Frailty Outcomes And Risk With Alzheimer's Related Dementia (FORWARD)

Statistical Analysis Plan (SAP)

Version 1.0 — November 3rd, 2025

NTC number: NCT06570109

Unique Protocol ID: N-20230068

Written by: Johannes Riis, MD, PhD-student.

Department of Geriatric Medicine, Aalborg University Hospital

Table of contents

1 Primary analysis: Association between frailty (Fried Phenotype) and ADL decline

1.1 Objectives & Hypotheses

1.1.1 Primary objective:

Estimate the association between baseline frailty (Fried Frailty Phenotype) and 12-month change in informant-rated Functional Activities Questionnaire score as a measure of IADL (FAQ-IADL).

The underlying hypothesis is that participants who are frail at baseline will have greater worsening (increase) in FAQ-IADL over 12 months than non-frail. The magnitude of the effect will be exceed minimal clinically significant change in FAQ-IADL over 1 year of follow- of 3 points

1.1.2 Secondary objectives:

1) Association between the primary exposure (Fried Frailty Scale) and self-report IADL function
2) Association between the primary exposure (Fried Frailty Scale) and cognitive decline measured by the Montreal Cognitive Assessment (MoCA).
3) Association between the primary exposure (Fried Frailty Scale) and informant-reported quality of life measured by the Quality of Life in Alzheimer's Disease questionnaire (QoL-AD).
4) Association between the primary exposure (Fried Frailty Scale) and self-reported quality of life measured by the Quality of Life in Alzheimer's Disease questionnaire (QoL-AD).

1.2 Outcomes

1.2.1 Functional ability

The primary outcome, informant reported FAQ-IADL, was chosen based on recommendations for outcomes measures in Alzheimer's disease with mild dementia (1). FAQ-IADL is a ordinal scale with 10 items and possible total scores from 0-30. Each item is scored on a 0–3-point scale, with 0 indicating that the participant has normal function on the item, 1 indicating that the participant can do it themselves with some difficulty, 2 indicating that the participant can do it with help, and 3 indicating dependence on others. FAQ-IADL has good reliability and is sensitive to changes, with a change of 3 points or more determined as clinically relevant.(2–5)

There is also a possibility of reporting that the item has not been performed and therefore is not relevant. Here the item is scored either as 0 if the participant has never done the activity but would be able to do it now, or 1 if the participant as never done the activity and would have difficulty now.

By the time of writing of this analysis (interim baseline data on 60 participants, no outcome data available), the frequency of having reached ceiling on the FAQ-IADL was 1 of 60. Ceiling here is defined as not being able to score 3 or more points higher on FAQ-IADL at follow-up. This number takes into account that the true maximum for some participants may be lower than 30 as some activities may never have been performed.

While informant reported FAQ-IADL is the primary outcome, self-reported FAQ-IADL is a secondary outcome.

1.2.2 Montreal Cognitive Assessment (MoCA)

A secondary outcome of cognitive status will be measured by the MoCA, which was reference papers on outcomes for Alzheimers disease, also over the most used test, the Mini Mental State Examination (1,6). The MoCA provides a global assessment of cognition across memory, visuospatial skills, verbal fluency, attention, and executive function, is fast to administer, and requires a minimum of training. The MoCA is more sensitive to smaller changes in cognitive function than the MMSE and has less tendency towards a ceiling effect; therefore, it is commonly recommended in mild cognitive impairment and mild dementia. Scores range from 0-30 with higher scores being better. (1)

1.2.3 Quality of Life in Alzheimer's Disease (QoL-AD)

The secondary outcome of quality of life will be measured by the QoL-AD, which was recommended as the main quality of life measure in Alzheimer's disease (1). The QOL-AD scale has a total score range of 13–52 with higher scores indicating better QOL. Subscale items are rated 1–4 (poor, fair, good, or excellent) on a variety of domains such as physical health, mood, memory, relationships, ability to complete tasks, and economy (7). Both patient and caregiver QoL-AD will be included as secondary outcomes.

