



## Research Paper

## Levels of disability in the older population of England: Comparing binary and ordinal classifications



Benedetta Pongiglione, Ph.D. Student <sup>a,\*</sup>, George B. Ploubidis, Ph.D. <sup>b</sup>,  
Bianca L. De Stavola, Ph.D. <sup>a</sup>

<sup>a</sup> Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK

<sup>b</sup> Centre for Longitudinal Studies, Department of Social Science, UCL Institute of Education, University College London, 55-59 Gordon Square, London, WC1H 0NU, UK

## ARTICLE INFO

## Article history:

Received 13 October 2016

Received in revised form

19 January 2017

Accepted 21 January 2017

## Keywords:

Disability

International classification of functioning

disability and health

England

Older population

Latent classes

Latent growth model

## ABSTRACT

**Background:** Recent studies suggest the importance of distinguishing severity levels of disability. Nevertheless, there is not yet a consensus with regards to an optimal classification.

**Objective:** Our study seeks to advance the existing binary definitions towards categorical/ordinal manifestations of disability.

**Methods:** We define disability according to the WHO's International Classification of Functioning, Disability and Health (ICF) using data collected at the baseline wave of the English Longitudinal Study of Aging, a longitudinal study of the non-institutionalized population, living in England. First, we identify cut-off points in the continuous disability score derived from ICF to distinguish disabled from non-disabled participants. Then, we fit latent class models to the same data to find the optimal number of disability classes according to: (i) model fit indicators; (ii) estimated probabilities of each disability item; (iii) association of the predicted disability classes with observed health and mortality.

**Results:** According to the binary classification criteria, about 32% of both men and women are classified disabled. No optimal number of classes emerged from the latent class models according to model fit indicators. However, the other two criteria suggest that the best-fitting model of disability severity has four classes.

**Conclusions:** Our findings contribute to the debate on the usefulness and relevance of adopting a finer categorization of disability, by showing that binary indicators of disability averaged the burden of disability and masked the very strong effect experienced by individuals having severe disability, and were not informative for low levels of disability.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

A large body of research on the conceptualization and measurement of disability has been published, accompanied by reviews of alternative disability models (with corresponding measurement methods<sup>1</sup>) and studies of the methodology for the measurement of disability. The challenges faced in this field derive from the fact that disability is often measured by self-reported responses to survey questions,<sup>2,3</sup> with problems including the wording of the questions, the time periods for which disability is reported, and the difficulties in administering surveys (see for example Freedman,<sup>4</sup> Jette<sup>5</sup>). On the other hand, the challenge of coding survey data in studies of disability, and of choosing between binary or more refined

classifications of disability has received remarkably less attention. At the same time, the relevance of identifying meaningful classification of disability is becoming clearer, while there is no general consensus in the literature with regards to the optimal number of disability grades/levels. For example, a recent study set in the UK<sup>6</sup> and studying trends in disability-free life expectancy has shown that the increase in number of years with any disability was higher for periods with mild disability than for those with moderate or severe disability. This suggests how important it is to distinguish different levels of severity of disability when assessing disability and mortality trends. Another example is the recent study of Wolf and colleagues (2015)<sup>7</sup> that identified three distinctive trajectories of disability, which differed with respect to their pace of decline. The authors acknowledged as a limitation of their work that they relied on a binary definition of disability and that a more finely graded measure might have led to different trajectories.

\* Corresponding author. Tel.: +44 7949131608.

E-mail addresses: [benedetta.pongiglione@lshtm.ac.uk](mailto:benedetta.pongiglione@lshtm.ac.uk) (B. Pongiglione), [g.ploubidis@ucl.ac.uk](mailto:g.ploubidis@ucl.ac.uk) (G.B. Ploubidis), [Bianca.DeStavola@lshtm.ac.uk](mailto:Bianca.DeStavola@lshtm.ac.uk) (B.L. De Stavola).

An attempt to categorize disability by severity level was made by Manton and Gu (2005).<sup>8</sup> They identified, among the older US population, six groups with distinctive aspects of disability ranging from active to frail and a seventh group comprising of nursing home residents. By looking at disability trends over seventeen years, the authors claimed that simply considering the 'per annum decline in chronic disability in the US elderly population (...) masks variation in per annum changes in the prevalence of disability and institutionalized residents' (Manton and Gu (2005), p.32<sup>8</sup>). Changes were found to be different across disability categories and by identifying where they occurred, the authors calculated changes in Medicare costs and related savings in population shifts from a severely disabled category to the non-disabled group.

The lack of equivalent evidence for the UK and, more generally, the scarcity of studies on more refined disability classifications, along with the potential policy relevance of addressing this gap, have motivated our work. Although the correct balance between the need for identifying a more refined disability grouping and the risk of over-classification is not easily achievable, this paper attempts to explore the advantages and disadvantages of binary versus multi-categorical classifications, by examining their discriminant power in terms of different health outcomes.

This study seeks to (1) produce binary and multi-categorical classifications of disability; (2) identify an optimal number of categories of disability using alternative criteria; (3) examine whether a multi-categorical classification may be advantageous compared to a binary.

To do so, we first measure disability as a continuous score and ask "How severe does continuous disability need to be for a person to be classified as disabled?". The rationale for finding a threshold for a binary classification, comes from the need for summarising the information, for example to provide population estimates of disability prevalence.

