means clustering to improve the initial clustering. Note that the hierarchical cluster analysis is performed on PCs following FAMD, and not directly on the original health deficit variables themselves. In addition to dimensionality reduction, this additional step is done to reduce noise in the data and generally yields a more stable cluster analysis [25]. We chose a two-cluster solution based on the higher relative loss of the sum of within-cluster inertia [19]. This entails that all participants are empirically divided into two cluster sub-groups.

- 3) Regression analysis for health deficit weighting: we used regression models to rank the relative importance or “weight” of the 93 health deficit variables. To this end, cluster sub-group belonging was used as independent binominal variable and each of the health deficit variables identified following the standard procedure as dependent variables. For binomial dependent variables, we used logistic regression models; for ordinal categorical variables (FAQ items), we used ordinal logistic regression; and for the continuous variables, we used linear regression.

To compute FI_r , we included health deficit variables with a false discovery rate (FDR)-adjusted $p < 0.05$ from the regression analyses in step 3 only (i.e., the highest ranked items). We assessed face validity of the refined set of health deficits included in FI_r against standard criteria for an FI [10] and core components of the frailty construct [1].

In order to compare our FIs with one created by others, we further calculated a published 40-item FI (FI_c) [26]. The items included in each FI is reported in Table 1. For all FIs, selected health deficit variables were binarized with a value of “1” given if a health deficit—or a marker thereof—was present, and “0” if absent. Continuous health deficit variables were dichotomized using established reference ranges, coding “1” as outside the normal range, “0” as within. For coding of continuous blood tests results, we used the ADNI laboratory reference ranges (including age

Table 1 Health deficit items included in FI_s, FI_r, and FI_c

Health deficit variables	Frailty Index (FI) variable		
	FI _s	FI _r	FI _c
On medication for arthritis	X		
On medication for elevated cholesterol	X		
On medication for hypertension	X		
On medication for hypothyroidism	X		
History of cardiovascular disease	X		X
History of endocrine-metabolic disease	X		X
History of head, eyes, ears, nose, and throat diseases	X		X
History of dermatologic-connective tissue disease	X		X
History of gastrointestinal disease	X		X
History of hematopoietic-lymphatic disease	X		X
History of hepatic disease	X		X
History of malignancies	X		X
History of musculoskeletal disease	X		X
History of neurological (non-AD) disease	X		X
History of psychiatric disease	X	X	X
History of renal-genitourinary disease	X		X
History of respiratory disease	X		X
History of other diseases	X		
History of major surgical procedures	X		
FAQ, writing checks, paying bills, or balancing checkbook	X	X	X
FAQ, assembling tax records, business affairs, or other papers	X	X	X
FAQ, heating water, making a cup of coffee	X	X	X
FAQ, traveling out of the neighborhood)	X	X	X
FAQ, preparing a balanced meal	X	X	X
FAQ, paying attention to and understanding a TV program, book, or magazine	X	X	X
FAQ, playing a game of skill such as bridge or chess	X	X	X
FAQ, shopping alone for clothes, household	X	X	X
NPI, agitation	X	X	X
NPI, anxiety	X	X	X
NPI, depression	X	X	
NPI, irritability	X	X	X
NPI, sleep disturbance	X	X	
NPI, apathy	X	X	X
NPI, aberrant motor behavior	X	X	X
NPI, change in appetite and eating	X	X	
Auditory impairment on physical exam	X		
Cranial nerve abnormality on physical exam	X		
Abnormal finger to nose test	X		X
Abnormal heel-knee test	X		
Abnormal gait on physical exam	X	X	X
Motor strength deficit on physical exam	X		
Abnormal plantar reflex on physical exam	X		
Abnormal tendon reflex on physical exam	X		
Sensory nerve abnormality on physical exam	X		
Tremor on physical exam	X		X

Table 1 (continued)

Health deficit variables	Frailty Index (FI) variable		
	FI _s	FI _r	FI _c
Vision impairment on physical exam	X		
Symptoms from abdomen	X		
Self-reported ankle swelling	X		
Self-reported shortness of breath	X		X
Self-reported chest pain	X		
Self-reported constipation	X		X
Self-reported cough	X		
Self-reported depressed mood	X	X	
Self-reported dizziness	X		X
Self-reported drowsiness	X	X	X
Self-reported dry mouth	X		
Self-reported low energy	X	X	X
Self-reported recent fall	X		X
Self-reported insomnia	X		X
Self-reported life satisfaction	X		
Self-reported muscle pain	X		X
Self-reported palpitations	X		
Self-reported urinary discomfort	X		X
Self-reported urinary frequency	X		
Self-reported vision disturbance	X		
Blood albumin (g/dL)	X		
Blood alkaline phosphatase (ALP; U/L)	X		
Blood alanine aminotransferase (ALT; U/L)	X		
Blood aspartate aminotransferase (AST; U/L)	X		
Blood vitamin B12 (pg/mL)	X		
Blood total bilirubin (mg/dL)	X		
Blood calcium (mg/dL)	X		
Blood cholesterol (mg/dL)	X		
Blood creatinine (mg/dL)	X		
Blood glucose (mg/dL)	X		
Blood hematocrit (%)	X		
Blood hemoglobin (g/dL)	X		
Blood mean corpuscular hemoglobin (MCH; pg)	X		
Blood mean corpuscular volume (MCV; fL)	X		
Blood neutrophil count (10 ³ /μl)	X	X	
Blood total protein (g/dL)	X		
Blood red blood cell count (RBC; 10 ⁶ /μl)	X	X	
Blood triglycerides (mg/dL)	X		
Blood urea nitrogen (BUN, mg/dL)	X		
Blood uric acid (mg/dL)	X		
Body mass index (BMI)	X		
Diastolic blood pressure (mmHg)	X		X
Heart rate (count)	X		
Mean arterial pressure (mmHg)	X		
Pulse pressure (mmHg)	X	X	

Table 1 (continued)

Health deficit variables	Frailty Index (FI) variable		
	FI _s	FI _r	FI _c
Systolic blood pressure (mmHg)	X		X
Number of medications (polypharmacy)	X	X	
Elevated geriatric depression scale (GDS) score	X	X	

X denotes that the variable is included in the respective frailty index (FI). *FI_s* = a 93-item frailty index created according to standard procedure. *FI_r* = a 26-item frailty index created according to a refined, data-driven procedure. *FI_c* = a 40-item FI created according to standard procedure by Canevelli, et al. [26]. *FAQ* = Functional Assessment Questionnaire. *NPI* = Neuropsychiatric Inventory

and/or sex-specific cut-points). For activities of daily living variables, the deficit item was recoded in an ordinal manner [26]: for instance, the Finances item of the FAQ was coded as follows: “0” for independent/normal functioning, “0.25” for difficulty, and “0.5” for requires assistance, and “1” for dependent. All cut points used for dichotomization are provided in Supplementary Tables S1 and S2. Note that an FI score was only calculated for individuals who had less than 20% missing variables. The same selection of health deficits used to compute the FIs in the development sample (i.e., the ADNI1-cohort) were subsequently used to calculate corresponding FI-variables for the validation sample (i.e., pooled ADNI2 and ADNI-GO cohorts). The validation sample FIs were tested and displayed along with the results from the development sample for out-of-sample verification.

Diagnostic prediction performance

To assess diagnostic performance of the two standard FIs (FI_s and FI_c) versus FI_r, we performed three sets of machine-learning based binary classification on the diagnostic groups (HC vs AD, MCI vs AD, and HC vs MCI) using linear discriminant analysis from the R packages *discrim* [27] and *tidymodels* [28]. In the main analyses, we ran tenfold internal cross-validation on the ADNI1 data (i.e., whereby 9 of the folds predict the remaining fold iteratively), repeated 100 times on randomly partitioned data. To assess reproducibility, we built our models based on the development cohort (ADNI1) which were then used to classify the diagnostic groups in the validation cohort (ADNI2 and ADNI-GO). To obtain an estimate of standard deviation for each of the binary classification analyses of the validation-sample, we performed

a pseudo-tenfold cross-validation with 100 repetitions, where first the validation-sample was split into 10 folds, where each of the folds was predicted with the same machine learning model (built on the whole development-sample, i.e., ADNI1 dataset) for each of the respective binary classification tasks. We computed area under the curve (AUC) as our main measure of model performance, but also report sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), and the F1-score. Due to differences in diagnostic group size (i.e., class imbalance), we conducted sensitivity analyses to account for imbalanced sampling in the development-sample. Here, for every binary classification task, we under-sampled the larger diagnostic group to match the lower diagnostic group. Then, as in the main analyses in the development sample, we ran tenfold cross-validation with 100 repetitions.