1.3 Exposure

The primary exposure is the Fried Frailty Phenotype, rated as frailty yes/no. The Fried Frailty Phenotype identifies frailty as a distinct physiological entity with ageing, which is overlapping but separate from disability and morbidity. Frailty by this definition is defined as weigh loss (>4.5 kg in one year), exhaustion, slow gait speed (below 20th percentile for sex and height), weakness (handgrip strength below 20th percentile for sex and BMI), and low physical activity (by the Physical Activity Scale for the Elderly). The presence of three or more out of five criteria is defined as frailty while one to two criteria is prefrailty (8). The measurement of physical activity is the only modification form the original operationalization of the phenotype (9), and is based on the cutoff previously published in the Osteoporotic Fractures in Men study(10)

Table 1: operationalization and cutoffs for Fried Frailty Criteria.

Domain	**Operationalization**
Weight loss	Yes/ no answer to the question: In the last year, have you lost more than 4.5 kg (10 pounds) unintentionally?
Exhaustion	**Based on answer to the two originally chosen** CES–D Depression Scale questions, "I felt that everything I did was an effort" or "I could not get going" in the last week. Reporting to either question that the participant had felt in that way most or a moderate amount of the time results in the domain being scored as present.
Physical Activity	Total Physical Activity Scale for the Elderly Score below 89.6 as previously published

	in the Osteoporotic Fractures in Men study(10)
Walking speed	**Men** height ≤173 cm and speed ≤0.6531 m/s height >173 cm and speed ≤0.762 m/s **Women** height ≤159 cm and speed ≤0.6531 m/s height >159 cm and speed ≤0.762 m/s
Grip strength	Average dominant hand grip strength over 3 trials measured by JAMAR hand held dynamometer. Cutoffs: **Men** BMI ≤24 kg and strength <29 kg BMI 24.1–28 and strength <30 kg BMI >28 and strength <32 kg **Women** BMI ≤23 and strength <17 kg BMI 23.1–26 and strength <17.3 kg BMI 26.1–29 and strength <18 kg BMI >29 and strength <21 kg

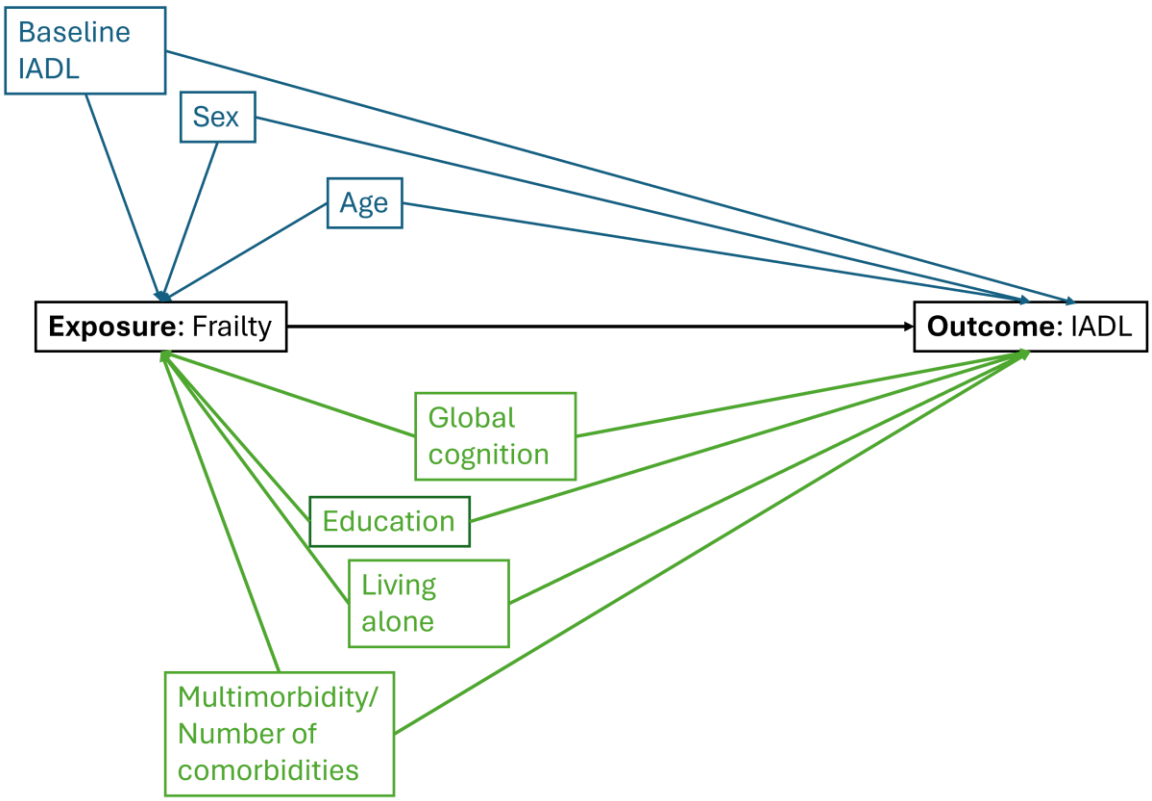
Weight loss, exhaustion and physical activity is based on information reported by participants.

