Robert French and Fiona Steele

Trajectories of functional disability for the elderly in Britain

Article (Accepted version) (Refereed)

Original citation: French, Robert and Steele, Fiona (2015) *Trajectories of functional disability for the elderly in Britain.* Longitudinal and Life Course Studies, 6 (3). ISSN 1757-9597

DOI: 10.14301/llcs.v6i3.317

Trajectories of frailty for the elderly in Britain

*Robert French, Centre for Multilevel Modelling, University of Bristol

Fiona Steele, Department of Statistics, London School of Economics and Political Science

R.French@bris.ac.uk

Abstract

This study uses an innovative approach to characterise trajectories of functional disability over the final stages of the life course. We use data from the British Household Panel Survey (BHPS), an annual household survey of all adults in a representative sample of British households from 1991-2008. The analysis focuses on the subsample of elderly household members who were aged from 65 to 84 in any of the 18 waves of data, with a final sample of 6,140 individuals contributing a total of 22,124 person years. As in previous research, we estimate latent growth curves, but extend the standard model to incorporate a measurement model for the latent outcome variable 'functional disability'. We identify accelerating trajectories of frailty for a representative sample of elderly individuals separately by gender. We show that socio-occupational classification is associated with the level of initial frailty and to a lesser extent the change in frailty with age. The contribution of this paper is to explore the use of a measurement model to exploit the variation between items in discriminatory power for identifying an individual's functional disability. Further we are able to test explicitly for temporal measurement invariance in frailty i.e. to what extent the items consistently measure the latent variable as people age.

Key terms. Ageing; Activities of daily living; Health trajectories; Britain; British Household Panel Survey (BHPS); Structural equation model (SEM); Growth model; Measurement model; Temporal measurement invariance

Funding: This work was supported by the UK Economic and Social Research Council (grant number RES-576-25-0032).

Introduction

The ageing population in the UK is a consequence of a reduced birth rate and delayed mortality. Delayed mortality may result in a change in the prevalence of morbidity, either increasing (Verbrugge, 1984) or decreasing (Fries, 1980), which has implications for health care costs. For a well-informed policy response to this ageing population one needs well-defined measures of health for the elderly, and to establish how these measures progress with age, and how the level and nature of change with age differs between individuals.

The ageing process is typically represented by a trajectory of declining health, defined by increasing disability (Grundy & Glaser, 2000), diminishing quality of life (Zaninotto, Falaschetti, & Sacker, 2009), self-rated health (Sacker, Worts, & McDonough, 2011), physical performance (Payette et al., 2011), or ability to carry out everyday activities (Haas, 2008). In this study we are concerned with a functional definition of health - how far health limits an individual's ability to enjoy a normal life - rather than a medical definition or diagnosis, since it allows comparability between individuals across a variety of different health conditions (Burchardt, 2000). This is typically measured using questions regarding individuals' ability to undertake everyday tasks over several domains. The core set of such questions is the activities of daily living (ADL) (Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963), the ADL term is also used generically to describe a wide variety of question sets that attempt to capture similar constructs. Extensions to ADL include the instrumental activities of daily living (iADL) (Lawton & Brody, 1969) which includes higher level tasks, and SF-36 (Ware & Sherbourne, 1992) which captures social functioning. Because ADLs are measured over different domains it is useful to combine these items into a single metric of functional disability (also referred to as frailty) for analysing changes in the health of elderly individuals.

We argue that the methods currently used to combine the ADL scores for models of change in functional disability have two important limitations. Firstly, these studies use simple aggregations of individual items, such as the sum of ADL scores, to create the single metric for analysing change, typically assigning equal or arbitrarily-chosen differential weights to each activity. This approach ignores variability between items in their relative difficulty (e.g. climbing stairs may be more difficult than walking down the street) and in their ability to discriminate between individuals with different levels of physical functioning. Secondly, previous studies have assumed that the difficulty and discriminatory power of items is the same at each age, widely referred to as *temporal / longitudinal measurement invariance* (or its complement *measurement equivalence*). Departures from this assumption due to changes in the relationship between the observed items and the underlying construct with age observed in cross-sectional studies (LaPlante, 2010) will confound attempts to identify growth patterns.

The primary aim of this paper is to describe how a type of longitudinal structural equation model (SEM) can be used to analyse functional disability trajectories. The model we propose treats physical functioning as a time-varying latent variable that is measured by a set of observed indicators (ADL items). A measurement model specifies the relationship between the latent variable and the ADL items, with parameters representing the difficulty and discriminatory power of each item. A major advantage of making explicit the relationship between the items and the latent construct (physical functioning) in this way is that it also allows exploration and testing of temporal measurement invariance. We describe how increasingly restrictive forms of measurement invariance can be tested by comparison of measurement models with different parameter constraints. We then show how, under the assumption of temporal measurement invariance, the measurement model can be combined with a growth model for latent physical functioning in which change can depend on individual characteristics.

We illustrate the application of the longitudinal SEM and testing of temporal measurement invariance in analyses of functional disability using data from the British Household Panel Survey (BHPS). We allow change in functional disability to differ by gender, age grouping and socio-economic status. Separate models are fitted for men and women because we expect functional disability trajectories for women to show worse health for biological, psychological and sociological factors (Nathanson, 1975). We also estimate separate models for two age groups, the 'early retirement' group aged 65 to 74 and the 'middle elderly' aged 75 to 84. Models for the 'oldest old' aged 85+ are not considered due to insufficient sample size in the BHPS. Socio-economic status (as measured by final occupation) is included as a covariate in the growth model. We allow trajectories to vary by socio-economic status as we expect a social gradient through accrued exposure to risk factors, both in terms of direct effects from certain types of employment, but also from the indirect risk factors and mediating factors associated with class (Nilsson, Avlund, & Lund, 2010).

Review of Approaches to Modelling Trajectories of Physical Functioning

Trajectories of functional disability can be estimated using either a multilevel model (MLM) or structural equation model (SEM). In their simplest form, these models are equivalent (Curran, 2003; Steele, 2008). Both allow for individual-specific trajectories with normally distributed latent variables representing individual departures from the intercept and slope of an overall growth curve, and both can be extended to allow for nonlinear growth. These latent variables are usually referred to as random effects in MLM and factors in SEM.

In an MLM for growth the repeated health measurements are viewed as a two-level hierarchical structure with occasions nested within individuals and age is treated as a time-varying explanatory variable (Goldstein & Woodhouse, 2001). The advantage of the MLM approach is that it is very flexible, with possible extensions to the basic growth curve including allowance for additional levels of clustering and between-individual variation in the timing of measurements at a given occasion. Individuals not present at all measurement points can be included under a 'missing at random' (MAR) assumption (Little and Rubin, 2002), however those with missing items within a wave require multiple imputation of missing values in order for that wave to be included.

In an SEM for growth, the measures at each occasion are treated as the observed indicators of the unobserved latent growth factors, i.e. latent variables for the individual-specific intercepts and slopes. The advantage of the SEM approach is the ability to include additional latent variables, for example to allow for measurement error in outcomes or covariates. It is also straightforward using SEM software to incorporate individuals with incomplete data within waves, again under an MAR assumption.

Several studies have fitted a latent growth curve model to trajectories of functional disability. Li (2005) estimates a two-level random effects model of 'ADL disability' using the Michigan Medicaid Waiver Program of individuals aged 65+ measured every few months from 1999 to 2003. They find evidence of an accelerating trajectory of ADL disability for the whole sample. Park et al. (2008) use a similar model of 'functional status' for the University of Alabama at Birmingham Ageing Study which surveyed individuals aged 65+ every 6 months from 1999 to 2004. They also find increasing and accelerating functional disability. Mendes de Leon et al. (2002) estimate a MLM of 'ADL disability' using the Women's Health and Ageing Study which followed women aged 65+ for 24 consecutive weekly assessments in 1992. They find a linear increase in functional disability over this much shorter time frame. Haas (2008) estimates a latent growth curve of 'functional limitations' using the Health and Retirement Study which followed individuals aged 61-71 at baseline annually from 1992-2002, and find an increasing and accelerating trajectory for functional limitations.

These studies use unconditional models as a baseline and to identify functional form, and conditional models to quantify how these trajectories differ by individual characteristics. Stuck et al. (1999) review the individual risk factors for ADL decline. In this paper we focus on two of the most common factors: gender and SES. Although females live longer than males, women generally have higher reported illness (Nathanson, 1975). There are many reasons for these differences, for example biological factors such as genes and hormones make males more susceptible to diseases that result in death, e.g. from heart disease, while women are more likely to suffer from conditions which impact on reported health but not death, e.g. arthritis (Case & Paxson, 2005). Moreover, there are gender differences in acquired risks, for example men are more likely to smoke and drink while females are more likely to be overweight and face stress (Verbrugge, 1989). The SES gradient in health arises from direct risk factors associated with occupation, e.g. physical hazards and psychosocial stressors at work, but also from risk behaviours associated with class, e.g. smoking and heavy alcoholic drinking (Feng et al., 2013). Over the life course we expect the SES effect to increase as exposure lengthens (Sacker, Clarke, Wiggins, & Bartley, 2005), but once an individual retires the SES effects accrued during the working life may diminish as exposure to certain risk factors associated with work cease (House, Kessler, & Herzog, 1990).