Splitting the population into two categories, however, may be too simplistic. Therefore, the second question we ask is "Would a finer classification of disability better capture clustering of disability items and of heterogeneity in later disease and mortality rates?". To answer this question, we identify boundaries among categories of disability severity in terms of their association with health function and mortality, adopting a similar approach as the one adopted by Serlin and colleagues<sup>9</sup> who delineated different levels of cancer pain severity. The rationale behind assessing the association between disability and selected health outcomes was derived from three hypotheses about why people with disabilities may have poorer mental and physical health than their non-disabled peers. First, the experience of living with a disability could lead to mental health problems and worse physical conditions; second, people with mental and physical health problems could be more likely to subsequently become disabled; and third, other factors, such as socioeconomic circumstances, might independently increase the risk of disability and mental ill health. Given that we measure disability at baseline and health outcomes over the course of the following ten years, we mainly explore the first hypothesis, supported by evidence on the association between disability and mental health,<sup>10</sup> as well as studies investigating the association between physical disability (or physical activity)<sup>11,12</sup> and muscle strength and quality and functional limitation.

To address these research questions, we rely upon the WHO's conceptualization of disability as defined in the International Classification of Functioning, Disability and Health (ICF),<sup>13</sup> and use a common set of disability items selected according to the ICF conceptual framework to derive first a continuous score of disability and relative cut-off points, and then to derive a multi-categorical disability summary performing latent class analysis (LCA).

## Material and methods

### Data source and sample

This study used data drawn from the English Longitudinal Study of Ageing (ELSA). Briefly, ELSA is a longitudinal study of the non-institutionalized population, living in England, who were aged 50 years or older at the time of interview. 11,391 core-member respondents were recruited at wave one in 2002/2003. For our analysis, we included all participants who had complete records on all disability items at wave 1, leaving us with a sample of 9715 (in the Appendix, section A, we justify this choice). We take advantage of the longitudinal nature of the study and also use outcome measures from the following five waves of the survey, i.e. up to 2010/2011. A detailed description of the ELSA cohort profile has been released recently.<sup>14</sup>

### Measures

#### Disability

Variables describing disability were chosen according to the WHO's ICF framework. In this study we rely on the screening and selection carried out in a previous work.<sup>15</sup> There, a total of fifty items was selected to model disability. Briefly, body function and structure were measured using variables such as self-rated eyesight and hearing, chronic conditions such as high blood pressure and arthritis, and questions about pain; activity limitations were measured through activity of daily living (ADLs) and mobility functions; instrumental activities of daily living (IADLs) and various limitations due to health problems, such as in working, were selected for assessing participation restrictions. Variables were all either dichotomous (i.e. yes/no answer) or ordered categorical. A detailed description of the items and selection process is available elsewhere.<sup>15</sup>

#### Health measures and death

Information on deaths that occurred from 2002 to 2011 was freely available, and for respondents who gave their consent to link their data to the National Health Service Central Register (NHSCR) time of death by year was also disclosed.

A number of health indicators were considered and selected to cover different spheres of health, including measures of mental health and anthropometric measures for physical domains. Here we present one outcome for each group, although analyses were replicated for more indicators and are available upon request. Mental health function was measured at every wave using eight items of the CES-D scale and treated as a continuous variable ranging from 0 to 8. Anthropometric measurements for physical functioning were assessed during the nurse visits, which took place every two waves, i.e. the second, fourth and sixth waves. Physical functioning was assessed through grip strength in the dominant hand and was estimated by the average of three measurements done using the Smedley dynamometer.

#### Confounders

When assessing the association of disability with mortality and selected health outcomes, we controlled for a number of confounders, all of them measured at wave 1. We only considered these early measures to avoid the issues arising from later values of these confounders lying on the causal pathway between disability and mortality: controlling for them would remove some of the association between exposure (disability) and outcome.<sup>16</sup> The confounders in the model included demographic characteristics such as ethnicity, marital and parental status and household size; socioeconomic position measured through income, wealth,

occupation and education; health-related behaviours including smoking, drinking and physical activity; and the presence of limiting long-lasting illness and socioeconomic background represented by father's occupation when respondents were fourteen.

### Analysis

The methods used are presented in separate paragraphs, according to the research question they address. All analyses were carried out separately for men and women.

#### Binary disability

In order to capture disability as a binary variable, we used the continuous score estimated in our previous study from the battery of 50 items described above using a latent variable measurement model. A full description of the model is available elsewhere.<sup>15</sup>

Here we intended to find a threshold in this continuous score that optimally discriminates disabled and non-disabled individuals. To do so, we adopted two approaches:

1. We follow the WHO's strategy<sup>13</sup> and replicated the WHO's approach looking at the average disability score observed among those reporting at least one limitation in any of the six disability questions selected by the Washington Group (WG).<sup>17</sup> The six disability domains identified as crucial by the WG include problems in seeing, hearing, walking or climbing steps, remembering or concentrating, washing all over or dressing and communicating, for example understanding or being understood.
2. We used the Receiver Operating Characteristic (ROC) methods<sup>18</sup> to assess the agreement of our score with an external gold standard. The external gold standard we chose consists of receiving health or disability benefits. The cut-off in the continuous disability score was then chosen using two alternative criteria<sup>19</sup> known as the point on the curve closest to the (0, 1) (where specificity = 1 and sensitivity = 1), and the Youden index.

The disability prevalence resulting from these approaches were then compared with national statistics on the proportion of disabled people in the UK in 2002. National data were collected in the General Lifestyle Survey.<sup>20</sup>

#### Multi-categorical disability

To produce categorical measures of disability, LCA was performed, using all 50 binary and categorical variables previously identified as indicators of disability. An individual is assigned to be member of a class according to his/her highest probability of being in that class, even though an individual may have several classes to which he/she is a partial membership.<sup>21</sup> We explore models with two through six latent classes of a latent variable (the algebraic notation is available in the Appendix, section B). The choice of the number of classes, i.e. of best-fitting model to represent categorical disability, is based on three sets of criteria. The first set consists of statistical indicators, including entropy, the Bayesian Information Criterion (BIC) and the bootstrap likelihood ratio test (BLRT).<sup>22</sup> The second set of criteria consisted of comparisons of each model's estimated probabilities of endorsing a disability item across classes and the assessment of whether they highlight informative differences. The third criterion used external validation of each model's predicted disability classes in terms of their association with mortality and health, as explained in the next paragraph.