Finally, we examined whether the discriminative ability of the three FI-variables changed according to different age and sex strata by dividing the development sample into young-old and old-old (± 75 years), and male and female. Here, we performed binary classification on the four resulting sub-groups for HC versus AD classification only to constrain the number of analyses run. We ran paired *t* tests to statistically compare the model performance of the two classifiers (either FI_s or FI_c, versus FI_r) in the main analyses, as well as for the sensitivity analyses with undersampling.

Prognostic performance

We evaluated how well the FI-variables predicted future dementia risk by survival analysis of subjects living with MCI at baseline. Analyses were performed

using the *survival* package [29], if not otherwise stated. We assessed time from baseline examination to the date of registered AD conversion. Participants who had not progressed at their last recorded visit were right-censored. We did not account for a competing mortality risk. Dementia-free survival in MCI participants was first assessed by different sample levels of frailty using FI-quartiles without adjusting for covariates by the Kaplan–Meier estimator. To test and compare the predictive performance of the three continuous FI-variables over time we also estimated time-dependent areas under receiver operating characteristic curves (AUC(t)) and their 95% confidence intervals and bands by means of the R package ‘*timeROC*’, using the iid-representation of the AUC estimator for inference [30].

Next, we tested associations between dementia-free survival and the three FIs as continuous variables by fitting multivariate Cox proportional hazards models, taking into account age, sex, education as well as baseline cognitive and functional performance. Here, FI-scores were multiplied by 100 in order to facilitate meaningful interpretation of the associated hazard ratios (HR). In the first model set, we included either one of the continuous FI variables, covarying for age, sex, and education. In the second model set, we assessed relationships between the FI scores and dementia risk when accounting for baseline cognitive and functional performance. To this end, we fitted Cox proportional hazards models with the Clinical Dementia Rating Sum of Boxes (CDRSB) score for each participant as an additional covariate. For the development sample, individual, and global Schoenfeld scaled residuals tests were all non-significant, suggesting proportionality of hazards. In the validation sample, Schoenfeld scaled residuals tests indicated violation of the proportionality of hazards assumption for the age covariate ($p < 0.05$). Thus, for the multivariate validation models, we estimated average hazard ratios [31] using Prentice weights with censoring correction and robust variance estimation as implemented in the *coxphw* package.

To examine predictive validity, we tested associations between the FI-variables and mortality risk—the endpoint for which the FI-approach was originally developed [4, 10]. Due to a relatively low mortality rate, particularly in the validation cohort ($n_{\text{deaths}} = 15$ AD, 10 HC and 20 MCI, respectively), subjects across diagnostic groups were pooled and

the analyses were considered explorative. We hypothesized that FI_t would be associated with mortality risk comparable to published literature, also when adjusting for confounders (age, sex, education, cognitive performance). Here, MMSE was used to account for variation in baseline cognition instead of CDRSB as the latter equals 0 in HC subjects. In the adjusted analyses, we assumed non-proportional hazards and estimated average hazard ratios as described above since Kaplan–Meier plots showed pronounced crossing of the survival curves for FI_s and FI_c (Supplementary Figure S4).

Results

Table 2 summarizes baseline characteristics for the development and validation samples. The validation sample was younger, more highly educated, and included fewer patients living with dementia at baseline compared with the development sample.

FAMD-based subject cluster analysis

We performed FAMD on the 93 health deficit variables identified by the standard procedure which yielded 93 PCs. A plot of the cumulative explained variance of all PCs, and a scree plot of the first 10 are shown in Supplementary Figure S1. The contribution of each continuous and categorical health deficit variable to the first 5 PCs are shown in Supplementary Figures S2 and S3, respectively. FAQ-items contributed most to the 1st PC. Number of prescription drugs contributed the most to the 2nd. The first 60 PCs, explaining 80% of cumulative variance, were fed into cluster analysis. A cluster dendrogram showing the empirical division of individual subjects into our chosen two-cluster solution is shown in Fig. 2. Baseline characteristics of the individuals in each cluster are shown in Table 3. Median degree of frailty, as measured by two standard FIs (FI_s , FI_c), was significantly greater in the smaller cluster (cluster 2, coined “frail”), comprising 230 subjects (149 AD, 81 MCI). The largest cluster, cluster 1 (coined “fit”), consisted of 589 subjects (229 HC, 316 MCI, and 44 AD). The participants in the two clusters were of comparable age, and had similar sex distributions. In cluster 2, more subjects were living with polypharmacy, more reported low levels of energy, and had more symptoms of depression compared with subjects in cluster 1.

Table 2 Baseline characteristics of development and validation samples

Sample	Development (ADNI1, $n = 819$)	Validation (ADNI2 + GO, $n = 815$)	p
Baseline diagnosis, n (%)			< 0.001
AD	193 (23.6%)	151 (18.5%)	
HC	229 (28.0%)	190 (23.3%)	
MCI	397 (48.5%)	474 (58.2%)	
Age in years, median [IQR]	75.5 [71.2;80.0]	72.6 [67.6;77.7]	< 0.001
Sex, n (%)			0.075
Female	342 (41.8%)	377 (46.3%)	
Male	477 (58.2%)	438 (53.7%)	
Education in years, median [IQR]	16.0 [13.0;18.0]	16.0 [14.0;18.0]	< 0.001
MMSE, median [IQR]	27.0 [25.0;29.0]	28.0 [26.0;29.0]	< 0.001
CDRSB, median [IQR]	1.5 [0.0;3.0]	1.0 [0.5;2.5]	0.400
FI_s			
Median [IQR]	0.168 [0.125;0.220]	0.174 [0.133;0.228]	0.019
Mean (SD)	0.176 (0.070)	0.183 (0.067)	
99th percentile	0.358	0.367	
FI_r			
Median [IQR]	0.154 [0.077;0.269]	0.154 [0.096;0.269]	0.125
Mean (SD)	0.187 (0.138)	0.192 (0.131)	
99th percentile	0.615	0.594	
FI_c			
Median [IQR]	0.200 [0.131;0.269]	0.206 [0.150;0.275]	0.349
Mean (SD)	0.211 (0.100)	0.214 (0.097)	
99th percentile	0.469	0.468	

AD=Alzheimer's disease dementia. FI_s =a 93-item frailty index (FI) created according to standard procedure by the authors. FI_r =a 26-item FI created by adding a data-driven supplement to the standard procedure. FI_c =a 40-item FI created according to standard procedure by Canevelli, et al. [26]. HC=cognitively normal control. IQR=interquartile range. MCI=mild cognitive impairment. SD=standard deviation. Sample group differences for continuous and categorical variables were assessed by Kruskal–Wallis and Pearson's chi-squared tests, respectively

Table 4 shows the empirical ranking of health deficit items, including the 26 health deficit variables correlating significantly ($p_{FDR} < 0.05$) with cluster belonging, which were used to compute FI_r . Sixteen out of the 26 health deficits used to compute FI_r were related to activities of daily living (FAQ) and neuropsychiatric symptoms (NPI) reported by next of kin. Three items were based on self-report (feeling depressed, low energy, drowsy). One item was based on depressive symptoms as rated on the geriatric depression scale (GDS), one was having abnormal gait on physical exam, one was number of prescription drugs, two were blood test results (neutrophil count, red blood cell count), the last item was pulse pressure. Except having an item count less than 30, FI_r was deemed in accordance with standard criteria for an FI [10].