Measurement of physical functioning in longitudinal studies

All of the methods used for studying longitudinal change in functional disability discussed above are based on a single health outcome variable derived from answers to a series of questions. The simplest approach to creating a single measure from multiple measures is to sum ADL scores on each question. For example, Li (2005) uses questions on eight activities, with responses coded between zero (no limitation) and four (maximum limitation). These scores are summed across the eight items to generate the functional disability outcome measure. Using the total ADL score is problematic since each component is given equal weight, thus ignoring variation in the discriminatory power of the different items. Other studies use ad hoc methods to assign different weights to the items. For example, Holstein et al. (2006) measure levels of difficulty for 12 ADL items, and use these to create four categories of functional disability: (i) individuals who can manage all items without difficulty, (ii) individuals who can manage every activity but some with difficulty, (iii) those who need help in at least one category, and (iv) those who need help with two or more activities. Such an approach compounds the problem of equal weighting of different items by then using arbitrary thresholds for categorisation; it also ignores much of the information contained in the responses. We propose to use a measurement model to generate a single metric for functional limitations, which has the advantage that it allows each of the activities to have its own relationship with the latent outcome variable, rather than imposing equal or arbitrary weights.

Methods

Data and measures

Data for the study are from the British Household Panel Survey (BHPS), an annual household survey of all adults in a representative sample of British households from 1991-2008 (ISER, 2010). Elderly household members (aged 65 or over) were asked additional questions on their ability to carry out activities of daily living and these formed the sample for analysis. Our analysis is based on 2,788 males and 3,352 females, contributing a total 22,124 person years.

Our observed indicators of physical functioning are the six ADL items: 'cut toenails', 'get up and down stairs or steps', 'walk down the road', 'get around the house', 'bath, shower or wash all over', and 'get in and out of bed'. The score for each ADL item was constructed from responses to two questions: whether the individual is able to carry out an ADL (Q1 coded unaided, aided, or not at all) and, for those who answered "unaided", the level of difficulty in performing the ADL (Q2 coded very easy, fairly easy, fairly difficult, or very difficult). Thus for each ADL we can construct a six-point score, ranging from zero for those with the least functional disability who could carry out the ADL unaided (Q1) and very easily

(Q2) to five for those with the greatest functional disability who could not carry out the ADL at all (Q1 only).

Although measurements of ADL were available for all individuals aged 65 and above, we consider only individuals at ages 65 to 84, split into two-subsamples aged 65-74 and 75-84. For example, with 18 years of data, an individual aged 65 at wave 1 may have ADL measurements until they were aged 83 (at wave 18), and will contribute 10 waves of data in the 65-74 age group and 8 waves of data in the 75-84 age group. We discard data for ages 85 and above. We fit separate models for each age group because measurement invariance is more likely to be satisfied within each group than for a single model fitted across the whole 20-year age range. For the same reason we also estimate separate models for male and females, thus avoiding the assumption that the measurement model has the same form for both genders.

We allow the level and rate of change of the trajectories to differ by SES. The measure of SES used is the National Statistics Socio-Economic Classification (NS-SEC) (Office for National Statistics, 2010) which categorises each individual's most recent occupation into eight classes. Occupation is measured for each wave, and we use the modal category across all waves. The 'never worked and long term unemployed' category is excluded from the main analysis because this group are likely to have pre-existing health issues and thus have trajectories that are different from the majority of the population. Analyses including the 'never worked' group as an additional SES category are presented in Appendix A. We find little impact on estimates of differences in functional disability trajectories among the seven employed categories after inclusion of never-worked individuals. Long-term unemployed individuals were excluded because of the small sample size in this category (fewer than 3 individuals in all age by gender subsamples) and for a more homogenous grouping.

Individuals not present for all waves of the survey were still included in the analysis, as were cases with missing data on some of the ADL measures for a particular year. Individuals with missing data are retained using full information maximum likelihood under a missing at random (MAR) assumption (Little & Rubin, 2002), as is standard in multivariate methods such as SEM. In the present study, MAR allows the probability of nonresponse on any ADL item in a given year to depend on observed responses on other items in the same year, observed responses in previous years, age, gender and SES. If missingness depends directly on the values that would have been observed, after accounting for observed responses, nonresponse is said to be 'not missing at random' (NMAR) or non-ignorable. For example, the MAR assumption may be questionable if participants leave the study due to sudden ill health that is not adequately captured by their physical functioning trajectory up to the time of drop out. (B. Muthén, Asparouhov, Hunter, & Leuchter, 2011) describe how approaches allowing for NMAR, such as selection and pattern mixture models, can be incorporated in the general SEM framework, while emphasising the importance of sensitivity analyses using a range of models due to untestable assumptions.

Longitudinal structural equation model of physical functioning

In this paper we use a type of SEM known as a multiple indicator growth model (Chan, 1998; Hancock, Kuo, & Lawrence, 2001; Wu, Liu, Gadermann, & Zumbo, 2010). The model consists of two simultaneously estimated components: a measurement model relating responses on the six observed ADL items to a latent variable representing physical functioning, and a growth model for change in the latent variable with age. As described above, separate SEMs were fitted for each of the four age-by-gender subsamples. All models were fitted using the Mplus software (L. K. Muthén & Muthén, 1998-2012).

Measurement model and tests for temporal measurement invariance

Let y_{rti} denote the response on item r at age t for individual i. A general longitudinal measurement model can be written

$$y_{rti} = \alpha_{rt} + \lambda_{rt} f_{ti} + \epsilon_{rti} \quad (1)$$

where f_{ti} is the latent 'frailty' at age t for individual i, α_{rt} are intercepts, λ_{rt} are coefficients or factor loadings, and ϵ_{rti} are residuals. The age-specific factors f_{ti} and residuals ϵ_{rti} are each assumed to follow multivariate normal distributions. We allow for autocorrelation in both frailty and individual items across ages. We assume that the covariance between items at a given age t is explained by the common factor f_{ti}, so that $\text{cov}(\epsilon_{rti}, \epsilon_{sti}) = 0$ for $r \neq s$. To fix the location and scale of f_{ti} we impose the identification constraints $\alpha_{1t} = 0$ and $\lambda_{1t} = 1$.

The model given by (1) treats responses on the items as continuous, and may produce underestimates of standard errors and inflated chi-square model fit statistics when the normal distribution assumption is violated, for example when y_{rti} are ordinal (Lei & Wu, 2007). There are two main methods for accommodating ordinal items: (i) the underlying variable model (UVM) based on polychoric correlations and estimated by weighted least squares (WLS), and (ii) the item response theory (IRT) model where equation (1) is replaced by a proportional odds model and estimated by maximum likelihood (Jöreskog & Moustaki, 2001). Although the IRT model is generally preferred, it is computationally heavy and goodness-of-fit measures are limited. The UVM approach is therefore more widely used, but a potential disadvantage is that WLS handles missing data by pairwise deletion which requires stronger assumptions than full information maximum likelihood (Asparouhov & Muthén, 2010). Another issue with methods for ordinal data is that assessment of measurement invariance is less straightforward than for the continuous case (Millsap & Yun-Tein, 2004). For these reasons, we adopt the common practice of treating the ordinal ADL items as continuous. This approach is supported by evidence from an extensive simulation study comparing the UVM approach with traditional factor analysis which suggests that biases in the parameter estimates tend to be small when items have six or more categories (Rhemtulla, Brosseau-Liard, & Savalei, 2012).

The model of equation (1), which we refer to as Model 1, allows for changes in the underlying structure of frailty with age through the inclusion of age–specific intercepts and loadings. However, under this model individual trajectories in f_{ti} are difficult to interpret because changes in the true level of physical functioning with age are confounded with changes in its measurement. Before estimating growth trajectories for f_{ti} we therefore test for temporal measurement invariance by considering two increasingly restricted forms of equation (1). In Model 2, factor loadings for the same item are constrained to be equal across ages ($\lambda_{rt} = \lambda_r$). This model assumes *metric invariance*, i.e. the strength of the relationship between each item and the underlying latent variable is constant over time. Metric invariance is required for the latent variable to have the same meaning at different ages, and is tested by comparing Model 2 with the unconstrained Model 1. We then consider Model 3 with the additional restriction that the intercepts for the same item are fixed across ages ($\alpha_{rt} = \alpha_r$), i.e. the intercepts in the relationships between the items and the latent variable are constant over time. Comparison of Model 3 and Model 2 tests for *scalar invariance*, a necessary condition for comparing the mean of the latent variable across ages. The combination of metric and scalar invariance in Model 3 is sometimes referred to as *strong invariance*, which is widely considered as an essential prerequisite for examining temporal change in f_{ti}.