1.4 Confounders

In identifying confounders, we had to take into consideration that some baseline variables may be mediators, as frailty may have been present for some time before the baseline visit. However, a core set of potential confounders were thought to be age (in years), sex, and baseline levels of the outcome.

Further, potential confounders in a more advanced model were identified to be global cognitive function, multimorbidity (as total number of comorbidities), education (in years), and living arrangement (alone or with spouse/partner). The below figure illustrates the directed acyclic graph. Factors written in green identify factors only included in the extended model, while factors in blue identify confounder included in both models.

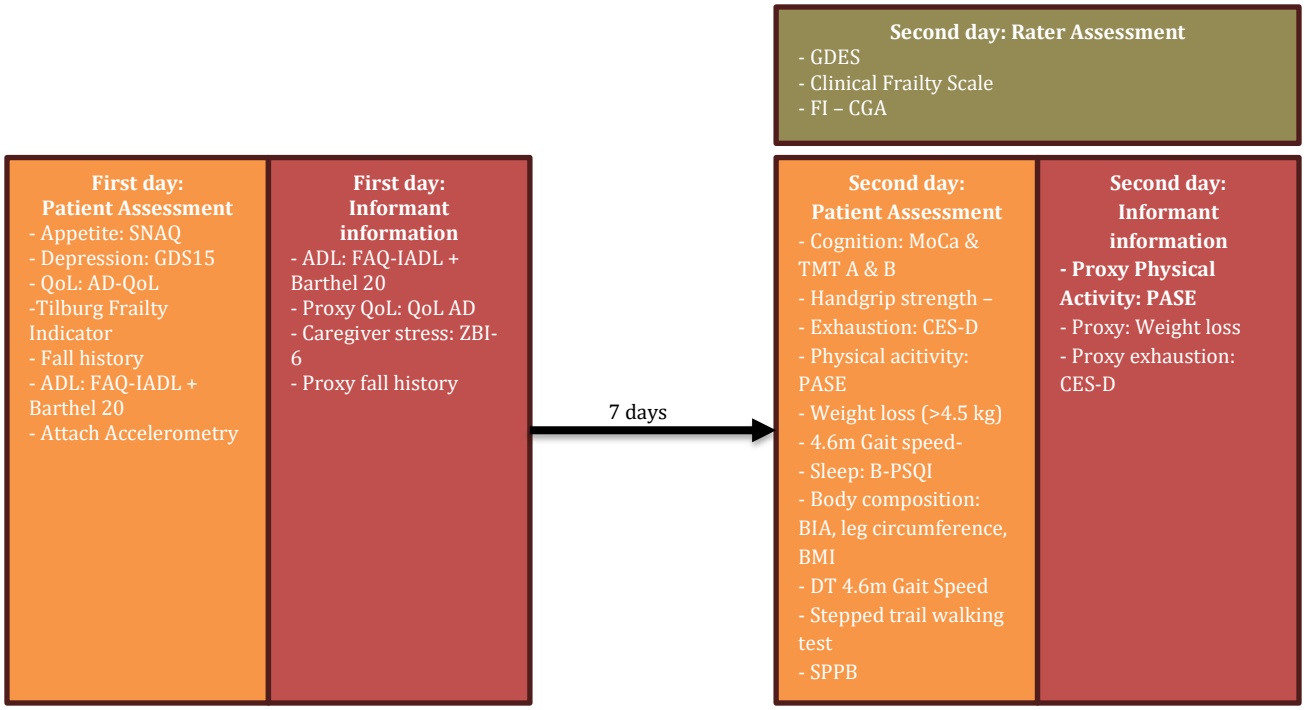
Factors which overlapped or were part of the frailty criteria were intentionally omitted from the model, as to not adjusted for frailty indirectly, and as they were identified as likely to be mediators. These were malnutrition (overlap with weight loss in frailty definition), physical activity, and depression (as the CES-D questions for exhaustion are form a depression scale, and it may be a mediator). Further tests of motor-cognitive function were not included as they were identified as potential mediators.