Frailty index characteristics

Density plots for all three FI-variables for the development and validation samples are shown in Fig. 3A, whereas central tendency, variability and 99th percentiles are quantified in Table 2. FI_r displayed a pattern of greater right-skew, variability and a higher upper FI-limit compared with standard FIs (FI_s , FI_c). As expected, median FI-scores increased with greater degree of cognitive impairment for all FIs (Fig. 3B). As shown in Supplementary Figure S5, the relationship with age appeared weaker for FI_r compared with standard FIs.

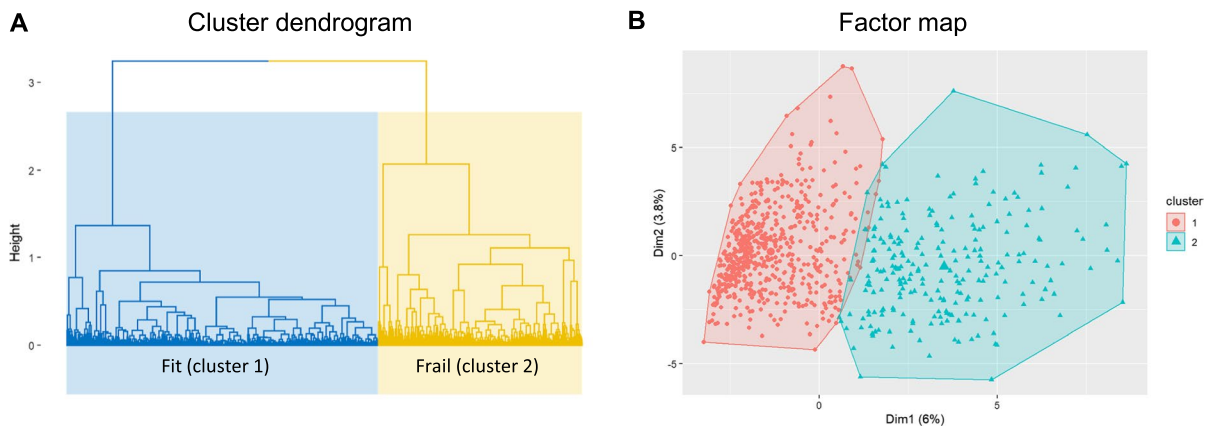


Fig. 2 **A** Dendrogram showing the hierarchical structure of the subject clustering solution. The blue part shows the first and largest cluster which we coined “fit” due to significantly lower frailty scores in this subgroup (c.fr. Table 3) compared with the second, smaller cluster (yellow), coined “frail”. **B**

Scatterplot showing subjects and their cluster belonging based on the two first principal components (PC, “Dim1”, “Dim2”) from factor analysis of mixed data (FAMD). The percentages denote explained variance of each PC

Table 3 Baseline characteristics of subject cluster sub-groups formed by FAMD-based cluster analysis

Cluster	Cluster 1 “Fit” (N= 589)	Cluster 2 “Frail” (N= 230)	p value
Diagnosis			< 0.001
-AD	44 (7.5%)	149 (64.8%)	
-HC	229 (38.9%)	0 (0.0%)	
-MCI	316 (53.7%)	81 (35.2%)	
Age	75.2 [71.3;79.7]	76.40 [70.70;81.80]	0.299
Sex			0.943
-Female	245 (41.6%)	97 (42.2%)	
-Male	344 (58.4%)	133 (57.8%)	
Education, years	16.0 [14.0;18.0]	16.0 [12.0;17.0]	0.001
CDRSB	0.50 [0.00;1.50]	3.50 [2.50;5.00]	< 0.001
FI _s	0.15 [0.11;0.19]	0.22 [0.18;0.27]	< 0.001
FI _c	0.17 [0.12;0.23]	0.30 [0.23;0.37]	< 0.001
Number of prescription drugs]	5.0 [3.0;7.0]	6.0 [4.0;8.0]	< 0.001
Self reported low energy			< 0.001
-Absent	504 (85.6%)	154 (67.0%)	
-Present	85 (14.4%)	76 (33.0%)	
NPI, depressive symptoms			< 0.001
-No	524 (89.1%)	140 (60.9%)	
-Yes	64 (10.9%)	90 (39.1%)	

AD= Alzheimer’s disease dementia. CDRSB=Clinical Dementia Rating Sum of Boxes. FI_s=a 93-item frailty index (FI) created according to standard procedure by the authors. FI_c=a 40-item FI created according to standard procedure by Canevelli, et al. [26]. HC=healthy cognitively normal control. MCI=mild cognitive impairment. Continuous variables are reported as median [interquartile range]. Sample group differences for continuous and categorical variables by Kruskal–Wallis and Pearson’s chi-squared tests, respectively

Table 4 Ranking of health deficit items

Rank	Health deficit variable	Adjusted <i>p</i> value
1	FAQ, writing checks, paying bills, or balancing checkbook	<0.001
2	FAQ, assembling tax records, business affairs, or other papers	<0.001
3	FAQ, shopping alone for clothes, household	<0.001
4	FAQ, traveling out of the neighborhood)	<0.001
5	FAQ, preparing a balanced meal	<0.001
6	FAQ, playing a game of skill such as bridge or chess	<0.001
7	FAQ, paying attention to and understanding a TV program, book, or magazine	<0.001
8	FAQ, heating water, making a cup of coffee	<0.001
9	NPI, anxiety	<0.001
10	NPI, apathy	<0.001
11	NPI, agitation	<0.001
12	NPI, depression	<0.001
13	NPI, irritability	<0.001
14	NPI, change in appetite and eating	<0.001
15	NPI, sleep disturbance	<0.001
16	Self-reported depressed mood	<0.001
17	NPI, aberrant motor behavior	<0.001
18	History of psychiatric disease	<0.001
19	Self-reported low energy	<0.001
20	Self-reported drowsiness	<0.001
21	Geriatric Depression Scale (GDS) score	<0.001
22	Abnormal gait on physical exam	<0.001
23	Number of medications (polypharmacy)	<0.001
24	Blood neutrophil count ($10^3/\mu\text{l}$)	<0.001
25	Blood red blood cell count (RBC; $10^6/\mu\text{l}$)	0.01
26	Pulse pressure (mmHg)	0.03
27	Blood glucose (mg/dL)	0.07
28	Self-reported dizziness	0.07
29	History of endocrine-metabolic disease	0.07
30	Heart rate (count)	0.07
31	Blood hematocrit (%)	0.07
32	Tremor on physical exam	0.12
33	Body mass index (BMI)	0.12
34	Blood mean corpuscular hemoglobin (MCH; pg)	0.16
35	Blood albumin (g/dL)	0.16
36	Blood mean corpuscular volume (MCV; fL)	0.20
37	Blood hemoglobin (g/dL)	0.20
38	Systolic blood pressure (mmHg)	0.24
39	Vision impairment on physical exam	0.28
40	Self-reported muscle pain	0.29
41	Blood alanine aminotransferase (ALT; U/L)	0.29
42	Blood alkaline phosphatase (ALP; U/L)	0.30
43	Blood creatinine (mg/dL)	0.30
44	Cranial nerve abnormality on physical exam	0.30
45	Blood cholesterol (mg/dL)	0.31

Table 4 (continued)