Assessing goodness of fit

We test the overall fit of Model 1 to determine whether the baseline for subsequent change in model fit is a good representation of the data. We use chi-squared (χ^2) tests, comparing with the saturated model which perfectly reproduces the sample means, variances and covariances of each observed ADL item at each age. Although the χ^2 test is widely used, there are several limitations relevant to our study: (i) the χ^2 test statistic is dependent on sample size and sensitive to the size of the correlations between the observed items, with large samples and correlations leading to higher values of χ^2, (ii) in a multi-group model (or repeated observation of the same group over time) the χ^2 test is sensitive to even minor deviations between the groups' sample covariance matrices, and (iii) the test is based on the assumption that the observed variables have a multivariate normal distribution, with departures from normality leading to higher values of χ^2 (Kline, 2005; Vandenberg & Lance, 2000). These problems with the χ^2 test have led to the development of numerous fit indices which are usually considered alongside the χ^2 test, many of which are based on the χ^2 with adjustments for sample size and model complexity.

For each of these alternative fit indices, Vandenberg and Lance (2000) specify the traditional values required to infer good model fit alongside the more stringent thresholds proposed by Hu and Bentler (1999). We consider both of these thresholds in our analysis. The first of the alternative tests is the Tucker-Lewis index (TLI) (Tucker & Lewis, 1973) which is less susceptible to sample size and favours parsimonious models. Values of the TLI range between 0 and 1 with higher values indicating better fit, and a traditional threshold of 0.9 or above and a more stringent threshold of 0.95 or above for a good model fit. The second alternative test of fit is the root mean square error of approximation (RMSEA) (Steiger, 1990) which does not require a null model and also adjusts for model complexity. The RMSEA also ranges from 0 to 1, but with values close to zero indicating a better fit. The traditional threshold value for an acceptable model fit is 0.08 or less, with a more stringent threshold of 0.06 or less. The third alternative test is the standardized root mean square residual (SRMR) (Bentler, 1995) which is sensitive to model specifications among the factor covariances. The SRMR again ranges from 0 to 1, with lower values indicating better model fit, the traditional threshold for good model fit is 0.10 or less, and a more stringent threshold of 0.08 or less.

Model comparison

In addition to the χ^2 test and alternative tests for assessing *absolute* model fit described above, Vandenberg and Lance (2000) suggest two ways for evaluating *relative* model fit. In our case, of particular interest are the changes in model fit arising from adding the temporal measurement invariance constraints of Models 2 and 3. Metric invariance is tested by comparing Models 1 and 2, and scalar invariance by comparing Models 3 and 2. The first test we consider is based on the change in the chi-squared ($\Delta\chi^2$), where a non-significant difference between models indicates that the additional temporal measurement invariance constraint does not lead to a deterioration in model fit. However, as discussed above, there are limitations with chi-squared tests for measurement invariance. A second approach is to examine the change in the comparative fit index ($\Delta\mathrm{CFI}$). Cheung and Rensvold (2002) provide guidelines on model fit suggesting that a ΔCFI value closer to zero than -0.01 indicates that the more restrictive model is an adequate fit (i.e. the invariance hypothesis should not be rejected), a $\Delta\mathrm{CFI}$ of between -0.01 and -0.02 indicates researchers should be suspicious about the invariance assumption, and $\Delta\mathrm{CFI}$ of less than -0.02 suggests that the invariance constraint should be rejected.

Latent growth model with SES effects

The measurement model shown in equation (1) specifies the relationship between an individual's latent frailty f_{ti} at age t and their responses on the observed ADL items. Age is centred at the baseline age for the age sample (65 or 75). The second part of the SEM (commonly referred to as the 'structural' model) is a growth model for change in this latent variable with age. We consider a nonlinear growth model in which f_{ti} changes as a quadratic function of age and additionally depends on dummy variables for SES denoted by x_{mi} $(m = 2, 3, \ldots, 7)$, taking the first category (routine occupations) as the reference. The growth model can be expressed as

$$f_{ti} = \beta_{0i} + \beta_{1i}t + \beta_{2i}t^2 + \sum_{m=2}^{7} \gamma_{0m}x_{mi} + \sum_{m=2}^{7} \gamma_{1m}x_{mi}\, t + \sum_{m=2}^{7} \gamma_{2m}x_{mi}t^2 + e_{ti} \quad (2)$$

The intercept and coefficients of the quadratic function in age, $\beta_{ki} = \beta_k + u_{ki}$ ($k = 0,1,2$), are composed of a fixed part β_k common to all individuals and an individual-specific random effect u_{ki}, where the random effects (u_{0i}, u_{1i}, u_{2i}) are assumed to follow a trivariate normal distribution. The e_{ti} are independent normally distributed time-varying residuals. The main effects of SES, the coefficients γ_{0m} of x_{mi}, allow baseline frailty (at $t = 0$, age 65 or 75) to depend on SES, while the coefficients of the interactions between SES and t and t^2 (γ_{1m} and γ_{2m}) allow the rate of change in functioning with age to vary by SES.

This SEM (Model 4) which combines the measurement model of equation (1) and growth model of equation (2) is the main model of interest. We also estimate a second SEM (Model 5) which constrains the factor loadings to be equal for all items. This is akin to modelling the growth of a frailty measure which is simply the sum of the scores on each of the items. Thus contrasting Model 4 with Model 5 allows us to see the effect of failure to allow for differences in the discriminatory power of the ADL items when modelling frailty trajectories.

Results

Measurement models and evidence for temporal measurement invariance

To test for temporal measurement invariance in the underlying latent 'frailty' variable, we estimate three versions of the measurement model with increasingly restrictive constraints. Model 1 is a simple measurement model with no measurement invariance constraints i.e. factor loadings and item intercepts are allowed to vary with age. Absolute model fit statistics for Model 1 are presented in the first panel of Table 1. For all age by gender groups the χ^2 test indicates significant differences between Model 1 and the saturated model (with parameters for the means, variances and covariances for the 6 ADL items measured at 10 time points) at the 1% level, which implies that the model is not a good fit to the data. The TLI gave weak evidence of good model fit with values below the more stringent threshold for both gender subsamples, with females just above the less stringent while males were below even this threshold. The RMSEA provided the strongest evidence of good model fit, with values for the well below the more stringent threshold for both samples. The SRMSR also provides evidence of good model fit, with values below the more stringent threshold except for older male group, though this is still well within the less stringent threshold. Although there is a suggestion that Model 1 could be improved, of most interest are comparisons with Models 2 and 3 for testing measurement invariance.

Insert 'Table 1: Tests for temporal measurement invariance' here

Model 2 is a restricted version of Model 1 with the factor loadings for each item constrained to be equal for all ages. Figure 1 shows that the estimated factor loadings of Model 1 are broadly similar across ages though with a slight upwards trend (which is consistent with all activities becoming more difficult as individuals get older), so it seems reasonable that constraining these to be equal over time may be a sensible assumption. We formally test whether this assumption holds by assessing the change in model fit between Models 1 and 2, in other words whether the differences in the factor loadings of the measurement model by age shown in Figure 1 are sufficiently large to lead to a significant change in model fit. The tests of change in model fit between Model 1 and Model 2 are shown in the second panel of Table 1, following overall goodness-of-fit statistics for each model. The $\Delta\chi^2$ between Models 1 and 2 suggests that imposing time invariant factor loadings leads to a significantly worse model fit (p<0.001). However we see only a small $\Delta\mathrm{CFI}$ indicating an insignificant change in model fit, far below the threshold for metric invariance. Based on $\Delta\mathrm{CFI}$ and the similarity in estimates of factor loadings by age (Figure 1), we therefore conclude that there is some evidence of metric invariance.

Insert 'Figure 1: Factor loadings when allowed to vary by age (Model 1)' here

Model 3 is a more restricted version of Model 2 in which the intercepts for each item are constrained to be the same for all ages (these were allowed to vary by age in Models 1 and 2). Figure 2 shows the estimates of the item intercepts by age for Model 2. These show a downward trend in the item intercepts over time. We formally test for scalar invariance - that is whether constraining the item intercepts to be equal across ages is a reasonable assumption - by examining the change in model fit statistics between Model 2 and Model 3 (see the third panel of Table 1). As seen for the contrast between Models 1 and 2 the $\Delta\chi^2$

indicates a significantly worse model fit (p<0.001), but a very small ΔCFI supports the assumption of scalar invariance.

Insert 'Figure 2: Item intercepts, when allowed to vary by age (Model 2)' here

Structural equation model

Model 4 is our main model of interest, a full SEM combining a measurement model of the same specification as Model 3 (assuming metric and scalar invariance) with a growth model. We interpret the measurement model parameters below and in the following section we contrast the growth model parameters of Model 4 with an alternative SEM which approximates a growth curve model fitted to an unweighted sum of scores on the ADL items (Model 5). Model 5 is similar to the growth models fitted in most previous research, but with frailty as a latent variable rather than a sum score.