We have chosen the same confounder set for secondary outcomes, as ADL function, cognition and quality of life are highly correlated in Alzheimer's dementia, and therefore likely share common confounder, however the baseline score of the outcome is adjusted for each analysis (from IADL), and analyses of quality of life and cognition are not adjusted for baseline IADL function as this is then identified as they were identified as potential mediators.

1.5 Other variables

There are also several other variables collected in the assessment battery at baseline and follow-up, including factors such as motor-cognitive tests, malnutrition tests, accelerometry, and caregiver burden. These will be used for characterization at baseline, and for ancillary analyses



Abbreviations: SNAQ: simplified nutritional appetite questionnaire, GDS15: Geriatric depression stale, AD-QOL: Quality of Life Alzheimer's Disease Scale, MOBID-2: Mobilization-Observation-Intensity-Dementia pain scale 2, FAQ ADL: Functional Assessment Questionnaire Activities of Daily living, ZBI: Zarit Burden Interview, MoCa: Montreal Cognitive Assessment,: TMT: Trail Making Test, PASE: Physical Acitvity Scale in the Elderly, BIA: Bioimpedance analysis, DT: dual task, SPPB: Short Physical Performance Battery.

1.6 Sample size and power

The a-priori sample size calculation was based on a two-sample t-test comparing FAQ-ADL between frail and non-frail individuals at follow-up. Based on the previous works by Wessel et al 2022 and Andrew et al 2019, a minimal clinically significant change in FAQ-IADL after 1 year of follow-up would be 3 points (2,3). Therefore for sample size estimation we assumed a mean change difference of 3 points on informant reported FAQ-IADL between participants with and without physical frailly at one year with a standard deviation of the difference between follow-up and baseline of 4 in each group calculated from previous reports (2). Further we assumed a prevalence of frailty in dementia of 25% and a drop-out rate of 20% (11). Based on these figure we estimate that at minimum of 130 participants (33 with physical frailty, 97 without) have to be recruited to obtain 90% power at a significance level of 5%. To obtain 80% a total of 98 participants would be required (25 with frailty, 73 without).

1.7 Statistical analysis

Even though the a priori power calculation was based on a t test of the difference between groups the main statistical analyses will instead use regression to adjust for baseline levels. This is to mitigate the risk of Lord's paradox associated with using change scores and should also result in increased power in the analysis (12). The primary statistical analysis

will therefore utilize generalized linear model to estimate the difference between groups adjusted for the two confounder sets detailed previously.

Continuous confounders will be modelled linearly in the primary analysis. However, non-linear effects will be examined by introducing restricted cubic splines, and this form will be retained for the respective confounder, if inclusion of their non-linear form with a restricted cubic spline in the model changes the effect estimate of the main exposure.

Selection bias due to participation and attrition will be mitigated through inverse probability weighting. Weights for participation will be generated from logistic regression using age, sex and MMSE score of non-participants, which is the only available data on non-participants. Weights for death and other sources of attrition will be calculated in separate models. The model for death will include the variables in the full model of primary analysis. Rates for other sources of dropout will further include depressive symptoms by GDS and caregiver burden by Zarit Burden Scale.