Rank	Health deficit variable	Adjusted <i>p</i> value
46	On medication for elevated cholesterol	0.31
47	On medication for arthritis	0.34
48	Diastolic blood pressure (mmHg)	0.42
49	Self-reported chest pain	0.49
50	On medication for hypertension	0.49
51	Self-reported dry mouth	0.50
52	Blood urea nitrogen (BUN, mg/dL)	0.52
53	Abnormal plantar reflex on physical exam	0.52
54	Mean arterial pressure (mmHg)	0.52
55	Blood triglycerides (mg/dL)	0.52
56	Self-reported recent fall	0.53
57	Auditory impairment on physical exam	0.53
58	History of musculoskeletal disease	0.55
59	History of hepatic disease	0.60
60	On medication for hypothyroidism	0.61
61	Self-reported cough	0.61
62	Abnormal finger to nose test	0.61
63	Self-reported constipation	0.62
64	Blood uric acid (mg/dL)	0.62
65	Self-reported palpitations	0.62
66	Abnormal tendon reflex on physical exam	0.68
67	History of gastrointestinal disease	0.71
68	Abnormal heel-knee test	0.72
69	Motor strength deficit on physical exam	0.72
70	Symptoms from abdomen	0.72
71	Sensory nerve abnormality on physical exam	0.72
72	Self-reported vision disturbance	0.72
73	Self-reported life satisfaction	0.72
74	History of head, eyes, ears, nose, and throat diseases	0.72
75	Self-reported urinary frequency	0.72
76	Self-reported shortness of breath	0.74
77	Blood calcium (mg/dL)	0.77
78	History of dermatologic-connective tissue disease	0.77
79	History of cardiovascular disease	0.77
80	Self-reported ankle swelling	0.77
81	Blood total protein (g/dL)	0.77
82	History of neurological (non-AD) disease	0.80
83	Blood total bilirubin (mg/dL)	0.84
84	Blood vitamin B12 (pg/mL)	0.89
85	History of respiratory disease	0.89
86	History of renal-genitourinary disease	0.89
87	History of other diseases	0.89
88	History of hematopoietic-lymphatic disease	0.89
89	History of major surgical procedures	0.97
90	Blood aspartate aminotransferase (AST; U/L)	0.97
91	History of malignancies	0.97
92	Self-reported urinary discomfort	0.97

Table 4 (continued)

Rank	Health deficit variable	Adjusted <i>p</i> value
93	Self-reported insomnia	0.99

The right column shows false discovery rate (FDR)-adjusted *p* values from regression analyses examining the association between 93 health deficits identified according to a set of standard criteria given by Searle, et al. [10], and subject clusters resulting from factor analysis of mixed data (FAMD)-based clustering analysis. Deficits were ranked according to the *p* values and a horizontal line below item 26 marks the 0.05 *p* value threshold. The top 26 items were used to compute FI_r. *Abbreviations: FAQ*= Functional Assessment Questionnaire. *NPI*= Neuropsychiatric Inventory

Diagnostic performance

Table 5 shows the results from pairwise diagnostic group classifications for the development and validation samples. For classification of HC versus AD, FI_r showed excellent performance in both development and validation samples (AUC 0.95 and 0.93, respectively) with relatively balanced sensitivity and specificity. For all other group classifications, poorly balanced sensitivity and specificity were seen, reflected by the imbalanced diagnostic group sizes. Class imbalance correction by undersampling led to more balanced sensitivity and specificity metrics, while overall model performance remained largely unchanged (Supplementary Table S3). For HC versus MCI, both FI_s and FI_c showed poor classification performance in both samples, whereas FI_r showed acceptable performance. For all comparisons, including results obtained by adjusting for class imbalance, classification performance was greater for FI_r compared with both FI_s and FI_c ($p < 0.001$). Assessment of discriminative ability across age and sex-strata (Supplementary Table S4) suggested poorer performance for FI_s in younger old (below 75 years) compared with old (75+) age groups, particularly for males. A similarly tendency was seen for FI_c, whereas overall performance for FI_r remained excellent across age and sex-strata. For females aged 75 and older, FI_r AUC for HC versus AD classification was outstanding (0.97, 95% CI 0.95 to 0.99).

Prognostic performance

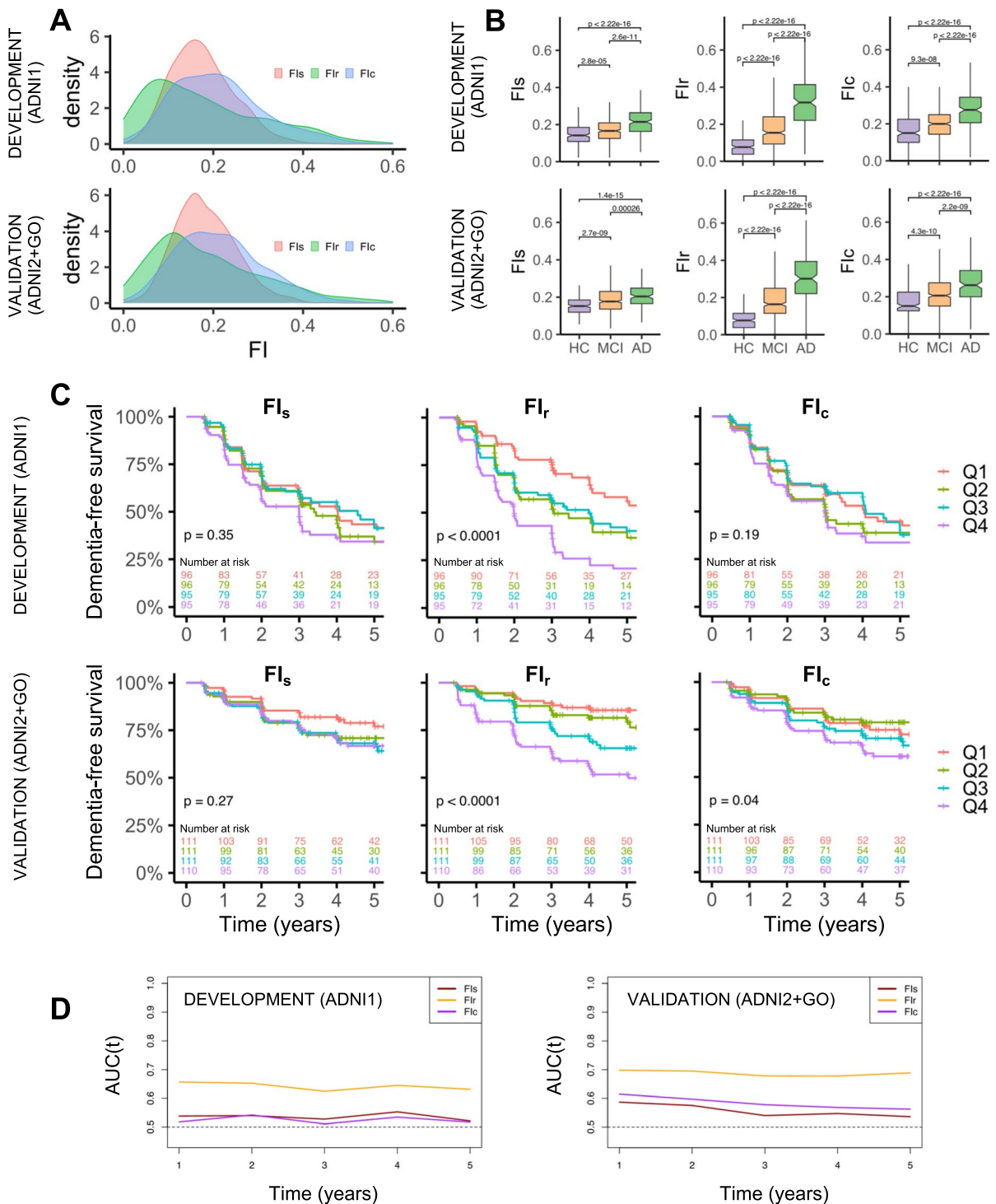
In the development sample, 382 out of 397 subjects with MCI at baseline had diagnostic follow-up data, with a median follow-up time of 744 days (interquartile range, 410 to 1461); 205 (54%) converted to dementia. In the validation sample, 443 out of 474 subjects with MCI at baseline had diagnostic follow-up data, with a median follow-up time of 1451 days (interquartile range, 731 to 2195); 111 (25%) converted to dementia.