An important consideration when evaluating differences in parameter estimates across subsamples (male vs. female and age 65-74 vs. 75-84) or model specifications (Model 4 vs. Model 5) is that these may be due in part to differences in the variance of the physical functioning factor. Suppose, for example, that we wish to compare the factor loading for a particular ADL item for two groups. Even if the underlying relationship between the ADL response and the factor is the same for each group, the estimated factor loading will be of smaller magnitude in the group with the largest factor variance. Standardised factor loadings and growth model coefficients can be computed to take account of such scaling effects (see Appendix B for details). We present unstandardised factor loadings and item intercepts for the measurement model component of the SEM in Table 2, and unstandardised model estimates for all growth model parameters of Models 4 and 5 in Tables 3a and 3b. Between-gender and between-age comparisons can be made as the factor variance is fairly similar across genders and age groups. However, because the factor variance changes according to whether or not the factor loadings in the measurement model are permitted to vary across ADL items, we present a separate set of standardised estimates for the overall SES effects of Models 4 and 5 in Table 4.

The measurement model component of the SEM

The factor loadings λ_r and item intercepts α_r of the measurement part of the growth model are shown in Table 2. The factor loadings are interpreted as the expected change in the observed ADL item for a one unit change in the factor, and represent the discriminatory power of the items in terms of the functional disability latent variable. The factor loading for the first item: 'cutting toenails' is constrained to 1 to fix the scale of the latent variable. 'Walking down the road' and 'managing steps' have the largest factor loadings indicating they are best at discriminating between individuals with different levels of functioning. 'Getting around the house' and 'getting in and out of bed' have the lowest factor loadings, i.e. these are the least discriminatory items for changes in functional disability.

Insert 'Table 2: Factor loadings and item intercepts for measurement part of SEM (Model 4)' here

The item intercepts represent the difficulty of the ADLs. We constrain the first item 'cutting toenails' to zero, and this is the least difficult activity because the estimated intercepts for the other items are all negative. For all age by gender subsamples 'walking down the road' is the most difficult activity, followed by 'bathing, showering and washing'. The other categories ('managing stairs or steps', 'getting in and out of bed' and 'getting around the house') have roughly equal values within each age by gender subsample. The intercepts are larger in magnitude for females (within age groups), which is consistent with the literature on poorer female health.

The growth model component of the SEM

The parameter estimates for the growth model component of the full SEM (Model 4) are shown in the left half of Table 3a (for males) and Table 3b (for females). The coefficients of the SES dummy variables are interpreted as contrasts with the reference group 'routine occupations' at the baseline age in the sample. Functional disability at baseline (β_0) is greater for females. The intercept variances, $\mathrm{var}(u_{0i})$, are interpreted as the between-individual variances in the level of physical functioning at $t = 0$ (age 65 or 75) for each gender. We see a slightly larger baseline variance for females.

Insert 'Tables 3a and 3b: Growth model parameters' here

Predicted trajectories for each gender and age group are presented in Figure 3 with separate curves for each SES group. These trajectories are calculated using the SES coefficients for frailty for someone at the mean of the distribution, in other words the individual random effects are set at their means of zero. For all SES groups we estimate a positive linear growth (β_1) in frailty, and the quadratic growth factor mean (β_2) shows a slight acceleration in growth for the older male group. The random effect variance associated with the linear age effect, $\mathrm{var}(u_{1i})$, is similar for men and women, though slightly smaller for males. There is a negative covariance between the individual intercepts and slopes suggesting that higher frailty at baseline is associated with slower increase in frailty. Note that the variance of the random effect for t^2 (u_{2i}) and its covariances with the other random effects were found to be negligible, and were therefore omitted from the structural model.

Insert 'Figure 3: Frailty trajectories by socio-economic status' here

SES is allowed to affect both the intercept and slope of frailty. For each age-by-gender group we find small but significant effects of SES on the intercept (γ_{0m}) compared with the reference category 'routine occupations', though the differences among lower status occupations ('small employers and own account workers', 'lower supervisory and technical occupations', and 'semi-routine occupations') are not consistently statistically significant. In terms of the social gradient in the change in frailty (γ_{1m} and γ_{2m}) males show a slight widening of the social gradient in functional disability with age, while females show a slight convergence with age (though from a more divergent baseline), though these relationships are statistically significant only for males and for the less routine occupations. In Appendix 2 Figure A3 we show the trajectories for the never worked group. For all age by gender groups we see high initial frailty which increases with age, for females this increase is in line with other SES categories, whereas for males we see a much higher growth rate in the post retirement age group, but a much lower growth rate for the 75-84 age group. Table 4 shows standardised estimates of SES effects at selected ages, calculated by dividing the unstandardised estimates of Tables 3a and 3b by an estimate of the standard deviation of the latent frailty variable at that age (see Appendix B for details). These standardised coefficients can be interpreted as differences in frailty between 'routine occupations' and other SES categories measured in standard deviation units. For example, at age 65 men in higher managerial occupations are predicted to have a frailty score 0.57 standard deviations lower than for men in routine occupations.

In the right-hand side of Table 4 we show estimates from a comparison model (Model 5) which proxies a growth model fitted to an unweighted sum of ADL scores. Table 5 shows standardised SES effects for Models 4 and 5 for the age by gender subsamples, calculated for selected ages 3 years apart. We would not expect the SES effects to be dramatically different given the factor loadings from the measurement part of Model 4 shown in Table 2 are relatively close to one another. This comparison shows the SES effects would be slightly underestimated (lower predicted frailty) when no measurement models is used for the post-retirement age group. The differences are greater for the 75-84 age groups, with the measurement model giving higher predictions of frailty for males, but much lower predictions for females.

Discussion

The general health of the elderly population is typically measured using questions relating to functional ability across a range of dimensions. When using these measures to model trajectories of functional disability as people age, researchers typically use simple methods to combine these indicators, such as the total score. We argue that these approaches are limited since they do not capture the difference in discriminatory power of these different items. We propose supplementing the growth model of functional disability with a measurement model to better capture the underlying latent variable functional disability that we wish to use as the outcome in the growth model.

Another advantage of specifying a measurement model is that it makes explicit and allows testing of the assumption of temporal measurement invariance. We estimated a sequence of three increasingly restricted models in order to test for measurement invariance for the gender-by-age subsamples. Vandenberg and Lance (2000) argue that assessing model fit using only a χ^2 test is limited because it is sensitive to sample size and differences in the covariance structure, and suggest using a suite of fit indices including TLI, RMSEA and SRMR to evaluate the degree of temporal measurement invariance. By recognising the strengths and weaknesses of each of these indices we are able to build a more robust assessment of the temporal measurement invariance assumption.

We estimate separate SEMs for the growth in latent frailty for subgroups defined by age and gender. Overall we see increasing functional disability, with accelerating growth for females but not for males. For both genders we find evidence of a social gradient in the baseline levels of frailty between the most routine occupational class (the reference category) and the least routine social classes. The social gradient in the rate of change of frailty was less statistically significant, though we do see large differences for example our model predicts that the functional disability of a male from the lowest SES group at aged 65 is equivalent to that of an individual from the highest SES group who was 10 years older.

Classification of females into SES groups using only an individual's occupation is an imperfect proxy for SES since it ignores non employment based determinants of SES, which for many women, especially for our older population, may be more related to their husband's occupation (Arber, 1997). In our analysis we focus on individual measures of SES and health, however this could be extended to include both the household cluster and household measures of SES.

Future research which estimates trajectories of functional disability for the elderly could benefit from adopting our approach of using an SEM to incorporate a measurement model which treats frailty as a latent variable. The SEM framework is extremely general and there are a number of generalisations of the model described here which allow a richer set of substantive questions to be addressed. One such extension is the growth mixture model (GMM) in which individuals with similar physical functioning trajectories are grouped into latent classes and the probability of class membership depends on individual characteristics such as SES (e.g. Jung & Wickrama, 2008; B. Muthén & Asparouhov, 2008). GMM is usually applied to longitudinal data on a single response. Latent transition analysis is a generalisation of GMM suitable for multivariate longitudinal data, such as our multiple ADL items, where a latent class is defined for each time point and individuals may move between classes over time (Collins & Lanza, 2013).