#### Association of disability with mortality and health

We considered the association between each version of

categorical disability, i.e. with classes from two to six classes, and each of the following outcomes: mortality by 10 years since entry, and the longitudinal trajectories of grip strength and mental health.

Association between disability and mortality was assessed using discrete-time survival analysis (D TSA) through pooled logistic regression models,<sup>23</sup> with measures of effect expressed as odds ratios.

The association with health outcomes was parameterised using latent growth models (LGM).<sup>24</sup> Fig. 1 illustrates the conceptual model of the association of latent categorical disability and mental health observed at each wave (i.e. six-time points) selected here as an illustrative outcome, controlling for all confounders (age and all the demographic and socioeconomic variables described in the previous section). The algebraic notation to express the top part of Fig. 1 (i.e. latent growth model) can be found in the Appendix, section C. Equivalent models were fitted for all the other health outcomes.

Missingness in outcome observations was assumed to be missing at random (MAR) and maximum likelihood estimation was used. This means that we assume that missingness in outcome data was explained by observed outcomes at other waves and the variables included in the model. These variables however also suffer from missingness; for this reason, we only include respondents that have complete confounder and exposure data.

## Results

### Sample characteristics

Of 9715 total participants included in the analyses, 54% were women, and the overall average age was 64 years (64.4 and 64.8 for men and women respectively). Over the course of the study (from 2002 to 2011), 21% of males and 16% of female respondents died corresponding to 1775 respondents in total. Of the 50 disability items, women had higher prevalence than men in almost all variables, with the exception for example of difficulty in communicating and being engaged in social activity. Prevalence of all disability indicators by gender are available in Appendix Table A1 and details on summary statistics of health outcomes data across waves are available in Table 1.

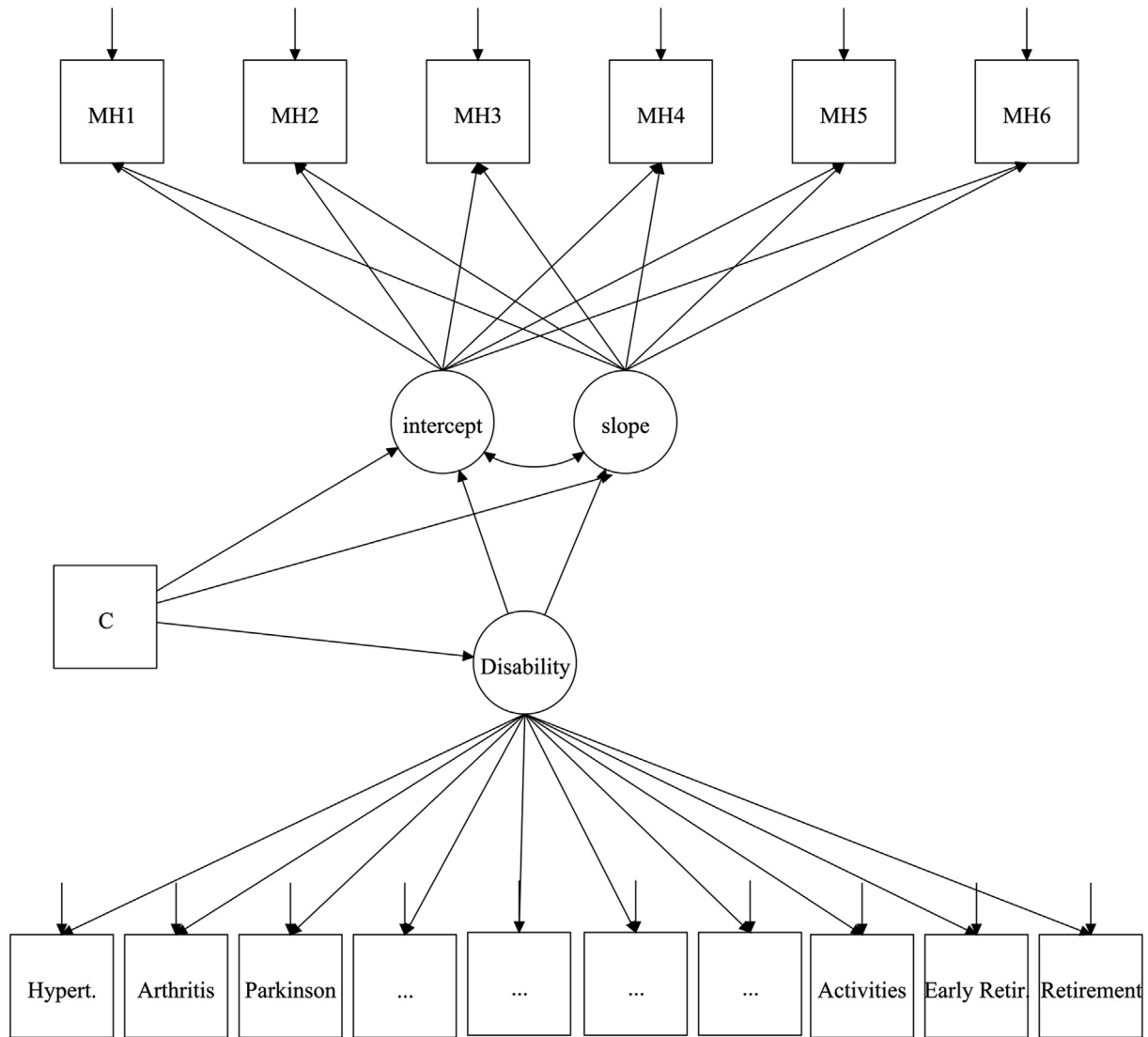
### Binary disability

According to national statistics for the 45 + UK population in 2002, 32.8% of men and 32.9% of women have disability.<sup>20</sup>

The continuous disability score developed in a previous work<sup>15</sup> ranged from -1.71 to 2.85 among men and from -1.85 to 3.42 among women (see Fig. 2). The average score among women was 0.08 and 0.09 among men.