Figure 3D shows Kaplan–Meier survival curves illustrating the probability of remaining stable MCI (dementia-free) at time of follow-up for different FI-quartiles calculated for each sample. The FI-quartile curves differed significantly in terms of future dementia risk across samples for FI_r only. Estimates of AUC(t) for prediction of AD conversion in subjects with MCI at baseline over 5-year follow-up for the three continuous FI variables are shown in Fig. 4E. AUC(t) 95% confidence bands (Supplementary Figure S6) for FI_s and FI_c both intersected the 0.5-line, suggesting poor to chance level discrimination. Average AUC(t) for FI_r ranged from 0.62–0.70 over time suggesting poor to acceptable performance in prediction of future dementia risk, outperforming FI_c and FI_s from year 2 in both samples (Supplementary Figure S7). In the multivariate survival analyses (Table 6), only FI_r associated with future dementia risk across samples and in a model adjusting for cognitive and functional baseline performance (as scored by CDRSB).

In exploratory analyses, we found a crude mortality hazard ratio for FI_r of 1.04 (1.02 to 1.05) which is in accordance with published estimates from FI meta-analysis (HR 1.04, 95% CI 1.03 to 1.04) [32]. Kaplan–Meier survival curves and associations with mortality risk for all three FI-variables (FI_s, FI_r, FI_c) are shown in Supplementary Figure S4 and Supplementary Table S5. In the Development sample, only FI_r associated with mortality risk (average hazard ratio 1.02, 95% CI 1.01 to 1.04) in the fully adjusted model.

Discussion

Several studies show promise for FIs in dementia risk prediction [2, 5, 6, 12], but the results have been variable and validation studies are lacking. We tested whether adding a data-driven health deficit selection step to the standard FI procedure improves prediction of current cognitive status and future conversion to dementia. The data-driven optimization procedure



may help researchers and clinicians streamline FI development and improve early detection of MCI and

AD and identify those at highest risk for clinical trial enrolment.

Fig. 3 **A** Density plots showing three frailty index (FI) distributions for ADNI1 (development) and ADNI2/GO (validation) cohorts. FI_s = a 93-item FI created according to standard procedure by the authors. FI_r = a 26-item FI created by adding a data-driven supplement to the standard procedure. FI_c = a 40-item FI created according to standard procedure by Canevelli, et al. [26]. **B** Boxplots illustrating central tendency and variability of the three different FI-variables for cognitively normal (healthy) controls (HC), and subjects living with mild cognitive impairment (MCI) or Alzheimer’s disease dementia (AD). *P* values are from Wilcoxon rank sum tests comparing diagnostic group differences in FI-scores. **C** Kaplan–Meier survival curves for sample quartiles calculated for each FI. The survival probabilities indicate the probability of remaining stable MCI at time of follow-up, and vertical lines through each line indicate censoring. **D** Estimated mean AUC(t) for prediction of AD conversion in subjects with MCI at baseline plotted over 5 years of follow-up for the three rival continuous FI variables in development and validation cohorts

Our results show that the data-driven FI_r outperforms standard FIs in terms of diagnostic and prognostic performance, both in development and validation samples. As an example of diagnostic performance, Canevelli and coworkers reported an AUC of 0.67 to 0.75 for a standard FI in discriminating participants with and without dementia [26]. Classification performance for FI_r in discriminating subjects with normal cognition and dementia was 0.95 to 0.96. While FI_r diagnostic performance remained excellent (mean AUCs from 94 to 97) across age and sex-strata, overall performance dropped in younger strata for standard FIs (FI_s , FI_c). As the FAMD-based clustering approach employed here aims to maximize explained variance of included health deficits, overlap with confounding entities is likely diminished.

In terms of prognostic performance, only FI_r associated with future dementia conversion when adjusting for cognitive status. Some studies of frailty and future dementia risk have not adjusted for baseline cognitive status [2, 6], others have shown conflicting results [5, 12, 33]. The present study is the first to include out-of-sample verification. Our data show that two FIs created according to standard procedure had little added value in prediction of dementia when adjusting for cognitive performance, even when health deficits strongly related to dementia (such as certain ADL-items and cognitive tests) were excluded. The results show that dementia prediction ability of the accumulation of deficits model is variable and depends on the way an FI is constructed. Our proposed data-driven assessment of health deficits revealed items particularly sensitive to

AD-related cognitive status and conversion to dementia. We argue that the standard procedure for creating an FI published in 2008 [10] could benefit from revision and that health deficit selection may be streamlined by FAMD-based subject clustering.

Characteristics of health deficits identified by cluster analysis

In line with previous studies, higher FI-scores were associated with greater degree of cognitive impairment [5, 33, 34]. FAQ items were amongst the top health deficit variables that differed most between cluster subgroups, including independence in assembling tax records, business affairs, or other papers, and writing checks, paying bills, or balancing checkbooks. Previous studies confirm that FAQ is sensitive to early cognitive decline [35, 36]. Polypharmacy (more than five different prescription medications daily) was also among the highest ranked health deficit variables following cluster analysis, a feature of frailty previously associated with dementia risk [37]. Cardiovascular disease and frailty are closely linked and may share similar causal mechanisms [38]. As such, most FIs to date consist of one or more items involving the cardiovascular system, including a history of ischemic heart disease, stroke, or heart failure [e.g., 2, 5, 16, 21, 26, 34]. Some also include biomarkers thereof, such as blood pressure [21, 26]. The question here seems not to be *whether* cardiovascular health deficits should be included in an FI, but rather *which* one to choose in order to maximize FI performance. While the standard procedure for creating an FI gives little advice on *optimal* item selection [10], the present data-driven approach ruled out all cardiovascular items except pulse pressure (PP). Often used as a surrogate of arterial stiffness, increased PP is a feature of aging that has been associated with blood–brain barrier dysfunction and cognitive impairment [39]. Two items representing hematopoietic and immune systems were also among the top-rated health deficits variables (Table 4) and were included in FI_r : blood neutrophil count and red blood cell count (RBC). Neutrophils are the most abundant leukocyte in the periphery and are gaining increasing attention as a prognostic AD biomarker [40]. Neutrophils are hypothesized to contribute to AD progression through systemic inflammation and disturbance of the blood–brain barrier [41]. RBC is one of several red

Table 5 Pairwise group classification

Classification/performance measure	Sample					
	Development (ADNI1)			Validation (ADNI2 + GO)		
	FI-variable					
	FI _s	FI _r	FI _c	FI _s	FI _r	FI _c
<i>HC vs. AD</i>						
Area under the curve	0.75 (0.00)	0.95 (0.00)	0.81 (0.00)	0.75 (0.00)	0.93 (0.00)	0.80 (0.00)
Sensitivity	0.81 (0.00)	0.94 (0.00)	0.83 (0.00)	0.82 (0.00)	0.95 (0.00)	0.82 (0.00)
Specificity	0.56 (0.01)	0.77 (0.00)	0.64 (0.00)	0.53 (0.00)	0.76 (0.00)	0.58 (0.00)
PPV	0.69 (0.00)	0.83 (0.00)	0.74 (0.00)	0.69 (0.00)	0.83 (0.00)	0.71 (0.00)
NPV	0.71 (0.01)	0.92 (0.00)	0.77 (0.01)	0.70 (0.00)	0.92 (0.00)	0.72 (0.00)
F1-score	0.74 (0.00)	0.88 (0.00)	0.78 (0.00)	0.75 (0.00)	0.89 (0.00)	0.76 (0.00)
<i>MCI vs. AD</i>						
Area under the curve	0.67 (0.00)	0.81 (0.00)	0.71 (0.00)	0.60 (0.00)	0.76 (0.00)	0.66 (0.00)
Sensitivity	0.97 (0.00)	0.89 (0.00)	0.92 (0.00)	0.95 (0.00)	0.89 (0.00)	0.91 (0.00)
Specificity	0.11 (0.01)	0.47 (0.00)	0.24 (0.00)	0.08 (0.00)	0.42 (0.00)	0.23 (0.00)
PPV	0.69 (0.00)	0.78 (0.00)	0.71 (0.00)	0.76 (0.00)	0.83 (0.00)	0.79 (0.00)
NPV	0.66 (0.05)	0.69 (0.01)	0.61 (0.02)	0.32 (0.00)	0.55 (0.00)	0.44 (0.00)
F1-score	0.81 (0.00)	0.83 (0.00)	0.80 (0.00)	0.85 (0.00)	0.86 (0.00)	0.84 (0.00)
<i>HC vs. MCI</i>						
Area under the curve	0.60 (0.00)	0.78 (0.00)	0.63 (0.00)	0.65 (0.00)	0.78 (0.00)	0.65 (0.00)
Sensitivity	0.00 (0.00)	0.41 (0.00)	0.06 (0.01)	0.01 (0.00)	0.31 (0.00)	0.07 (0.01)
Specificity	1.00 (0.00)	0.89 (0.00)	0.97 (0.00)	1.00 (0.00)	0.92 (0.00)	0.97 (0.00)
PPV	0.15 (0.03) [§]	0.68 (0.01)	0.57 (0.06)	1.00 (0.00)	0.60 (0.00)	0.45 (0.00)
NPV	0.63 (0.00)	0.73 (0.00)	0.64 (0.00)	0.71 (0.00)	0.77 (0.00)	0.72 (0.00)
F1-score	0.15 (0.03) [§]	0.51 (0.00)	0.18 (0.00) [§]	0.01 (0.00)	0.41 (0.01)	0.12 (0.05)