Another direction for future work is the use of richer datasets which would allow a wider set of items to measure functional disability, and a wider set of controls. For example the English Longitudinal Study of Ageing includes measures of iADL and mobility to supplement the ADL, and has better measures of SES to improve identification of the social gradient in frailty trajectories (which could be used in a second measurement model for a latent SES measure). To date, these datasets have only been used to model functional disability cross-sectionally (Gjonça, Tabassum, & Breeze, 2009). Time-varying measures of social status would allow us to explore the relationship between change in SES and change in functional disability, and determine whether the changes in SES effects with age are real or simply a function of increasing time since the measure was taken. We know that there may be reverse causality in this relationship as health status could also impact on social status (Steele, French, & Bartley, 2013). Longitudinal data on both health and SES would allow us to identify the direction of these effects. Residential status is another time-varying characteristic of policy relevance (because of the cost of residential care) which may be included as a determinant of frailty trajectories. Such a model could be extended to identify the effect of residential status on individuals where care needs (including moves into residential care) are not met (Scott, Evandrou, Falkingham, & Rake, 2001). Finally, studies that incorporate this approach over shorter term periods would be able to capture aspects of recovery as well as the longer term increase in functional disability found in this study. Importantly, a shorter time span would also make it easier to satisfy the temporal measurement invariance assumption.

References

Arber, S. (1997). Comparing Inequalities in Women's and Men's Health: Britain in the 1990s. *Social Science & Medicine, 44*(6), 773-787.

Asparouhov, T., & Muthén, B. (2010). Weighted Least Squares Estimation with Missing Data. *MplusTechnical Appendix*.

Bentler, P. M. (1995). *Eqs Structural Equations Program Manual* (Vol. 6).

Burchardt, T. (2000). The Dynamics of Being Disabled. *Journal of Social Policy, 29*(04), 645-668. doi: 10.1017/S0047279400006097

Case, A., & Paxson, C. (2005). Sex Differences in Morbidity and Mortality. *Demography, 42*(2), 189-214. doi: 10.2307/4147343

Chan, D. (1998). The Conceptualization and Analysis of Change over Time: An Integrative Approach Incorporating Longitudinal Mean and Covariance Structures Analysis (Lmacs) and Multiple Indicator Latent Growth Modeling (Mlgm). *Organizational Research Methods, 1*(4), 421-483. doi: 10.1177/109442819814004

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling, 9*(2), 233-255.

Collins, L. M., & Lanza, S. T. (2013). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences* (Vol. 718): John Wiley & Sons.

Curran, P. J. (2003). Have Multilevel Models Been Structural Equation Models All Along? *Multivariate Behavioral Research, 38*(4), 529-569. doi: 10.1207/s15327906mbr3804_5

Feng, Q., Zhen, Z., Gu, D., Wu, B., Duncan, P. W., & Purser, J. L. (2013). Trends in Adl and Iadl Disability in Community-Dwelling Older Adults in Shanghai, China, 1998–2008. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68*(3), 476-485. doi: 10.1093/geronb/gbt012

Fries, J. F. (1980). Aging, Natural Death, and the Compression of Morbidity. *N Engl J Med., 303*(3), 130-135. doi: 10.1056/NEJM198007173030304

Gjonça, E., Tabassum, F., & Breeze, E. (2009). Socioeconomic Differences in Physical Disability at Older Age. *Journal of Epidemiology and Community Health (1979-), 63*(11), 928-935. doi: 10.2307/20721090

Goldstein, H., & Woodhouse, G. (2001). Modelling Repeated Measures. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel Modelling of Health Statisitcs* (pp. 13-26). Chichester: Wiley.

Grundy, E., & Glaser, K. (2000). Socio-Demographic Differences in the Onset and Progression of Disability in Early Old Age: A Longitudinal Study. *Age and Ageing, 29*(2), 149-157. doi: 10.1093/ageing/29.2.149

Haas, S. (2008). Trajectories of Functional Health: The 'Long Arm' of Childhood Health and Socioeconomic Factors. *Social Science & Medicine, 66*(4), 849-861. doi: 10.1016/j.socscimed.2007.11.004

Hancock, G. R., Kuo, W.-L., & Lawrence, F. R. (2001). An Illustration of Second-Order Latent Growth Models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 470-489. doi: 10.1207/S15328007SEM0803_7

Holstein, B. E., Avlund, K., Due, P., Martinussen, T., & Keiding, N. (2006). The Measurement of Change in Functional Ability: Dealing with Attrition and the Floor/Ceiling Effect. *Archives of Gerontology and Geriatrics, 43*(3), 337-350. doi: 10.1016/j.archger.2005.12.004

House, J. S., Kessler, R. C., & Herzog, A. R. (1990). Age, Socioeconomic Status, and Health. *The Milbank Quarterly, 68*(3), 383-411. doi: 10.2307/3350111

Hu, L. t., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi: 10.1080/10705519909540118

Institute for Social and Economic Research. (2010). *British Household Panel Survey: Waves 1-18, 1991-2009 [Computer File] 7th Edition*.

Jöreskog, K. G., & Moustaki, I. (2001). Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivariate Behavioral Research, 36*(3), 347-387.

Jung, T., & Wickrama, K. (2008). An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass, 2*(1), 302-317.

Katz, S., Ford, A., Moskowitz, R., Jackson, B., & Jaffe, M. (1963). Studies of Illness in the Aged. The Index of Adl: A Standardized Measure of Biological and Psychosocial Function. *JAMA, 185*, 914 - 919. doi: 10.1001/jama.1963.03060120024016

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modelling* (2 ed.). New York: The Guilford Press.

LaPlante, M. P. (2010). The Classic Measure of Disability in Activities of Daily Living Is Biased by Age but an Expanded Iadl/Adl Measure Is Not. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 65B*(6), 720-732. doi: 10.1093/geronb/gbp129

Lawton, M. P., & Brody, E. M. (1969). Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. *Gerontologist, 9*, 179-186.

Lei, P. W., & Wu, Q. (2007). Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Educational measurement: Issues and practice, 26*(3), 33-43.

Li, L. W. (2005). Trajectories of Adl Disability among Community-Dwelling Frail Older Persons. *Research on Aging, 27*(1), 56-79. doi: 10.1177/0164027504271348

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*: Wiley.

Mendes de Leon, C. F., Guralnik, J. M., & Bandeen-Roche, K. (2002). Short-Term Change in Physical Function and Disability: The Women's Health and Aging Study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57*(6), S355-S365. doi: 10.1093/geronb/57.6.S355

Millsap, R. E., & Yun-Tein, J. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research, 39*(3), 479-515.

Muthén, B., & Asparouhov, T. (2008). Growth Mixture Modeling: Analysis with Non-Gaussian Random Effects. *Longitudinal data analysis*, 143-165.

Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth Modeling with Nonignorable Dropout: Alternative Analyses of the Star* D Antidepressant Trial. *Psychological Methods, 16*(1), 17.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide. Seventh Edition.* Los Angeles, CA: Muthén & Muthén.

Nathanson, C. A. (1975). Illness and the Feminine Role: A Theoretical Review. *Social Science & Medicine (1967), 9*(2), 57-62. doi: 10.1016/0037-7856(75)90094-3

Nilsson, C. J., Avlund, K., & Lund, R. (2010). Social Inequality in Onset of Mobility Disability among Older Danes: The Mediation Effect of Social Relations. *Journal of Aging and Health, 22*(4), 522-541. doi: 10.1177/0898264309359684

Office for National Statistics. (2010). *The National Statistics Socio-Economic Classification: (Rebased on the Soc2010) User Manual* (Vol. 3). Southampton: Palgrave Macmillan.

Park, N. S., Klemmack, D. L., Roff, L. L., Parker, M. W., Koenig, H. G., Sawyer, P., & Allman, R. M. (2008). Religiousness and Longitudinal Trajectories in Elders' Functional Status. *Research on Aging, 30*(3), 279-298. doi: 10.1177/0164027507313001

Payette, H., Gueye, N. D. R., Gaudreau, P., Morais, J. A., Shatenstein, B., & Gray-Donald, K. (2011). Trajectories of Physical Function Decline and Psychological Functioning: The Québec Longitudinal Study on Nutrition and Successful Aging (Nuage). *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 66B*(suppl 1), i82-i90. doi: 10.1093/geronb/gbq085

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical Sem Estimation Methods under Suboptimal Conditions. *Psychological Methods, 17*(3), 354.

Sacker, A., Clarke, P., Wiggins, R. D., & Bartley, M. (2005). Social Dynamics of Health Inequalities: A Growth Curve Analysis of Aging and Self Assessed Health in the British Household Panel Survey 1991–2001. *Journal of Epidemiology and Community Health, 59*(6), 495-501. doi: 10.1136/jech.2004.026278

Sacker, A., Worts, D., & McDonough, P. (2011). Social Influences on Trajectories of Self-Rated Health: Evidence from Britain, Germany, Denmark and the USA. *Journal of Epidemiology and Community Health, 65*(2), 130-136. doi: 10.1136/jech.2009.091199

Scott, A., Evandrou, M., Falkingham, J., & Rake, K. (2001). Moves into Residential Care Amongst Older People in Britain. *Paper Presented to the 2001 British Household Panel Survey Research Conference; 5-7 July, 2001, Colchester, UK, 2012*(November). https://www.iser.essex.ac.uk/files/conferences/bhps/2001/docs/pdf/papers/scott.pdf

Steele, F. (2008). Multilevel Models for Longitudinal Data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 171*(1), 5-19. doi: 10.1111/j.1467-985X.2007.00509.x

Steele, F., French, R., & Bartley, M. (2013). Adjusting for Selection Bias in Longitudinal Analyses Using Simultaneous Equations Modeling: The Relationship between Employment Transitions and Mental Health. *Epidemiology, 24*(5), 703-711 710.1097/EDE.1090b1013e31829d32479.