To reproduce the WHO's strategy for finding the disability cut-off, we estimated first the proportions of respondents reporting at least one limitation in the six WG activities. 62.4% of men and 60.7% of women were found to have at least one limitation in the six WG activities. The mean disability score among the respondents belonging to this group was 0.44 and 0.47 for men and women respectively. Setting the cut-off point at these values led to 31.5% of male respondents and 31.7% of female respondents being classified as having disability.

When we used ROC analysis to set the threshold for distinguishing disabled and non-disabled individuals, we found that the cut-off for men was 0.51, according to both criteria, i.e. the Youden index and minimization of  $r$ . For women, the cut-off was 0.50 when using the minimum value of  $r$  and 0.58 based on the Youden index. According to these values, 29.2% of male respondents and either 30.3% -according to Youden index- or 27.9%



**Fig. 1.** Conceptual model including measurement model of disability and latent growth model for mental health measured on six occasions. MH = Mental Health; C = confounders (it includes all confounders listed in the text); "..." only some of the 50 disability items were reported to ease the reading of the model.

-according to minimum value of *r*-of women were considered disabled.

A graphical summary of these results is provided in Fig. 2. It shows the distribution of the continuous disability score by gender and the cut-off points obtained using the two approaches (with two results presented when using the ROC method for women).

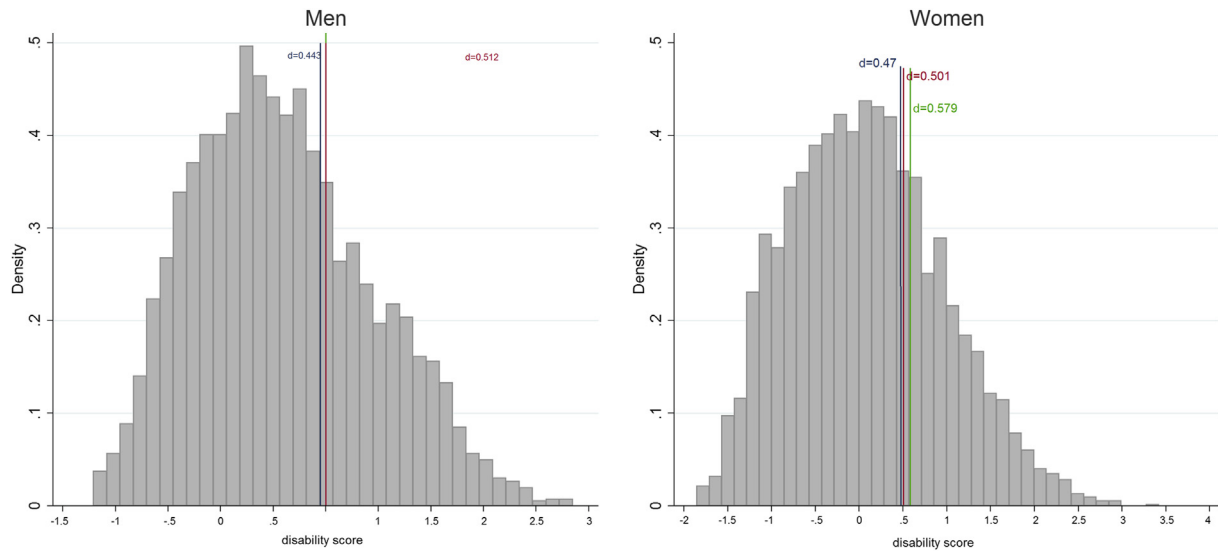
*Categorical disability*

Latent class models were fitted separately by gender; their fit indicators are presented in Appendix (Table A2). Overall, all models had entropy higher than 0.88, indication of good allocation quality for both for men and women. For both genders, the highest entropy

**Table 1**  
Sample size and mean scores of health outcomes by wave at measurement.

	Mental health score <sup>a</sup>		Grip strength <sup>b</sup>		Coagulation <sup>c</sup>	
	n	Mean score (s.d.)	n	Mean score (s.d.)	n	Mean score (s.d.)
Wave 1	9715	1.51 (1.92)				
Wave 2	7743	1.52 (1.91)	6839	29.4 (11.43)	5382	3.2 (0.72)
Wave 3	6630	1.46 (1.92)				
Wave 4	5792	1.36 (1.86)	5021	28.1 (11.29)	3513	3.4 (0.56)
Wave 5	5389	1.48 (1.91)				
Wave 6	4889	1.28 (1.8)	4238	26.9 (10.55)	2966	3 (0.54)

<sup>a</sup> Scale from 0 (no symptoms in any CES-D item).  
<sup>b</sup> Measured in kg and averaged over three measurements.  
<sup>c</sup> Measured in g/l.



**Fig. 2.** Observed distributions of continuous disability score and cut-off points identified by the two approaches, by gender. Blue line: cut-off according to WHO's approach. Red line: cut-off according to ROC curve, minimizing  $r$ . Green line: cut-off according to ROC curve, maximizing  $j$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was observed in the model with three classes, whilst the lowest BIC was in the 6-classes model. When assessing the BLRT, problems with local maxima occurred, therefore BLRT was not used as a criterion for model selection. The proportions of respondents assigned to different categories of disability by each model are reported in Table A3 in the appendix.

