**Methods**

**Data**

We analysed data from the English Longitudinal Study of Ageing (ELSA) - a nationally representative panel study of men and women who were aged 50 and over and living in private households in England at baseline (Banks, 2021; Steptoe et al., 2013). ELSA collects both subjective and objective measures of mental and physical health, finances, and attitudes around ageing. The data are collected face-to-face using computer-assisted interviewing (CAPI), combined with self-completion questionnaires completed using pen and paper (PAPI). Ethical approval for the ELSA study was granted by the National Research Ethics Service (Steptoe et al., 2013).

To date, nine waves of data collection have taken place over a period of 17 years (2002-2019), with biennial follow-ups. For the current analyses, we used data spanning 11 years from Wave 4 (2008-2009) to Wave 9 (2018-2019), meaning participants were aged XX and over at Wave 4 and XX and over at age Wave 9. Information on relevant health-risk behaviours (i.e., SNAP behaviours) was collected from Wave 3 onwards. However, we chose Wave 4 as the baseline wave to maximise the number of sample members for whom longitudinal weights were available (Ref).

To answer our research questions, a longitudinal sample of those who were interviewed in-person at all six waves (i.e. Waves 4 to 9) was needed. Thus, six separate datasets containing ELSA responses to the core self-completion questionnaire (covering questions on wellbeing, health behaviours and health outcomes, among others) for Waves 4 to 9 were merged into a single dataset using unique respondent identification numbers. Thus, a total of 5429 participants who answered the core questionnaire across all six waves (i.e. Waves 4 to 9) were included in the current analyses; participants who had not responded to the questionnaire in one or more of the included waves were excluded.

**Health behaviour measures**

We used an online Qualtrics survey (see Appendix A) to gather consensus from experts on how SNAP behaviours should be *defined* (i.e. which measure to choose from the options available in the ELSA dataset) and *categorised* to reflect risk status (i.e., how many categories to divide each behaviour into and what cut-offs to use). Researchers who had published at least one relevant article, as the first author, in the last three years on any of the four SNAP behaviours were identified through the research team’s personal networks. Out of an opportunistic sample of 20 researchers, 14 experts participated in our survey. The results of the survey indicated a consensus (defined as agreement of 70% or greater) on the most appropriate measure of fruit and vegetable intake and physical activity behaviours. Regarding behaviours for which there was no clear consensus (such as alcohol drinking and smoking), the authors engaged in internal discussions and reached a conclusion.

***Smoking***

We used current smoking status as a measure of smoking. The data comprised participants’ binary (yes/no) response to the question ‘Do you smoke at all nowadays?’

***Fruit and vegetable intake***

In Wave 4, fruit and vegetable intake was assessed with 13 items such as ‘How much of the following did you eat yesterday’ for categories including ‘small glass of fruit juice’ and ‘salad (cereal bowlfuls)’. In Waves 5 to 8, however, participants were asked ‘how many portions of vegetables – excluding potatoes –do you eat on a given day?’ and ‘how many portions of fruits do you eat on a given day?’

To make fruit and vegetable intake consistent across waves, answers from Wave 4 were recoded to match the measure used in Waves 5 to 8. To do this, we used a conversion method that has been previously used to conduct longitudinal analyses on fruit and vegetable intake from the ELSA study (Hackett et al., 2018; Kojima et al., 2020; conversion rates adapted from Kojima et al. (2020) are shown in Appendix B).

Finally, the portions of fruits and vegetables reported were added to create a single variable representing the total number of fruit and vegetable portions (i.e., servings) consumed per day for each Wave. This data was subsequently divided into two categories (<5 or ≥5 portions per day) (Kojima et al., 2020).

***Alcohol Consumption***

We included data on alcohol consumption over the last week: Specifically, i) whether the participant consumed alcohol in the last week (yes/no), ii) the number of days that alcohol was consumed in the last week (1-7 days), and iii) the volume of different types of alcoholic beverages (e.g. beer, wine, spirits) consumed in the last week.