AD=Alzheimer's disease dementia, AUC=area under the curve. FI_s=a 93-item frailty index (FI) created according to standard procedure by the authors. FI_r=a 26-item FI created by adding a data-driven supplement to the standard procedure. FI_c=a 40-item FI created according to standard procedure by Canevelli, et al. [26]. HC=healthy cognitively normal control. MCI=mild cognitive impairment. NPV=negative predictive value. PPV=positive predictive value

[§]Error margin is reported as standard error

blood cell indices associated with AD and cognitive decline [42], and may reflect an array of pathological disturbances affecting brain function, i.e., B-vitamin deficiencies [43], anemia [44], and chronic kidney disease [45].

Cluster analysis revealed a higher degree of depressive symptomatology in the smaller, "frail" cluster 2, compared with cluster 1 (see Table 3). In turn, FI_r included both self-reported depressed mood, the NPI depression-item and GDS score as health deficits, reinforcing the strong link between frailty and depressive syndromes [46]. Depression is also a common manifestation in AD with prevalence estimates up to 50% [47], with associations with AD pathology including amyloid- β accumulation [48]. The present findings are intriguing as the ADNI

study was designed to rule out subjects with clinical depression, and points to an important role even for subclinical symptoms. Indeed, one study found that even subthreshold symptoms of geriatric depression were related to AD-related neurodegeneration, which appeared to be independent of amyloid burden [49]. Self-reported low levels of energy and abnormal gait were also among the top variables that differed between the cluster subgroups. These characteristic frailty components might represent targets of prevention, as a randomized controlled clinical trial found that exercise was effective in reducing cognitive frailty [50]. Overall, the clustering approach identified a nuanced pattern of frailty-related health factors jointly contributing to the predictive ability of FI_r.

Table 6 Associations between three different FI-variables and conversion to dementia within follow-up for subjects with MCI at baseline

Sample/FI-variable*	Model 1 ^a		Model 2 ^b	
<i>Development</i> (ADNI1)	HR (95% CI)	<i>p</i> value	HR (95% CI)	<i>p</i> value
FI _s	1.02 (1.00–1.04)	0.04	1.00 (0.98–1.03)	0.68
FI _r	1.03 (1.02–1.05)	< 0.001	1.02 (1.01–1.03)	< 0.001
FI _c	1.01 (1.00–1.03)	0.06	1.00 (0.99–1.02)	0.72
<i>Validation</i> (ADNI2/GO) ^c	AHR (95% CI)	<i>p</i> value	AHR (95% CI)	<i>p</i> value
FI _s	1.02 (0.99–1.05)	0.19	0.99 (0.96–1.02)	0.40
FI _r	1.04 (1.03–1.06)	< 0.001	1.02 (1.00–1.04)	0.02
FI _c	1.02 (1.00–1.05)	0.04	1.00 (0.97–1.02)	0.80

AHR=average hazard ratio. HR=hazard ratio. CI=confidence interval. FI_s=a 93-item frailty index (FI) created according to standard procedure by the authors. FI_r=a 26-item FI created by adding a data-driven supplement to the standard procedure. FI_c=a 40-item FI created according to standard procedure by Canevelli, et al. [26]

*All FI-variables were multiplied by 100 before entered into the models

^aModel 1 included age, sex, and education as covariates

^bModel 2 included age, sex, education, and Clinical Dementia Rating scale, Sum of Boxes (CDRSB)

^cDue to non-proportionality of hazards for age in model 1 and 2 of the validation cohort, we estimated average hazard ratios (AHRs) instead of HRs using Prentice weights with censoring correction and robust variance estimation

The relationship between frailty and future dementia risk varies across FI-variables

The degree of frailty predicted conversion from MCI to dementia across samples for the data-driven FI_r only. Although standard generated FIs (FI_s, FI_c) associated with future dementia risk in age, sex, and education-adjusted models, these results failed out-of-sample validation and the association did not remain when adjusting for baseline CDRSB scores. We are aware of few FI-studies of dementia risk with external validation, and most lack correction for baseline cognition using more comprehensive tools such as CDRSB. Many geriatric outpatient clinics have started implementing frailty assessment by means of FIs created using standard procedure. Given that validated tools for cognitive and functional assessment such as CDR exist, our findings question the addition of time-consuming assessment by means of standard FIs for clinical dementia workup unless their predictive abilities are improved following, e.g., a data-driven optimization procedure.

Several studies have employed prediction models of dementia due to AD [51], but to the best of our knowledge, this is the first study that does so based on subject clustering of health deficits related to frailty.

Interestingly, only FI_r and not FI_s and FI_c were predictive of conversion to AD when including CDRSB as an additional covariate, suggesting that the clustering approach to selecting health deficits yields a frailty measure with added predictive value beyond global cognitive functioning. Based on the specific health differences between cluster subgroups, this finding is in line with the literature. For instance, differences in low energy and gait abnormalities fit well with the frailty construct, and studies indicating that physical activity associates with AD risk [e.g., 52].

Limitations

The current findings should be interpreted with the following limitations taken into consideration. The selection of health variables was based on availability in the ADNI database, which was not originally designed to estimate frailty. Thus, our selection did not include certain phenotypical frailty measures such as weight loss, poor grip strength, or walking speed. The cut-points for the individual health deficits used to estimate FI are somewhat arbitrary, such as for blood pressure, and also applying data-driven approaches to different reference ranges might have improved the utility of the FI [53].

Another limitation is the lack of validation against traditional frailty endpoints, such as mortality, hospitalizations and falls. In exploratory analyses, we found that crude FI_r HR-estimates for mortality in ADNI were comparable to those in the literature [32]. Although FI_r was the only FI-variable correlating with mortality in fully adjusted analysis (Supplemental Table S5), superiority in the prediction of mortality and other frailty endpoints needs to be assessed in future research employing larger-scale databases. Furthermore, the dataset employed here did not allow for comparison of our data-driven revision of the accumulation of deficits model with results obtained by using the rival phenotype or physical frailty model [54].

To our knowledge, using FAMD-based subject clustering as an approach for selecting out features, such as health deficits is novel. In particular, selecting health deficits based on regression analyses of cluster belonging using FDR-adjusted p-value threshold has not been tested before. The sensitivity to differences in sample size might call for further development of our method using unbiased deficit selection for different population and sample sizes. One could even argue that a more appropriate approach would be to select features using a supervised approach [55]. However, an advantage to our unsupervised approach is that it more likely captures variance related to frailty per se compared to supervised methods, and is not bound to categorical definitions of diagnoses. Another advantage is that the present approach does not need information about the primary endpoint (e.g., mortality, incident dementia) for model training, importantly enabling FI development on novel datasets and clinical cohorts where prospective endpoint data are not yet available.

Conclusion

Adding a data-driven supplemental step to the standard procedure for creating an FI improves prediction of cognitive status and future dementia risk. While the data-driven procedure employed here reduced the number of items included in an FI, the remaining selection adhered well with standard criteria outlined by Searle and colleagues [10] and included items reflecting core components of the frailty construct [1].