Low levels of missing baseline data is expected, and if this is below 5% complete case analysis will be conducted. In case of more than 5% missing baseline data on variables in the main analysis, multiple imputation with chained equations (50 imputations) will be used to account for missing data, but complete case results will still be presented.

1.8 Pre-planned sensitivity analyses

Preplanned sensitivity analysis includes combining the FAQ-IADL with an inverse of the Barthel 20. This gives a combined ADL score and serves to mitigate potential ceiling effects on the FAQ-IADL, as participants with more impairment on FAQ-IADL then would be able to experience decline on the Barthel.

Further, analysis with the frailty phenotype stratified into three groups of robust, prefrail and frail will be reported. Finally, analysis without inverse probability weighting will be reported.

2 Preplanned ancillary analysis A: institutionalization and death

2.1 Objective

After the 1-year follow-up schedule for the primary outcome, participants will undergo a five-year pre planned administrative chart-based follow-up. From the chart information on movement to nursing home and death will be collected. The objective is the investigate the association between the Fried Frailty Phenotype and death/institutionalization.

2.2 Statistical analysis

The preplanned ancillary analysis will be performed using survival analysis. Analysis with death as the outcome will include an unadjusted Kaplan Meier analysis, and Cox regression adjusting for the two covariate models detailed in section 1.4. For institutionalization a unadjusted cumulative incidence function will be used, as well as cause specific Cox regression, also with the two adjustment model detailed in section 1.4.

The inverse probability weighting for selection described in section 1.7 will be used. It is not expected to have a large loss to follow-up as participants who decline from in person follow-up will still be included in administrative chart-based follow-up and therefore this weighting will only be used if more than 5% of participants decline administrative follow-up.

It is assumed that the main confounders for decline in functional ability, death, and institutionalization overlap, and therefore there is no change in confounders. This also increases comparability to the primary analysis.

Preplanned sensitivity analyses include an analysis with institutionalization and death as a composite outcome. Also, all analyses will be conducted with inverse probability weighting as sensitivity analyses.

3 Preplanned ancillary analysis B: reliability of the Fried Frailty measure and its components.

3.1 Objective

The Fried Frailty Phenomena is based on five components including three self-report items (exhaustion, weight loss, physical activity) and two measurements (grip strength, gait speed). Due to cognitive deficits and fluctuations in people with Alzheimer's dementia it is necessary to confirm reliability of the Fried Frailty Phenotype in this population, as well as explore agreement with modifications relying on informant report or less complex self-report questions. Therefore, this analysis aims to:

1) Compare test-retest reliability of frailty classification by the Fried Frailty Phenotype based on self-report and informant report information
2) Compare reliability of components of the Fried Frailty Phenotype to identify components with potential poor reliability based on self-report.
3) Explore agreements between potential modifications and the original Frailty definition components which are based on self-report.

3.2 Statistical analysis

Cohens kappa will be used to compare test retest of the self-report and informant report fried frailty criteria (based on frailty yes/no). Similarly, Cohens kappa will be used to compare reliability of components of the frailty criteria (and for subjective items this will both be for self-report and informant). These components are score 0 or 1 based on the criteria in the table in section 1.3.

Agreement will be assessed also with Cohens kappa as well as percentage agreement and prevalences meeting each criteria. Potential modifications that will be tested for the three self-report items include the informant report item for each of them.

For physical activity, which is the most extensive self-report item requiring , this also includes objective measures of activity based on accelerometry and the simpler question "How often do you engage in activities that require a low or moderate level of energy such as gardening, cleaning the car, or doing a walk?" scored a s1 if answer is one to three times a month or hardly ever or never and 0 if more than once a week or once a week.

For weight loss a modification will be in participants reported appetite as poor or very poor, similarly to how others have modified the criteria (9). Also, modifications include a score the FI-CGA rater assessed nutrition as poor and a BMI below 22 (inspired by the GLIM criteria.. Finally, for the exhaustion criteria the item from the Tilburg Frailty Criteria on physical tiredness will be tested as a modification, as well as FI-CGA rater assessed poor motivation (though a slightly different construct).