Steiger, J. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research, 25*(2), 173-180. doi: 10.1207/s15327906mbr2502_4

Stuck, A. E., Walthert, J. M., Nikolaus, T., Büla, C. J., Hohmann, C., & Beck, J. C. (1999). Risk Factors for Functional Status Decline in Community-Living Elderly People: A Systematic Literature Review. *Social Science & Medicine, 48*(4), 445-469. doi: 10.1016/S0277-9536(98)00370-0

Tucker, L. R., & Lewis, C. (1973). A Reliability Coefficient for Maximum Likelihood Factor Analysis. *Psychometrika, 38*(1), 1-10. doi: 10.1007/bf02291170

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods, 3*(1), 4-70. doi: 10.1177/109442810031002

Verbrugge, L. M. (1984). Longer Life but Worsening Health? Trends in Health and Mortality of Middle-Aged and Older Persons. *The Milbank Memorial Fund Quarterly. Health and Society, 62*(3), 475-519. doi: 10.2307/3349861

Verbrugge, L. M. (1989). The Twain Meet: Empirical Explanations of Sex Differences in Health and Mortality. *Journal of Health and Social Behavior, 30*(3), 282-304. doi: 10.2307/2136961

Ware, J. E. J., & Sherbourne, C. D. (1992). The Mos 36-Item Short-Form Health Survey (Sf-36). I. Conceptual Framework and Item Selection. *Med Care, 30*(6), 473-483. doi: 10.1097/00005650-199206000-00002

Wu, A., Liu, Y., Gadermann, A., & Zumbo, B. (2010). Multiple-Indicator Multilevel Growth Model: A Solution to Multiple Methodological Challenges in Longitudinal Studies. *Social Indicators Research, 97*(2), 123. doi: 10.1007/s11205-009-9496-8

Zaninotto, P., Falaschetti, E., & Sacker, A. (2009). Age Trajectories of Quality of Life among Older Adults: Results from the English Longitudinal Study of Ageing. *Quality of Life Research, 18*(10), 1301-1309. doi: 10.1007/s11136-009-9543-6

Tables & Figures

Table 1: Tests for temporal measurement invariance

	Males		Females	
	Aged 65-74	Aged 75-84	Aged 65-74	Aged 75-84
Absolute fit of Model 1:				
Chi-square test statistic, 1395 df (χ^2)	4830	4141	4316	3511
TLI	0.892	0.849	0.910	0.907
RMSEA	0.038	0.043	0.033	0.033
SRMSR	0.078	0.084	0.070	0.070
Absolute fit of Model 2:				
Chi-square test statistic, 1440 df (χ^2)	4945	4238	4448	3602
TLI	0.894	0.851	0.910	0.908
RMSEA	0.038	0.042	0.033	0.033
SRMSR	0.079	0.083	0.072	0.070
Absolute fit of Model 3:				
Chi-square test statistic, 1485 df (χ^2)	5107	4330	4618	3757
TLI	0.893	0.853	0.909	0.906
RMSEA	0.038	0.042	0.033	0.033
SRMSR	0.078	0.080	0.069	0.067
Change in model fit between Model 1 and Model 2:				
Chi-square test statistic, 45 df ($\Delta\chi^2$)	116	97	132	91
Change in CFI ($\Delta\mathrm{CFI}$)	-0.002	-0.002	-0.002	-0.002
Change in model fit between Model 2 and Model 3:				
Chi-square test statistic, 45 df ($\Delta\chi^2$)	161	93	169	155
Change in CFI ($\Delta\mathrm{CFI}$)	-0.002	-0.002	-0.003	-0.004
n	1712	1076	1958	1394

Table 2: Factor loadings λ_r and item intercepts α_r for the measurement part of the SEM (Model 4). Standard errors are given in brackets.

	Males				Females			
	Age 65-74		Age 75-84		Age 65-74		Age 75-84	
	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)
Factor loadings (λ_r):								
Cut toenails	1		1		1		1	
Get up and down stairs or steps	1.034	(0.023)	1.168	(0.039)	1.026	(0.024)	1.272	(0.043)
Walk down the road	1.013	(0.023)	1.240	(0.042)	1.058	(0.024)	1.407	(0.047)
Get around the house	0.716	(0.016)	0.798	(0.027)	0.664	(0.016)	0.884	(0.03)
Bath, shower or wash all over	0.871	(0.019)	1.011	(0.034)	0.840	(0.02)	1.107	(0.038)
Get in and out of bed	0.755	(0.017)	0.798	(0.027)	0.703	(0.017)	0.842	(0.029)
Item intercepts (α_r):								
Cut toenails	0		0		0		0	
Get up and down stairs or steps	-0.425	(0.039)	-1.192	(0.098)	-0.612	(0.049)	-1.759	(0.128)
Walk down the road	-0.519	(0.039)	-1.459	(0.103)	-0.842	(0.05)	-2.234	(0.142)
Get around the house	-0.442	(0.027)	-1.051	(0.066)	-0.597	(0.031)	-1.585	(0.089)
Bath, shower or wash all over	-0.489	(0.033)	-1.234	(0.084)	-0.693	(0.04)	-1.800	(0.113)
Get in and out of bed	-0.444	(0.028)	-1.021	(0.067)	-0.595	(0.033)	-1.444	(0.086)

Table 3a: Male subsample growth model parameters and model fit statistics for Models 4 and 5, SEMs with unequal and equal factor loadings across ADL items.

	Model 4: Growth model from SEM with unequal factor loadings for ADL items				Model 5: Growth model from SEM with equal factor loadings for ADL items			
	Age 65-74		Age 75-84		Age 65-74		Age 75-84	
Parameter estimates								
Intercept growth factor mean (β_0)	1.098***	(0.067)	1.909***	(0.086)	1.102***	(0.058)	1.928***	(0.081)
Slope growth factor mean (β_1)	0.060***	(0.023)	0.074**	(0.033)	0.048**	(0.019)	0.061**	(0.031)
Quadratic growth factor mean (β_2)	-0.001	(0.002)	0.003	(0.004)	-0.001	(0.002)	0.003	(0.004)
Effects of NS-SEC on intercept (γ_{0m}):								
Routine occupations (reference)								
Semi-routine occupations	0.028	(0.096)	-0.060	(0.116)	0.027	(0.079)	-0.055	(0.105)
Lower supervisory and technical occupations	-0.047	(0.097)	0.018	(0.105)	-0.046	(0.079)	0.032	(0.095)
Small employers and own account workers	-0.041	(0.091)	0.045	(0.118)	-0.031	(0.074)	0.037	(0.107)
Intermediate occupations	-0.386***	(0.136)	-0.192	(0.141)	-0.305***	(0.111)	-0.158	(0.128)
Lower managerial and professional occupations	-0.185**	(0.088)	-0.156	(0.106)	-0.133*	(0.071)	-0.137	(0.097)
Higher managerial and professional occupations	-0.247**	(0.103)	-0.233*	(0.126)	-0.188**	(0.084)	-0.206*	(0.114)
Effects of NS-SEC on coefficient of t (γ_{1m}):								
Routine occupations (reference)								
Semi-routine occupations	-0.030	(0.035)	0.024	(0.050)	-0.027	(0.029)	0.023	(0.048)
Lower supervisory and technical occupations	0.009	(0.036)	-0.082*	(0.044)	0.012	(0.030)	-0.077*	(0.041)
Small employers and own account workers	-0.060*	(0.034)	-0.016	(0.050)	-0.052*	(0.029)	-0.014	(0.047)
Intermediate occupations	-0.034	(0.048)	-0.048	(0.056)	-0.024	(0.041)	-0.045	(0.053)
Lower managerial and professional occupations	-0.063*	(0.032)	-0.019	(0.043)	-0.051*	(0.027)	-0.011	(0.041)
Higher managerial and professional occupations	-0.107***	(0.038)	-0.026	(0.050)	-0.087***	(0.032)	-0.021	(0.048)
Effects of NS-SEC on coefficient of t^2 (γ_{2m}):								
Routine occupations (reference)								
Semi-routine occupations	0.001	(0.004)	-0.006	(0.006)	0.002	(0.003)	-0.006	(0.006)
Lower supervisory and technical occupations	-0.002	(0.004)	0.006	(0.005)	-0.002	(0.003)	0.006	(0.005)
Small employers and own account workers	0.005	(0.004)	0.001	(0.006)	0.005	(0.003)	0.001	(0.006)
Intermediate occupations	0.005	(0.005)	0.009	(0.007)	0.004	(0.004)	0.009	(0.006)
Lower managerial and professional occupations	0.005	(0.003)	0.001	(0.005)	0.004	(0.003)	0.001	(0.005)
Higher managerial and professional occupations	0.009**	(0.004)	0.000	(0.006)	0.007**	(0.004)	0.000	(0.005)
Intercept growth factor variance, $\mathrm{var}(u_{0i})$	0.001	(0.004)	-0.006	(0.006)	0.002	(0.003)	-0.006	(0.006)
Slope growth factor variance, $\mathrm{var}(u_{1i})$	-0.002	(0.004)	0.006	(0.005)	-0.002	(0.003)	0.006	(0.005)
Covariance between factor mean and slope, $\mathrm{cov}(u_{0i}, u_{1i})$	0.005	(0.004)	0.001	(0.006)	0.005	(0.003)	0.001	(0.006)
Residual variance for the factor, $\mathrm{var}(e_i)$	0.005	(0.005)	0.009	(0.007)	0.004	(0.004)	0.009	(0.006)
Model fit								
Chi-square test statistic	5531	1885 df	4999	1885 df	6489	1890 df	5717	1890 df
TLI	0.898		0.849		0.871		0.815	
RMSEA	0.034		0.039		0.038		0.043	
SRMSR	0.075		0.082		0.125		0.129	
n	1712		1076		1712		1076	