The two identified subject clusters from cluster analysis showed a unique constellation of health deficits which contributed to the stronger predictive ability of diagnosis and disease progression. In particular, our data-driven clustering analysis suggested a strong contribution of activities of daily living, polypharmacy, and tests reflecting immune, hematopoietic and cardiovascular systems, as well as several items of depressive symptomatology in AD risk stratification, even within a sample well-screened to rule out clinical depression. The results point to frailty—when measured using an FI with data-driven health deficit assessment—as a putative modifiable AD risk factor. The proposed data-driven procedure warrants further testing on other often-used frailty endpoints, such as mortality.

Acknowledgements Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Authors' contributions A.E., L.T.W., and L.A.M. conceptualized the study. A.E., N.T.D., and L.A.M. performed investigation, visualization, and analysis. L.T.W. supervised the study. A.E. and L.A.M. wrote the original draft. A.E. revised, edited, and submitted the final manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by University of Oslo (incl Oslo University Hospital) The European Research Council under the European Union's Horizon 2020 research and Innovation program (ERC StG, Grant 802998). This work was performed on the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT) (tsd-drift@usit.uio.no). Computations were also performed on resources provided by UNINETT Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway.

Declarations

Competing interests The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *The Lancet*. 2013;381(9868):752–62. [https://doi.org/10.1016/s0140-6736\(12\)62167-9](https://doi.org/10.1016/s0140-6736(12)62167-9).
- Song X, Mitnitski A, Rockwood K. Age-related deficit accumulation and the risk of late-life dementia. *Alz Res Therapy*. 2014;6(54). <https://doi.org/10.1186/s13195-014-0054-5>.
- Engvig A, Wyller TB, Skovlund E, Ahmed MV, Hall TS, Rockwood K, Njaastad AM, Neerland BE. Association between clinical frailty, illness severity and post-discharge survival: a prospective cohort study of older medical inpatients in Norway. *Eur Geriatr Med*. 2022;13(2):453–61. <https://doi.org/10.1007/s41999-021-00555-8>.
- Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of Deficits as a Proxy Measure of Aging. *Sci World J*. 2001;1:323–36. <https://doi.org/10.1100/tsw.2001.58>.
- Trebbastoni A, Canevelli M, D'Antonio F, Imbriano L, Podda L, Rendace L, et al. The impact of frailty on the risk of conversion from mild cognitive impairment to Alzheimer's disease: evidences from a 5-year observational study. *Front Med*. 2017;4(178). <https://doi.org/10.3389/fmed.2017.00178>.
- Ward DD, Wallace LMK, Rockwood K. Cumulative health deficits, APOE genotype, and risk for later-life mild cognitive impairment and dementia. *J Neurol Neurosurg Psychiatry*. 2021;92(2):136–42. <https://doi.org/10.1136/jnnp-2020-324081>.
- Puts MTE, Toubasi S, Andrew MK, Ashe MC, Ploeg J, Atkinson E, Ayala AP, Roy A, Rodríguez Monforte M, Bergman H, et al. Interventions to prevent or reduce the level of frailty in community-dwelling older adults: a scoping review of the literature and international policies. *Age Ageing*. 2017. <https://doi.org/10.1093/ageing/afw247>.
- Theou O, Van Der Valk AM, Godin J, Andrew MK, McElhaney JE, McNeil SA, Rockwood K. Exploring clinically meaningful changes for the frailty index in a longitudinal cohort of hospitalized older patients. *J Gerontol: Series A*. 2020;75(10):1928–34. <https://doi.org/10.1093/gerona/glaa084>.
- Theou O, Jayanama K, Fernández-Garrido J, Buigues C, Pruijboom L, Hoogland AJ, et al. Can a prebiotic formulation reduce frailty levels in older people? *J Frailty Aging*. 2019;8:48–52. <https://doi.org/10.14283/jfa.2018.39>.
- Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8(1):24. <https://doi.org/10.1186/1471-2318-8-24>.
- Rockwood K, Howlett SE. Age-related deficit accumulation and the diseases of ageing. *Mech Ageing Dev*. 2019;180:107–16. <https://doi.org/10.1016/j.mad.2019.04.005>.
- Ward DD, Wallace LMK, Rockwood K. Frailty and risk of dementia in mild cognitive impairment subtypes. *Ann Neurol*. 2021;89(6):1221–5. <https://doi.org/10.1002/ana.26064>.
- Chao YS, Wu CJ, Wu HC, et al. Composite diagnostic criteria are problematic for linking potentially distinct populations: the case of frailty. *Sci Rep*. 2020;10(2601). <https://doi.org/10.1038/s41598-020-58782-1>.
- Chao Y-S, Wu H-C, Wu C-J, Chen W-C. Index or illusion: The case of frailty indices in the Health and Retirement Study. *PLoS ONE*. 2018;13(7):e0197859. <https://doi.org/10.1371/journal.pone.0197859>.
- Song X, Mitnitski A, Macknight C, Rockwood K. Assessment of individual risk of death using self-report data: an artificial neural network compared with a frailty index. *J Am Geriatr Soc*. 2004;52(7):1180–4. <https://doi.org/10.1111/j.1532-5415.2004.52319.x>.
- Lin S-Y, Lee W-J, Chou M-Y, Peng L-N, Chiou S-T, Chen L-K. Frailty index predicts all-cause mortality for middle-aged and older taiwanese: implications for active-aging programs. *PLoS ONE*. 2016;11(8):e0161456. <https://doi.org/10.1371/journal.pone.0161456>.
- Chao Y-S, Wu C-J. Principal component-based weighted indices and a framework to evaluate indices: results from the Medical Expenditure Panel Survey 1996 to 2011. *PLoS ONE*. 2017;12(9):e0183997. <https://doi.org/10.1371/journal.pone.0183997>.
- Pagès J. Analyse factorielle de données mixtes. *Revue de Statistique Appliquée*. 2004;52(4):93–111.
- Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 2008;25(1):1–18. <https://doi.org/10.18637/jss.v025.i01>.