4 Preplanned ancillary analysis C: comparison of frailty measures and intrinsic capacity

4.1 Objective

Aside from the from primary objective investigating the Fried Frailty Phenotype, the project also includes a preplanned comparison of different frailty measures and the newer concept of intrinsic capacity(13–15). Data is collected on the Fried Frailty Phenotype, the Tilburg Frailty Indicator, and the Frailty Index by standardized Comprehensive Geriatric Assessment. The three frailty model are theoretically different, and it is expected that they will classify frailty differently in the included sample. However, as research on frailty in older people with Alzheimer's dementia is extremely limited, it is not possible to determine which model is optimal for clinical and research use in this setting. Therefore, it is deemed relevant to compare the different frailty models andintrinsic capacity and investigate their association to functional and cognitive ability at baseline as well as1 year decline.

Overall to achieve better understanding on which paradigm may be most useful for older adults with Alzheimer's dementia, the exploratory analysis aims to:

1) Compare agreement in frailty classification by the Fried Frailty Phenotype (Frail if score ≥3 of 5), the Tilburg Frailty indicator (Frail if score ≥5 of 15), and the Frailty Index based on standardized Comprehensive Geriatric Assessment (FI-CGA, frail if index score ≥0.25)
2) Compare baseline correlations between the "raw scores" of the Fried Frailty (0-5), Tilburg Frailty Indicator (0-15), FI-CGA (0-1), and intrinsic capacity (0-10) and also functional ability and cognition and
3) Repeat the analyses described for the Fried Frailty Phenotype in section 1 for FI-CGA, Tilburg Frailty Indicator and Intrinsic Capacity for the outcomes of informant reported IADL and cognition by MoCA score to compare predictive ability of the different constructs.

4.2 Operationalization of Intrinsic Capacity

No standard operationalization of intrinsic capacity exists yet. Instead intrinsic capacity is typically rated form 0-2 on 5 domains: cognition, locomotion, mood, hearing and vision, and vitality. All participants in the study has cognitive impairment, but instead of scoring all as having low cognitive capacity, this is graded to indicate the degree of cognitive impairment. The operationalization will be as follows:

Cognition: scored 2 if MoCA above 17, 1 if MoCA 10 to 17, and 0 if MoCA below 10.

Locomotion: scored 2 if SPPB above 9, 1 if SPPB 7-9, and 0 if SPPB < 7.

Mood: scored 2 if GDS below 6, 1 if GDS is from 6-10, and 0 if GDS > 10.

Hearing and vision: 2 if both rated at least fair. Subtract one if vision is self-rated poor or blind. Subtract one if hearing is self-rated poor or deaf, or if two or less words are heard on both ears during whisper test.

Vitality: One point if no weight loss is reported using the question from the Fried Frailty weight lool criteria. Further one point if appetite is not reported as poor or very poor.

4.3 Statistical analysis

The first objective will be analyzed by comparing percentage classification of frailty, displayed as a Venn diagram. Further pairwise agreement will be measured using Cohens kappa.

Correlations in the second aim will be calculated pairwise in a matrix using Spearmans rho.

The analysis in the third aim will follow the same approach as in section 1 including adjustment, inverse probability weighting, and sensitivity analyses.

5 References

1. Ellison TS, Cappa SF, Garrett D, Georges J, Iwatsubo T, Kramer JH, et al. Outcome measures for Alzheimer's disease: A global inter-societal Delphi consensus. Alzheimer's & Dementia [Internet]. 2023 Jun 1 [cited 2023 Sep 14];19(6):2707–29. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12945

2. Andrews JS, Desai U, Kirson NY, Zichlin ML, Ball DE, Matthews BR. Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. Alzheimer's & Dementia : Translational Research & Clinical Interventions [Internet]. 2019 Jan 1 [cited 2023 Sep 6];5:354. Available from: /pmc/articles/PMC6690415/