The estimated probabilities of a positive item response for each disability class provide a first description of each disability class. They are illustrated below in Fig. 3a and b for 3-class and 4-class models and available in Appendix for 5-class and 6-class models (Fig. A2). These probabilities were quite distinct across disability groups in the 3-class model both for men and women. We labelled the category with lowest probabilities “no disability”, the intermediate “mild disability” and the group with highest probability “severe disability”. The largest differences were found for the items concerning pains and mobility, such as walking a quarter of mile, or climbing stairs. In the 4-class model, categories were labelled as “no disability”, “mild disability”, “moderate disability” and “severe disability”. “No disability” and “mild disability” overall presented similar estimated probabilities, but some differences were still noticeable among impairment items. “Moderate disability” had higher estimated probability than “no disability” and “mild disability” for all items, and the largest differences were observed, again, for some of the impairment variables, such as having pain, and in the mobility items. Severe disability had the highest estimated probabilities for all items and the gap with moderate disability was particularly large with regards to the activity domain as well as for items describing participation. In models with five and six classes (see Appendix, Fig. A2), the groups having the lowest levels of disability (i.e. groups 1, 2 and 3) tended to have very similar estimated probabilities, suggesting that there was not much difference in the endorsement of disability items among these groups.

#### *Association of alternative specifications of disability levels with health and mortality*

Before showing whether alternative specifications of disability appear to explain some outcome variation over time and in which direction, we briefly present the results of the association of each alternative specification of disability measured at baseline with

health outcomes measured at wave 6; corresponding tables by gender are available in the appendix (Table A4). We use this analysis as a preliminary step to see whether disability affects health measured after a ten-year lag, regardless of how this relationship changed over this interval, to facilitate the interpretation of the results from the LGM presented below. In synthesis, there seemed to be a severity gradient, where those belonging to the most disabled group performed the worst compared to non-disabled, but the number of significant disability levels strongly varied depending on the outcome considered.

Table 2 shows the association of binary and multi-categorical specifications (i.e. 3-class and 4-class models) of disability with the estimated growth parameters of selected outcomes, controlling for the complete set of confounders, while model estimated means of latent growth factors and the results for 5-class and 6-class models are available in the appendix (Tables A5 and A6 respectively). Binary disability (however specified) (first three rows) was associated with a significant worse intercept (i.e. mean outcome at baseline) for disabled versus not disabled men and women. The trajectories of grip strength in women as well as the trajectories of mental health in men appeared to converge with those of non-disabled ones, as shown by the opposite sign of the disability effect on the intercepts and slopes estimated in the growth model.

To assess the existence of a severity gradient of disability, we looked at multi-categorical disability and observe whether there was evidence of heterogeneity or linearity of associations. A disability gradient was found both in the 3-class and 4-class models. In the 3-class model, at baseline (as measured by the intercept coefficient estimated in the LGM) those having mild and severe disability presented lower grip strength, higher mental health problems and higher probability of dying compared to non-disabled, with the disadvantage being larger for those suffering from severe disability. In the 4-class model, all disability categories presented worse health conditions and higher odds of dying at each time point compared to no disability. The size of the intercept coefficients of moderate disability in the 4-class model was close to that of mild disability in the 3-class model. Results for 5 and 6-class models are available in the appendix and overall indicate that the intercepts and slopes of all outcomes among individuals assigned to disability groups 2 and 3 were not significantly different from those assigned to group 1 (i.e. no disability), and for higher severity of