20. Passarino G, Montesanto A, De Rango F, Garasto S, Berardelli M, Domma F, Mari V, Feraco E, Franceschi C, De Benedictis G. A cluster analysis to define human aging phenotypes. *Biogerontology*. 2007;8(3):283–90. <https://doi.org/10.1007/s10522-006-9071-5>.
21. Wallace LMK, Theou O, Godin J, Andrew MK, Bennett DA, Rockwood K. Investigation of frailty as a moderator of the relationship between neuropathology and dementia in Alzheimer's disease: a cross-sectional analysis of data from the Rush Memory and Aging Project. *The Lancet Neurology*. 2019;18(2):177–84. [https://doi.org/10.1016/S1474-4422\(18\)30371-5](https://doi.org/10.1016/S1474-4422(18)30371-5).
22. Franzin A, Sambo F, Di Camillo B. bnstruct: an R package for Bayesian Network structure learning in the presence of missing data. *Bioinformatics*. 2016;33(8):1250–52. <https://doi.org/10.1093/bioinformatics/btw807>.
23. de Bont J, Márquez S, Fernández-Barrés S, Warembourg C, Koch S, Persavento C, Fochs S, Pey N, de Castro M, Fossati S, et al. Urban environment and obesity and weight-related behaviours in primary school children. *Environ Int*. 2021;155:106700. <https://doi.org/10.1016/j.envint.2021.106700>.
24. Han L, Shen P, Yan J, Huang Y, Ba X, Lin W, et al. Exploring the clinical characteristics of COVID-19 clusters identified using factor analysis of mixed data-based cluster analysis. *Front Med*. 2021;8(644724). <https://doi.org/10.3389/fmed.2021.644724>.
25. Husson F, Josse J, Pages J. Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data. Technical report. Agrocampus Ouest, Applied Mathematics Department. 2010;1–17. http://www.sthda.com/english/upload/hcpc_husson_josse.pdf.
26. Canevelli M, Arisi I, Bacigalupo I, Arighi A, Galimberti D, Vanacore N, D'Onofrio M, Cesari M, Bruno G. Biomarkers and phenotypic expression in Alzheimer's disease: exploring the contribution of frailty in the Alzheimer's Disease Neuroimaging Initiative. *GeroScience*. 2021;43(2):1039–51. <https://doi.org/10.1007/s11357-020-00293-y>.
27. Fried TR, O'Leary J, Towle V, Goldstein MK, Trentalange M, Martin DK. Health Outcomes associated with polypharmacy in community-dwelling older adults: a systematic review. *J Am Geriatr Soc*. 2014;62(12):2261–72. <https://doi.org/10.1111/jgs.13153>.
28. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. 2020. <https://www.tidymodels.org>.
29. Therneau TM. A package for survival analysis in R. R package version 3.4-0. 2022. <https://CRAN.R-project.org/package=survival>.
30. Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381–97. <https://doi.org/10.1002/sim.5958>.
31. Schemper M, Wakounig S, Heinze G. The estimation of average hazard ratios by weighted Cox regression. *Stat Med*. 2009;28(19):2473–89. <https://doi.org/10.1002/sim.3623>.
32. Kojima G, Iliffe S, Walters K. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age Ageing*. 2017;47(2):193–200. <https://doi.org/10.1093/ageing/afx162>.
33. Gray SL, Anderson ML, Hubbard RA, LaCroix A, Crane PK, McCormick W, Bowen JD, McCurry SM, Larson EB. Frailty and incident dementia. *J Gerontol A Biol Sci Med Sci*. 2013;68(9):1083–90. <https://doi.org/10.1093/geronol/glt013>.
34. Kelaiditi E, Canevelli M, Andrieu S, Del Campo N, Soto ME, Vellas B, Cesari M. Frailty Index and cognitive decline in Alzheimer's disease: data from the impact of cholinergic treatment use study. *J Am Geriatr Soc*. 2016;64(6):1165–70. <https://doi.org/10.1111/jgs.13956>.
35. Brown PJ, Devanand DP, Liu X, Caccappolo E. Alzheimer's Disease Neuroimaging I: Functional impairment in elderly patients with mild cognitive impairment and mild Alzheimer disease. *Arch Gen Psychiatry*. 2011;68(6):617–26. <https://doi.org/10.1001/archgenpsychiatry.2011.57>.
36. Li K, Chan W, Doody RS, Quinn J, Luo S. Alzheimer's Disease Neuroimaging I: Prediction of conversion to Alzheimer's disease with longitudinal measures and time-to-event data. *J Alzheimers Dis*. 2017;58(2):361–71. <https://doi.org/10.3233/JAD-161201>.
37. Kristensen RU, Norgaard A, Jensen-Dahm C, Gasse C, Wimberley T, Waldemar G. Polypharmacy and potentially inappropriate medication in people with dementia: a nationwide study. *J Alzheimers Dis*. 2018;63(1):383–94. <https://doi.org/10.3233/JAD-170905>.
38. Ferrucci L, Fabbri E. Inflammaging: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nat Rev Cardiol*. 2018;15(9):505–22. <https://doi.org/10.1038/s41569-018-0064-2>.
39. Levin RA, Carnegie MH, Celermajer DS. Pulse pressure: an emerging therapeutic target for dementia. *Front Neurosci*. 2020;14:669. <https://doi.org/10.3389/fnins.2020.00669>.
40. Zhang YR, Wang JJ, Chen SF, Wang HF, Li YZ, Ou YN, Huang SY, Chen SD, Cheng W, Feng JF, et al. Peripheral immunity is associated with the risk of incident dementia. *Mol Psychiatry*. 2022;27(4):1956–62. <https://doi.org/10.1038/s41380-022-01446-5>.
41. Dong Y, Lagarde J, Xicota L, Corne H, Chantran Y, Chaigneau T, Crestani B, Bottlaender M, Potier MC, Aucouturier P, et al. Neutrophil hyperactivation correlates with Alzheimer's disease progression. *Ann Neurol*. 2018;83(2):387–405. <https://doi.org/10.1002/ana.25159>.
42. Winchester LM, Powell J, Lovestone S, Nevado-Holgado AJ. Red blood cell indices and anaemia as causative factors for cognitive function deficits and for Alzheimer's disease. *Genome Medicine*. 2018;10(1):51. <https://doi.org/10.1186/s13073-018-0556-z>.
43. Mikkelsen K, Stojanovska L, Tangalakis K, Bosevski M, Apostolopoulos V. Cognitive decline: a vitamin B perspective. *Maturitas*. 2016;93:108–13. <https://doi.org/10.1016/j.maturitas.2016.08.001>.
44. Weiss A, Beloosesky Y, Gingold-Belfer R, Leibovici-Weissman Y, Levy Y, Mulla F, Issa N, Boltin D, Koren-Morag N, Meyerovitch J, et al. Association of anemia with dementia and cognitive decline among

- community-dwelling elderly. *Gerontology*. 2022. <https://doi.org/10.1159/000522500>.
45. Deckers K, Camerino I, van Boxtel MP, Verhey FR, Irving K, Brayne C, Kivipelto M, Starr JM, Yaffe K, de Leeuw PW, et al. Dementia risk in renal dysfunction: A systematic review and meta-analysis of prospective studies. *Neurology*. 2017;88(2):198–208. <https://doi.org/10.1212/WNL.0000000000003482>.
 46. Lohman M, Dumenci L, Mezuk B. Depression and frailty in late life: evidence for a common vulnerability. *J Gerontol B Psychol Sci Soc Sci*. 2016;71(4):630–40. <https://doi.org/10.1093/geronb/gbu180>.
 47. Cortes N, Andrade V, Maccioni RB. Behavioral and neuropsychiatric disorders in Alzheimer's disease. *J Alzheimers Dis*. 2018;63(3):899–910. <https://doi.org/10.3233/JAD-180005>.
 48. Chi S, Yu JT, Tan MS, Tan L. Depression in Alzheimer's disease: epidemiology, mechanisms, and management. *J Alzheimers Dis*. 2014;42(3):739–55. <https://doi.org/10.3233/JAD-140324>.
 49. Donovan NJ, Hsu DC, Dagley AS, Schultz AP, Amariglio RE, Mormino EC, Okereke OI, Rentz DM, Johnson KA, Sperling RA, et al. Depressive symptoms and biomarkers of Alzheimer's disease in cognitively normal older adults. *J Alzheimers Dis*. 2015;46(1):63–73. <https://doi.org/10.3233/jad-142940>.
 50. Liu Z, Hsu F-C, Trombetti A, King AC, Liu CK, Manini TM, et al. Effect of 24-month physical activity on cognitive frailty and the role of inflammation: the LIFE randomized clinical trial. *BMC Med*. 2018;16(185). <https://doi.org/10.1186/s12916-018-1174-8>.
 51. Hou X-H, Feng L, Zhang C, Cao X-P, Tan L, Yu J-T. Models for predicting risk of dementia: a systematic review. *J Neurol Neurosurg Psychiatry*. 2019;90(4):373. <https://doi.org/10.1136/jnnp-2018-318212>.
 52. Scarmeas N, Luchsinger JA, Schupf N, Brickman AM, Cosentino S, Tang MX, Stern Y. Physical activity, diet, and risk of Alzheimer disease. *JAMA*. 2009;302(6):627–37. <https://doi.org/10.1001/jama.2009.1144>.
 53. Stubbings G, Farrell S, Mitnitski A, Rockwood K, Rutenberg A. Informative frailty indices from binarized biomarkers. *Biogerontology*. 2020;21(3):345–55. <https://doi.org/10.1007/s10522-020-09863-1>.
 54. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146–56. <https://doi.org/10.1093/gerona/56.3.m146>.
 55. Zucchelli A, Marengoni A, Rizzuto D, Calderón-Laraña A, Zucchelli M, Bernabei R, Onder G, Fratiglioni L, Vetrano DL. Using a genetic algorithm to derive a highly predictive and context-specific frailty index. *Aging*. 2020;12(8):7561–75. <https://doi.org/10.18632/aging.103118>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.