*** p <0.01, ** p<0.05, * p<0.10 . Standard errors in parentheses.

Table 3b: Female subsample growth model parameters and model fit statistics for Models 4 and 5, SEMs with unequal and equal factor loadings across ADL items.

	Model 4: Growth model from SEM with unequal factor loadings for ADL items				Model 5: Growth model from SEM with equal factor loadings for ADL items			
	Age 65-74		Age 75-84		Age 65-74		Age 75-84	
Parameter estimates								
Intercept growth factor mean (β_0)	1.443***	(0.065)	2.626***	(0.073)	1.478***	(0.055)	2.645***	(0.073)
Slope growth factor mean (β_1)	0.053**	(0.021)	0.043*	(0.022)	0.034*	(0.018)	0.038*	(0.023)
Quadratic growth factor mean (β_2)	0.002	(0.002)	0.004*	(0.002)	0.002	(0.002)	0.004	(0.003)
Effects of NS-SEC on intercept (γ_{0m}):								
Routine occupations (reference)								
Semi-routine occupations	-0.023	(0.082)	-0.084	(0.088)	-0.017	(0.065)	-0.081	(0.088)
Lower supervisory and technical occupations	0.205*	(0.120)	-0.059	(0.122)	0.157*	(0.095)	-0.092	(0.121)
Small employers and own account workers	-0.122	(0.125)	-0.199	(0.131)	-0.088	(0.099)	-0.222*	(0.130)
Intermediate occupations	-0.168*	(0.086)	-0.329***	(0.092)	-0.121*	(0.068)	-0.323***	(0.091)
Lower managerial and professional occupations	-0.160*	(0.084)	-0.251***	(0.098)	-0.125*	(0.066)	-0.254***	(0.096)
Higher managerial and professional occupations	-0.536**	(0.209)	-0.603**	(0.255)	-0.380**	(0.163)	-0.571**	(0.252)
Effects of NS-SEC on coefficient of t (γ_{1m}):								
Routine occupations (reference)								
Semi-routine occupations	-0.002	(0.030)	0.027	(0.031)	-0.001	(0.025)	0.027	(0.032)
Lower supervisory and technical occupations	0.003	(0.043)	0.015	(0.043)	0.002	(0.035)	0.017	(0.044)
Small employers and own account workers	-0.031	(0.046)	0.045	(0.044)	-0.023	(0.038)	0.045	(0.046)
Intermediate occupations	-0.047	(0.031)	0.029	(0.031)	-0.036	(0.026)	0.031	(0.032)
Lower managerial and professional occupations	-0.021	(0.031)	0.001	(0.034)	-0.011	(0.025)	0.011	(0.035)
Higher managerial and professional occupations	-0.031	(0.075)	0.109	(0.091)	-0.026	(0.062)	0.122	(0.094)
Effects of NS-SEC on coefficient of t^2 (γ_{2m}):								
Routine occupations (reference)								
Semi-routine occupations	-0.001	(0.003)	-0.004	(0.003)	-0.001	(0.003)	-0.004	(0.004)
Lower supervisory and technical occupations	-0.006	(0.004)	-0.001	(0.005)	-0.005	(0.004)	-0.001	(0.005)
Small employers and own account workers	0.001	(0.005)	-0.005	(0.005)	0.001	(0.004)	-0.005	(0.005)
Intermediate occupations	0.004	(0.003)	-0.004	(0.003)	0.003	(0.003)	-0.004	(0.004)
Lower managerial and professional occupations	-0.001	(0.003)	0.002	(0.004)	-0.001	(0.003)	0.000	(0.004)
Higher managerial and professional occupations	0.005	(0.008)	-0.011	(0.011)	0.004	(0.007)	-0.012	(0.011)
Intercept growth factor variance, var(u_{0i})	0.667	(0.047)	0.763	(0.064)	0.367	(0.022)	0.704	(0.041)
Slope growth factor variance, var(u_{1i})	0.009	(0.001)	0.010	(0.001)	0.005	(0.001)	0.010	(0.001)
Covariance between factor mean and slope, cov(u_{0i}, u_{1i})	-0.007	(0.005)	-0.016	(0.006)	-0.006	(0.003)	-0.020	(0.006)
Residual variance for the factor, var(e_i)	0.213	(0.010)	0.127	(0.009)	0.156	(0.004)	0.145	(0.006)
Model fit								
Chi-square test statistic	5304	1885 df	4211	1885 df	6645	1890 df	5078	1890 df
TLI	0.906		0.909		0.870		0.875	
RMSEA	0.030		0.030		0.036		0.035	
SRMSR	0.072		0.072		0.131		0.108	
N	1958		1394		1958		1394	

*** p <0.01, ** p<0.05, * p<0.10 . Standard errors in parentheses.

Table 4: Comparison of standardised SES effects for growth model component of Models 4 and 5, SEMs with unequal and equal factor loadings across ADL items.

	Model 4: Growth model from SEM with unequal factor loadings for ADL items				Model 5: Growth model from SEM with equal factor loadings for ADL items			
Age	65	68	71	74	65	68	71	74
Males:								
Routine occupations	0	0	0	0	0	0	0	0
Semi-routine occupations	0.031	-0.059	-0.122	-0.153	0.037	-0.050	-0.083	-0.064
Lower supervisory and technical occupations	-0.052	-0.043	-0.069	-0.122	-0.064	-0.039	-0.060	-0.118
Small employers and own account workers	-0.046	-0.197	-0.233	-0.167	-0.043	-0.198	-0.214	-0.111
Intermediate occupations	-0.431	-0.496	-0.432	-0.273	-0.422	-0.475	-0.400	-0.232
Lower managerial and professional occupations	-0.206	-0.369	-0.404	-0.330	-0.184	-0.348	-0.387	-0.316
Higher managerial and professional occupations	-0.276	-0.546	-0.596	-0.457	-0.260	-0.537	-0.600	-0.476
Females:								
Routine occupations	0	0	0	0	0	0	0	0
Semi-routine occupations	-0.025	-0.040	-0.067	-0.100	-0.024	-0.040	-0.074	-0.118
Lower supervisory and technical occupations	0.219	0.167	0.007	-0.209	0.217	0.162	-0.014	-0.254
Small employers and own account workers	-0.130	-0.215	-0.257	-0.263	-0.122	-0.203	-0.239	-0.236
Intermediate occupations	-0.179	-0.285	-0.289	-0.219	-0.167	-0.277	-0.288	-0.223
Lower managerial and professional occupations	-0.171	-0.242	-0.304	-0.353	-0.173	-0.229	-0.286	-0.337
Higher managerial and professional occupations	-0.571	-0.609	-0.512	-0.337	-0.525	-0.579	-0.493	-0.320