The data was categorised on the basis of the number of units of alcohol that they consumed. To do this, the amount of alcohol consumed was converted into the number of UK units (1 UK unit = 10 ml of pure alcohol). The alcohol by volume (ABV) assumptions for converting alcoholic beverages into the number of UK units were calculated using standard assumptions for beverage strength (Ref). Consumption was divided into four levels based on units consumed per week: *harmful* (>50 units for men, >35 units for women); *hazardous* (15-50 units for men and 15-35 units for women); *moderate* (14 units or less); *and abstainers* (0 units) (Department of Health, 2016).

***Physical Activity***

ELSA recorded physical activity by asking participants how often they took part in each of three different types of physical activity: vigorous-intensity (e.g., running/ jogging, swimming, cycling, aerobics/gym workout, tennis, and digging with a spade), moderate-intensity (gardening, cleaning the car, walking at moderate pace, dancing) and low-intensity (laundry and home repairs). The response categories were: hardly ever/never, one to three times a month, once a week and more than once a week. For the present analysis, a summary index of physical activity was derived by adding responses to all three physical activity questions based on the classification in other ELSA studies (Dhalwani et al., 2016; Hamer et al., 2014). These were: *sedentary* (no activity on a weekly basis); *low* (only mild activity at least once a week); *moderate* (moderate but no vigorous activity at least once a week); and *high* (any vigorous activity at least once a week).

***Sociodemographic variables***

All socio-demographic variables were based on Wave 4 responses. *Age* was derived from participants’ date of birth. For *sex*, participants were asked to code their responses as either female or male. *Parental occupation* (i.e., main carer that the participants lived with at the age of 14 years) was categorized as high (managerial, professional and administrative occupations or business owners), intermediate (trade and services related occupations) and low (manual, casual occupations and other occupations). Participants who stated their parent’s occupation as retired (n=19), unemployed (n=36) and / or sick/disabled (n=27) were marked as missing, since they could not be accurately categorised. Participants’ *own occupation* was measured using the three-class version of the National Statistics—Socioeconomic Classification Scheme (Elias & McKnight, 2003) and was categorized as high (managerial and professional occupations), intermediate (intermediate occupations) and low (semi-routine and routine occupations). Participants with occupations listed as other (n=12) – those that either had never worked, were full-time students, had occupations not stated or inadequately described, or were not classifiable for other reasons – were marked as missing. *Education* was categorized according to the Harmonized education scale used in ELSA, which is a simplified version of 1997 International Standard Classification of Education (ISCED-97) codes (Phillips et al., 2014). Education was therefore categorized as [high (Tertiary education), intermediate (Upper secondary & vocational training) and low (Less than lower secondary education). *Wealth*, i.e., total net non-pension household wealth at benefit unit level (benefit unit is a couple or single person with any dependent children) was grouped into tertiles.

***Health outcomes***

ELSA records data on a range of physical and mental health conditions. The present study used reports of health conditions from ELSA Wave 9. Participants were asked whether they still had the condition diagnosed by a doctor that they had reported in previous Wave(s) and if not, whether they had a new condition to report (for the list of condition recorded, refer to the first column of Table 1). We categorised these conditions into pathologies of eight body systems : eye disorders; circulatory disorders; nervous disorders; mental and behavioural problems; neoplasms; respiratory disorders; endocrine, nutritional and metabolic disorders; and musculoskeletal and connective system disorders. These are based on the grouping system used by Singer et al. (2019) examining complex multimorbidity using the ELSA dataset, where body systems were defined as per the chapters of the International Classification of Diseases 10th Revision system (see the second column of Table 1).

Finally, we calculated 10 binary variables representing the self-reported presence or absence of following health outcomes: i) pathologies of each of eight body systems, ii) multimorbidity (having two or more of 25 health conditions) and iii) complex multimorbidity (defined by (Harrison et al., 2014) as three or more chronic conditions co-occurring within one person that affect three or more different body systems).