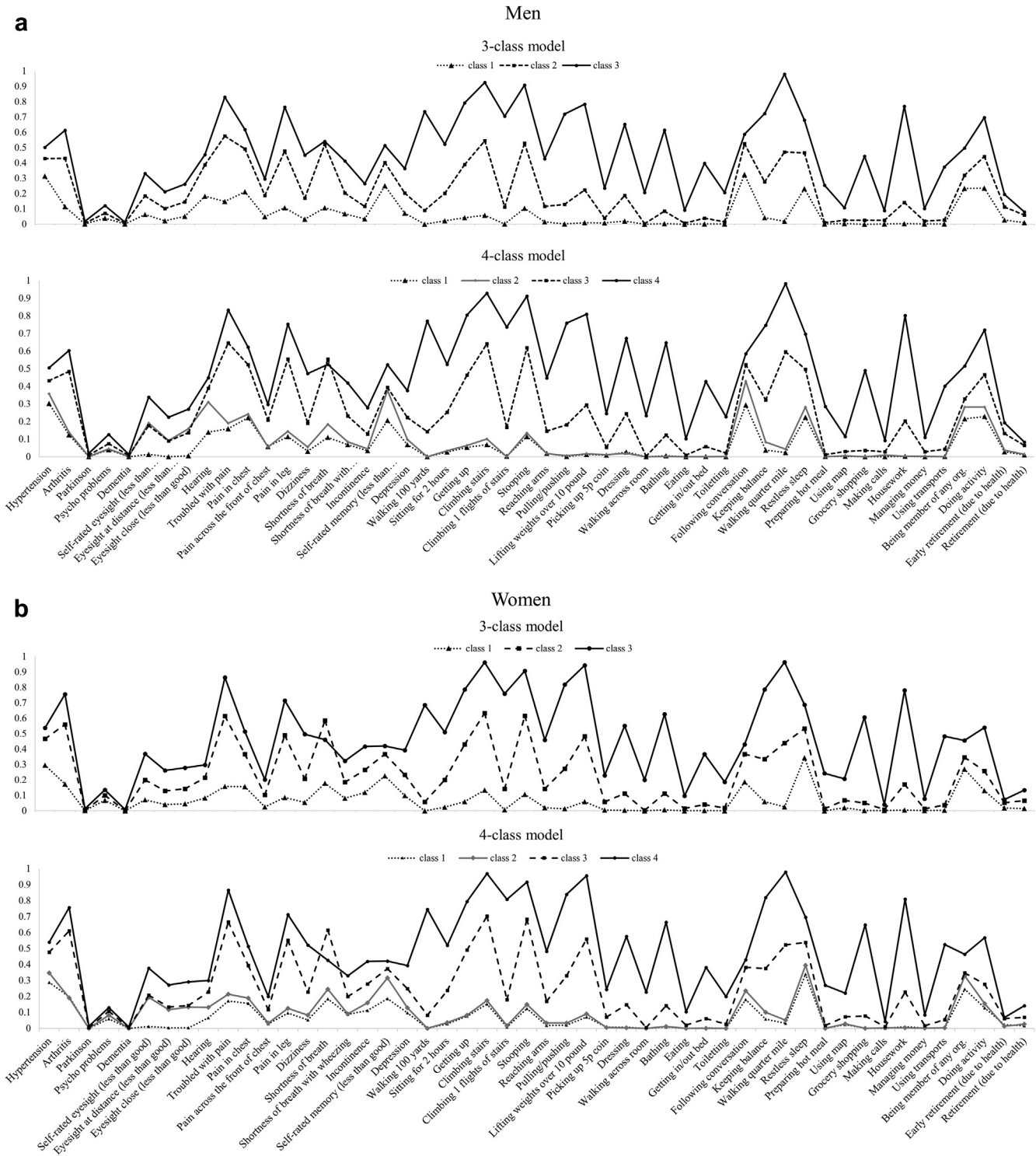


Fig. 3. a Probability of each disability item estimated by the 3 and 4-class models, men. b Probability of each disability item estimated by the 3 and 4-class models, women.

disability groups, the differences in grip strength, mental health at the first wave and odds of dying relative to non-disabled group was progressively larger as disability increased (intercept coefficients).

Therefore, comparing the results of the growth models for each disability specification, we observe that estimates for binary disability in terms of size were in between the coefficients observed for mild and severe disability in multi-categorical disability. Intercept coefficients of the association between mild disability in the 3-

class model and each outcome were very close to those of moderate disability in the 4-class model; and intercept coefficients of severe disability in 3-class and 4-class models were close to each other. The lowest disability group in the 4-class model (i.e. mild disability) presented significant worse health and mortality conditions compared to no disability group, but the disadvantage was smaller than in the other disability groups. Such a small but significant disadvantage was not captured when identifying only 3 categories

**Table 2**  
Association of disability groups with mortality and a number of mental and physical health indicators for a selection of models<sup>a</sup>.

Women						
Specification of disability level		Mean grip strength		Mental Health		Death
		Intercept	Slope	Intercept	Slope	β coef
2 classes (WHO)	disabled	-3.387*** (-3.9; -2.88)	0.259*** (0.13; 0.39)	1.099*** (0.98; 1.22)	-0.006 (-0.04; 0.03)	1.456*** (1.219; 1.740)
2 classes (ROC curve min r)	disabled	-3.439*** (-3.96; -2.92)	0.236*** (0.11; 0.37)	1.108*** (0.98; 1.23)	-0.008 (-0.05; 0.03)	1.486*** (1.244; 1.774)
2 classes (ROC curve max j)	disabled	-3.702*** (-4.23; -3.17)	0.285*** (0.15; 0.42)	1.115*** (0.99; 1.24)	-0.016 (-0.05; 0.02)	1.452*** (1.217; 1.732)
2 classes - LCA	disabled	-3.68*** (-4.21; -3.15)	0.281*** (0.15; 0.42)	1.223*** (1.1; 1.35)	-0.025 (-0.06; 0.01)	1.407*** (1.18; 1.678)
3 classes - LCA	mild	-2.092*** (-2.56; -1.62)	0.053 (-0.06; 0.17)	0.827*** (0.71; 0.94)	0.014 (-0.02; 0.05)	1.333*** (1.094; 1.624)
	severe	-5.398*** (-6.16; -4.63)	0.363*** (0.17; 0.56)	1.971*** (1.8; 2.15)	-0.072*** (-0.12; -0.02)	1.779*** (1.390; 2.276)
4 classes - LCA	mild	-0.594** (-1.12; -0.07)	0.003 (-0.12; 0.13)	0.191*** (0.07; 0.31)	-0.002 (-0.04; 0.03)	1.212 (0.941; 1.563)
	moderate	-2.604*** (-3.14; -2.07)	0.118* (-0.01; 0.25)	0.94*** (0.81; 1.07)	0.003 (-0.03; 0.04)	1.477*** (1.176; 1.855)
	severe	-5.829*** (-6.69; -4.97)	0.32*** (0.1; 0.54)	1.994*** (1.8; 2.18)	-0.082*** (-0.14; -0.02)	1.850*** (1.406; 2.433)