Age 75-84	Model 4: Growth model from SEM with unequal factor loadings for ADL items				Model 5: Growth model from SEM with equal factor loadings for ADL items			
Age	75	78	81	84	75	78	81	84
Males:								
Routine occupations	0	0	0	0	0	0	0	0
Semi-routine occupations	-0.067	-0.045	-0.124	-0.263	-0.068	-0.048	-0.143	-0.307
Lower supervisory and technical occupations	0.020	-0.185	-0.242	-0.186	0.040	-0.176	-0.230	-0.161
Small employers and own account workers	0.050	0.006	-0.014	-0.014	0.046	0.005	-0.012	-0.007
Intermediate occupations	-0.214	-0.271	-0.146	0.084	-0.197	-0.257	-0.112	0.152
Lower managerial and professional occupations	-0.174	-0.217	-0.219	-0.196	-0.170	-0.195	-0.180	-0.142
Higher managerial and professional occupations	-0.260	-0.331	-0.364	-0.372	-0.256	-0.326	-0.358	-0.363
Females:								
Routine occupations	0	0	0	0	0	0	0	0
Semi-routine occupations	-0.089	-0.041	-0.064	-0.139	-0.088	-0.040	-0.064	-0.142
Lower supervisory and technical occupations	-0.063	-0.024	-0.005	-0.004	-0.100	-0.055	-0.026	-0.018
Small employers and own account workers	-0.211	-0.116	-0.106	-0.167	-0.241	-0.146	-0.134	-0.195
Intermediate occupations	-0.349	-0.296	-0.291	-0.330	-0.351	-0.294	-0.285	-0.323
Lower managerial and professional occupations	-0.266	-0.245	-0.168	-0.067	-0.276	-0.244	-0.191	-0.136
Higher managerial and professional occupations	-0.639	-0.399	-0.335	-0.432	-0.620	-0.346	-0.275	-0.390

Figure 1: Trajectories of factor loadings (λ_r), when allowed to vary by age (Model 1)

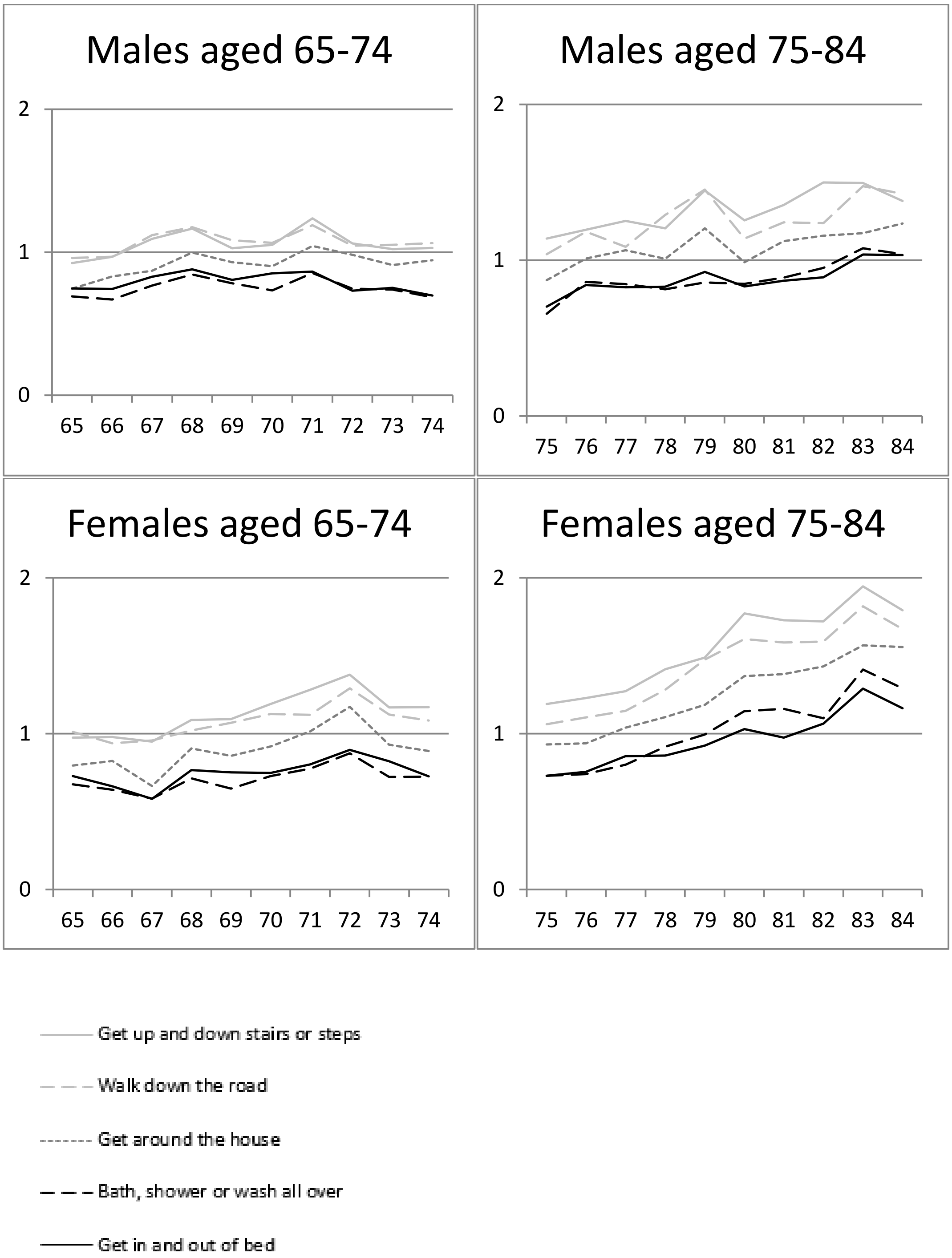

Figure 2: Item intercepts (α_r), when allowed to vary by age (Model 2)

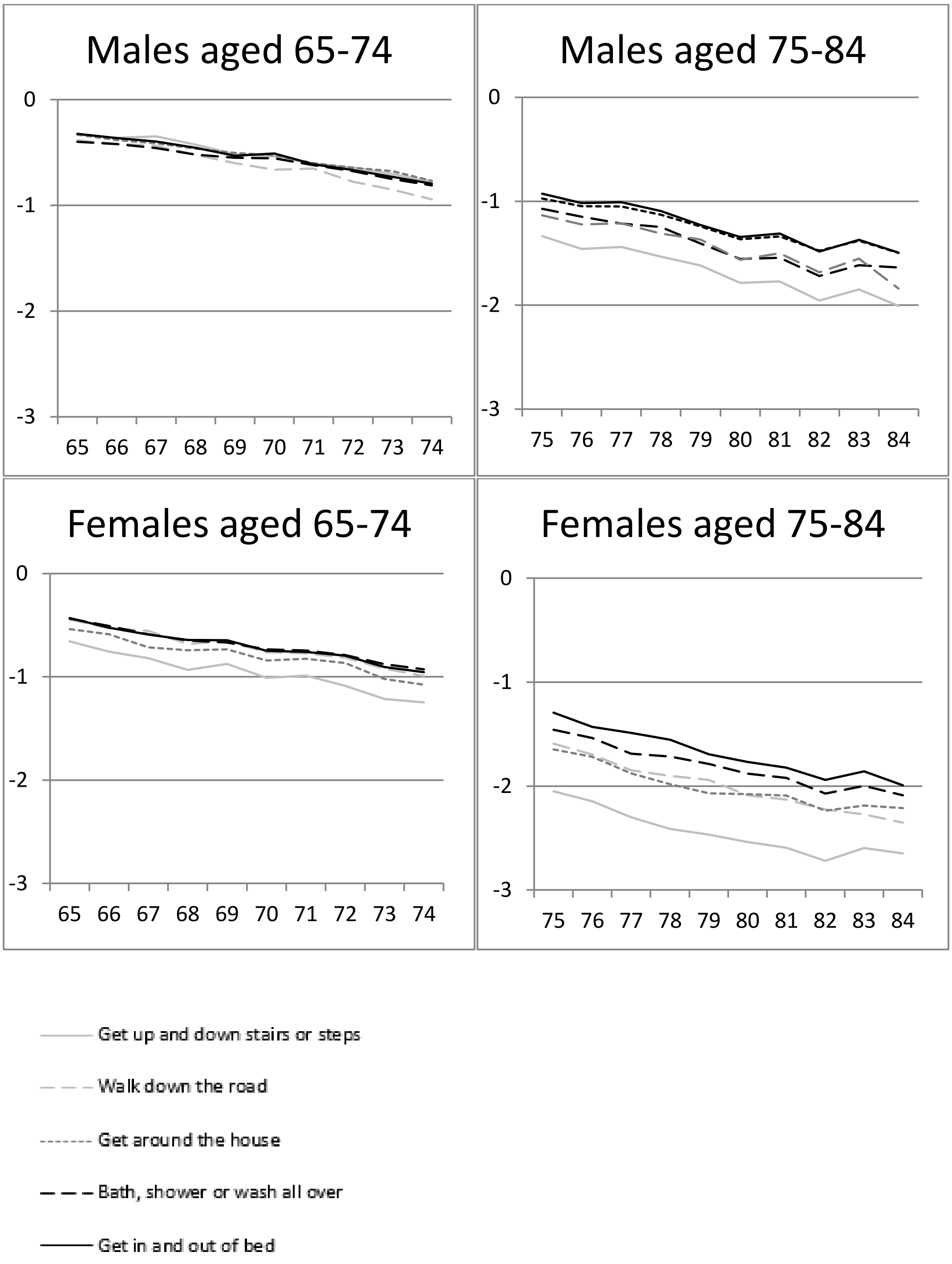

Figure 3: Frailty trajectories by socio-economic status (Model 4)