3. Wessels AM, Rentz DM, Case M, Lauzon S, Sims JR. Integrated Alzheimer's Disease Rating Scale: Clinically meaningful change estimates. Alzheimer's & Dementia : Translational Research & Clinical Interventions [Internet]. 2022 [cited 2023 Sep 6];8(1). Available from: /pmc/articles/PMC9169866/

4. Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol [Internet]. 1982 [cited 2023 Sep 14];37(3):323–9. Available from: https://pubmed.ncbi.nlm.nih.gov/7069156/

5. Juva K, Mäkelä M, Erkinjuntti T, Sulkava R, Ylikoski R, Valvanne J, et al. Functional assessment scales in detecting dementia. Age Ageing [Internet]. 1997 Sep [cited 2023 Sep 14];26(5):393–400. Available from: https://pubmed.ncbi.nlm.nih.gov/9351484/

6. Montero-Odasso M, Almeida QJ, Bherer L, Burhan AM, Camicioli R, Doyon J, et al. Consensus on Shared Measures of Mobility and Cognition: From the Canadian Consortium on Neurodegeneration in Aging (CCNA). J Gerontol A Biol Sci Med Sci [Internet]. 2019 May 16 [cited 2023 Sep 12];74(6):897. Available from: /pmc/articles/PMC6521916/

7. Bhattacharya S, Vogel A, Hansen MLH, Waldorff FB, Waldemar G. Generic and disease-specific measures of quality of life in patients with mild Alzheimer's disease. Dement Geriatr Cogn Disord [Internet]. 2010 Oct [cited 2025 Oct 17];30(4):327–33. Available from: https://pubmed.ncbi.nlm.nih.gov/20924186/

8. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci [Internet]. 2001 [cited 2023 Sep 14];56(3). Available from: https://pubmed.ncbi.nlm.nih.gov/11253156/

9. Theou O, Cann L, Blodgett J, Wallace LMK, Brothers TD, Rockwood K. Modifications to the frailty phenotype criteria: Systematic review of the current literature and investigation of 262 frailty phenotypes in the survey of health, ageing, and retirement in Europe. Ageing Res Rev. 2015 May 1;21:78–94.

10. Cawthon PM, Marshall LM, Michael Y, Dam TT, Ensrud KE, Barrett-Connor E, et al. Frailty in Older Men: Prevalence, Progression, and Relationship with Mortality. J Am Geriatr Soc [Internet]. 2007 Aug 1 [cited 2025 Oct 17];55(8):1216–23. Available from: /doi/pdf/10.1111/j.1532-5415.2007.01259.x

11. Koria LG, Sawan MJ, Redston MR, Gnjidic D. The Prevalence of Frailty Among Older Adults Living With Dementia: A Systematic Review. J Am Med Dir Assoc [Internet]. 2022 Nov 1 [cited 2023 Aug 22];23(11):1807–14. Available from: https://pubmed.ncbi.nlm.nih.gov/35283085/

12. Tennant PWG, Tomova GD, Murray EJ, Arnold KF, Fox MP, Gilthorpe MS. Lord's "paradox" explained: the 50-year warning on the use of "change scores" in observational data. 2023 Feb 3 [cited 2025 Oct 18]; Available from: https://arxiv.org/pdf/2302.01822

13. de Oliveira VP, Ferriolli E, Lourenço RA, González-Bautista E, de Souto Barreto P, de Mello RGB. The sensitivity and specificity of the WHO's ICOPE screening tool, and the prevalence of loss of intrinsic capacity in older adults: A scoping review. Maturitas. 2023 Nov 1;177:107818.

14. Zhang W, Wang A ping. Functional ability of older adults based on the World Health Organization framework of healthy ageing: a scoping review. Journal of Public Health (Germany) [Internet]. 2023 Nov 1 [cited 2024 Apr 8];1–19. Available from: https://link.springer.com/article/10.1007/s10389-023-02121-x

15. Koivunen K, Schaap LA, Hoogendijk EO, Schoonmade LJ, Huisman M, van Schoor NM. Exploring the conceptual framework and measurement model of intrinsic capacity defined by the World Health Organization: A scoping review. Ageing Res Rev. 2022 Sep 1;80:101685.