Men						
Specification of disability level		Mean grip strength		Mental Health		Death
		Intercept	Slope	Intercept	Slope	Beta
2 classes - WHO	disabled	-2.396*** (-3.2; -1.59)	0.038 (-0.17; 0.24)	1.164*** (1.05; 1.28)	-0.071*** (-0.11; -0.04)	1.548*** (1.306; 1.836)
2 classes - ROC curve	disabled	-2.452*** (-3.28; -1.62)	0.022 (-0.19; 0.24)	1.175*** (1.06; 1.29)	-0.073*** (-0.11; -0.04)	1.529*** (1.291; 1.810)
2 classes - LCA	disabled	-3.115*** (-4.02; -2.21)	0.029 (-0.21; 0.27)	1.311*** (1.19; 1.44)	-0.068*** (-0.11; -0.03)	1.439*** (1.213; 1.708)
3 classes - LCA	mild	-1.867*** (-2.67; -1.06)	0.09 (-0.11; 0.29)	0.854*** (0.74; 0.97)	-0.055*** (-0.09; -0.02)	1.349*** (1.131; 1.608)
	severe	-4.691*** (-5.94; -3.44)	0.288* (-0.05; 0.62)	1.978*** (1.81; 2.14)	-0.147*** (-0.2; -0.09)	1.726*** (1.375; 2.165)
4 classes - LCA	mild	-0.669* (-1.45; 0.12)	-0.046 (-0.24; 0.14)	0.234*** (0.13; 0.34)	0.002 (-0.03; 0.03)	1.087 (0.885; 1.335)
	moderate	-2.22*** (-3.13; -1.31)	0.041 (-0.19; 0.27)	1*** (0.87; 1.13)	-0.054*** (-0.09; -0.02)	1.529*** (1.252; 1.868)
	severe	-5.474*** (-6.84; -4.11)	0.426** (0.06; 0.79)	2.179*** (2; 2.36)	-0.19*** (-0.25; -0.13)	1.978*** (1.544; 2.533)

β coef, β coefficient; 95% confidence intervals in brackets (); \*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1.  
<sup>a</sup> All models are adjusted for all confounders.

of disability.

Finally, regarding the rate of change over time, as captured by the slopes, in most cases multi-categorical definitions of disability did not seem to identify significant differences in changes in health conditions over time. In the few cases where there was evidence of association, we observed a 'protective effect' of disability, where the worsening of the condition was slower among disabled -whatever the level of disability was-compared to their non-disabled counterpart, indicating a convergence of the trajectories over time.

**Discussion**

In this paper we sought to examine whether a finer classification of disability in the 50 + English population may be useful for both descriptive and health policy planning purposes. To the best of our knowledge, this is the first study attempting to identify categories of disability based on empirical evidence, rather than upon a priori theoretical classification, for the English population. Along with the identification of the most adequate number of disability categories, the results produced in this study also help to understand the relevance of classifying disability correctly and describe the main characteristics of each category. Our conclusion suggests that the best classification of disability consists of four classes. While

acknowledging that some arbitrariness and subjectivity affect this conclusion, there are also multiple sources of evidence lending support to our decision. Below we elaborate why we think that the optimal number of disability levels is four and discuss why it is important to go beyond a binary definition of disability.

The optimal number of disability classes was chosen based on: (i) fit indices; (ii) estimated probabilities of disability items conditional on class membership; and (iii) external validation based on the association of disability classes with health and mortality. The first criterion did not point to a preferred model. When we looked at probabilities of endorsing disability items for each class, some evidence for making a decision emerged. We wanted to identify as many grades of disability as each group presented a different probability of having problems with each disability item. The support to the 4-class model came considering and comparing all the class models, and therefore observing that in the 5 and 6-class models four patterns appeared, and with 4 classes the distinction between the two lowest levels was small, but still appreciable. The third criterion confirmed this conclusion by showing the existence of a disability gradient associated with mortality and health consisting of four levels of disability. The identification of three categories of disability and a no-disability group (i.e. 4-class model) enabled us to capture the strong effect of severe disability as well as

the intermediate impact of moderate disability, and also the small but significant disadvantage in health and mortality experienced by those affected by mild disability compared to the non-disabled. This small but significant effect experienced by the mildly disabled had a correspondence in the finding that the item profiles of non-disabled and mildly disabled in the 4-class model were very similar to each other, but with some appreciable differences, especially in the sensory functions in the impairment domain. Therefore, our results suggest that the best classification of disability has four classes, consisting of “no disability”, “mild disability” which presents characteristics more similar to no disability than the other disability levels, and “moderate” and “severe disability”.

In the introduction, we pointed out the relevance of going beyond a binary definition of disability, and mentioned recent studies<sup>6,7</sup> that have stressed the importance of identifying more finely graded measures of disability. Our findings contribute to this debate by showing the loss of information due to too broad categorizations of disability, compared to a more refined scale of severity. Using the 4-class model as a reference, we observed that the magnitude of the association between binary disability and all outcomes lied in between the results observed for the groups suffering from the most severe disability (moderate and severe). In the growth model for the relationship between disability and health outcomes, the intercept coefficients of binary disability were just slightly higher than the intercept coefficient of moderate disability. The binary indicator averaged the burden of disability and masked the very strong effect experienced by individuals having severe disability, and was not informative for low levels of disability. As shown by Jagger et al. (2015),<sup>6</sup> life expectancy with mild-disability is expanding, meaning that the number of years expected to be lived in mild-disability is increasing over time and the proportion of life expected to be lived without mild disability has decreased. If adequate grades of severity are not identified, it is not possible to monitor low disability, and equally it is impossible to assess the strong impact of severe disability.

Finally, we report some limitations that affect this work. First, we used estimated probabilities of class membership as a covariate in regression analysis and this ignores misclassification error. Various 3-step methods have been proposed to account for this.<sup>25,26</sup> We acknowledge this here, but it was beyond the scope of this work to explore in details this technical aspect. Moreover, the fact that entropy was high and adequate separation between the identified latent classes was found suggests that results would be very similar. Second, while latent variable modelling was in many respects a natural approach to measure disability, one issue is that it is data dependent which asks for further research to compare findings across different settings. A possible step to validate our results would be to replicate the analysis using the ELSA's sister studies, such as the Health and Retirement Study (HRS), Survey of Health, Ageing and Retirement (SHARE), which offer a substantially similar set of variables and are targeted on the same type of population of ELSA (adults aged 50+) but in other countries.

Lastly, we mention a feature of the study which limits the interpretation of results and calls for future research. While we investigated the association between disability and health outcomes to assess the grades of disability, we came across interesting results, especially concerning the effect of disability over time. Particularly unexpected was the finding that in some cases the health status of disabled individuals seemed to deteriorate at a slower pace compared to their non-disabled counterparts (where slope coefficients were significant and with opposite sign compared to intercept coefficients). We hypothesize that it may be due to the fact that non-disabled people are more likely to become disabled over time, and this would explain why slope coefficients

were significant and suggesting a protective effect of disability especially among the most disabled group, which cannot become any more disabled, and therefore the impact of their level of disability at baseline on health and mortality over time is more likely to remain the same. The nature of the data limits a more detailed investigation of this finding, and most research on the association between disability and anthropometric measures for physical domains and mental health is cross-sectional limiting the conclusions that can be drawn about causality. This opens the doors to further research to robustly explore our hypothesis, for example treating disability as a time varying variable and observing its trajectory and its impact on health.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Conflict of interest

The authors declare that they have no conflict of interest.

### Funding

Charges for publication would be supported by the Economic and Social Research Council (grant number ES/J500021/1).

### Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.dhjo.2017.01.005>.

### References

- Palmer M, Harley D. Models and measurement in disability: an international review. *Health Policy Plan.* 2012;27(5):357–364.
- Wolf D. Chapter 4-late-life Disability Trends and Trajectories. *Handbook of Aging and the Social Sciences.* eighth ed. San Diego: Academic Press; 2016:77–99.
- Wolf DA, Hunt K, Knickman J. Perspectives on the recent decline in disability at older ages. *Milbank Q.* 2005;83(3):365–395.
- Freedman VA. Implications of asking “ambiguous” difficulty questions: an analysis of the second wave of the asset and health dynamics of the oldest old study. *J Gerontol Ser B Psychol Sci Soc Sci.* 2000;55(5):S288–S297.
- Jette AM. How measurement techniques influence estimates of disability in older populations. *Soc Sci Med.* 1994;38(7):937–942.
- Jagger C, Matthews FE, Wohland P, et al. A comparison of health expectancies over two decades in England: results of the cognitive function and ageing study 1 and 2. *Lancet.* 2015;387(10020):779–786.
- Wolf DA, Freedman VA, Ondrich JL, Seplaki CL, Spillman BC. Disability trajectories at the end of life: a “countdown” model. *J Gerontol B Psychol Sci Soc Sci.* 2015;70(5):745–752.
- Manton KG, Gu X. Disability declines and trends in medicare expenditures. *Ageing Horiz.* 2005;2(1):25–34.
- Serlin RC, Mendoza TR, Nakamura Y, Edwards KR, Cleeland CS. When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. *Pain.* 1995;61(2):277–284.
- Honey A, Emerson E, Llewellyn G. The mental health of young people with disabilities: impact of social conditions. *Soc Psychiatry Psychiatr Epidemiol.* 2011;46(1):1–10.
- Feyzi V, Hasheminejad N, Jafariroodbandi AS, Zolala F, Asadi A. Relationship between hand grip strength and functional disability of the elderly living in nursing homes in Kerman, Southeastern Iran. *Int J Occup Hyg.* 2016;8(1):39–44.
- Hairi NN, Cumming RG, Naganathan V, et al. Loss of muscle strength, mass (sarcopenia), and quality (specific force) and its relationship with functional limitation and physical disability: the concord health and ageing in men project. *J Am Geriatr Soc.* 2010;58(11):2055–2062.
- World Health Organization. *World Bank. World Report on Disability.* Geneva, Switzerland: World Health Organization; 2011.
- Stephens A, Breeze E, Banks J, Nazroo J. Cohort profile: the english longitudinal study of ageing. *Int J Epidemiol.* 2013;42(6):1640–1648.
- Pongiglione B, De Stavola BL, Kuper H, Ploubidis GB. Disability and all-cause mortality in the older population: evidence from the english longitudinal study of ageing. *Eur J Epidemiol.* 2016;31(8):735–746.
- Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JA. Methods for dealing with time-dependent confounding. *Stat Med.* 2013;32(9):1584–1618.
- Centers for Disease Control and Prevention, Washington Group on Disability Statistics 2015 [cited 2016 25 Feb], Available from: <http://www.cdc.gov/nchs/>



- washington\_group/wg\_questions.htm.
18. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29–36.
  19. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr*. 2011;48(4):277–287.
  20. Office for National Statistics, General Lifestyle Survey Overview, a report on the 2011 General Lifestyle Survey 2013 [cited 2016 April 22, 2016], Available from: <http://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/compendium/generallifestylesurvey/2013-03-07/chapter7generalhealthgenerallifestylesurveyoverviewareportonthethe2011generallifestylesurvey>.
  21. Vermunt JK. Multilevel latent class models. *Sociol Methodol*. 2003;33(1):213–239.
  22. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Struct Equ Model A Multidiscip J*. 2007;14(4):535–569.
  23. D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham heart study. *Stat Med*. 1990;9(12):1501–1515.
  24. Muthén LK, Muthén BO. *Growth Modeling with Latent Variables Using Mplus: Introductory and Intermediate Growth Models*. 2010. Available from: .
  25. Bakk Z, Tekle FB, Vermunt JK. Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociol Methodol*. 2013;43(1):272–311.
  26. Bakk Z, Vermunt JK. Robustness of stepwise latent class modeling with continuous distal outcomes. *Struct Equ Model A Multidiscip J*. 2016;23(1):20–31.