**Table 1.**

*Data on morbidities used to ascertain basic multimorbidity and complex multimorbidity*

|  |  |  |
| --- | --- | --- |
|  | Morbidities | Body systems |
| 1 | High blood pressure | **1. Eye disorders** |
| 2 | Angina | 1.1. Glaucoma |
| 3 | Congested heart failure | 1.2. Macular degeneration |
| 4 | Heart murmur | 1.3. Cataracts |
| 5 | Abnormal heart rhythm | **2. Circulatory disorders** |
| 6 | Heart attack | 2.1. High blood pressure |
| 7 | Diabetes | 2.2. Angina |
| 8 | Stroke | 2.3. Heart Attack |
| 9 | Lung disease | 2.4.Congestive heart failure |
| 10 | Asthma | 2.5. Heart murmur |
| 11 | Arthritis | 2.6. Abnormal heart rhythm |
| 12 | Osteoporosis | 2.7. Stroke |
| 13 | Cancer | **3. Endocrine, nutritional and metabolic** |
| 14 | Parkinson’s disease | 3.1. Diabetic eye disease |
| 15 | Dementia | 3.2. Diabetes |
| 16 | Alzheimer’s disease | **4. Musculoskeletal and connective system** |
| 17 | Hallucinations | 4.1. Osteoporosis |
| 18 | Anxiety | 4.2. Arthritis |
| 19 | Depression | **5. Respiratory** |
| 20 | Emotional problems | 5.1. Lung disease |
| 21 | Mood swings | 5.2. Asthma |
| 22 | Glaucomes | **6. Neoplasms** |
| 23 | Diabetic eye disease | 6.1. Cancers |
| 24 | Macular degeneration | **7. Nervous disorders** |
| 25 | Cataracts | 7.1. Parkinson’s disease |
|  |  | 7.2. Alzheimer’s disease |
|  |  | 7.3. Hallucinations |
|  |  | **8. Mental and behavioural** |
|  |  | 8.1. Anxiety |
|  |  | 8.2. Depression |
|  |  | 8.3. Emotional problems |
|  |  | 8.4. Mood swings |
|  |  | 8.5 Dementia |

*Note.* Adapted from Singer, L., Green, M., Rowe, F., Ben-Shlomo, Y., Kulu, H., & Morrissey, K. (2019). Trends in multimorbidity, complex multimorbidity and multiple functional limitations in the ageing population of England, 2002–2015. *Journal of comorbidity*, *9*, 1-9.

**Creating the data sets for analysis**

Some participants (*n* = 1642) were excluded from the merged dataset (*n* = 5429), due to missing values on one or more socio-demographic variables, leaving a final sample of *n*=3787. This is because MPlus v8.5 software package (Muthén & Muthén, 2017) cannot handle missing data on predictors of latent classes and uses listwise deletion for cases that have missing data on any of the predictors. Since we took advantage of the strengths of the new three-step manual LCA approach (explained subsequently) and had a complex data set, we were unable to apply advanced approaches to addressing missing data.

**Statistical analysis**

Repeated measures latent class analysis (RMLCA) using the full information maximum likelihood (FIML) algorithm was conducted in MPlus v8.5 software package and R version v4.0.3 (Muthén & Muthén, 2017). RMLCA was used to identify whether distinct latent classes described those with similar patterns of four health behaviours over time, hereafter denoted as classes or latent classes (Aim 1). For this, we iteratively examined a series of LCA models with increasing number of classes till the model failed to converge, to identify the fewest number of classes that best represented the latent variable.These models were compared using several fit indices and classification diagnostics: Consistent Akaike’s Information Criterion (CAIC), Bayesian Information Criterion (BIC), adjusted Bayesian Information Criterion (aBIC), Approximate Weight of Evidence Criterion (AWE), Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (VLMR-LRT)(Schreiber, 2017). Lower values of the BIC, CAIC, saBIC, and AWE indicate a better fitting model. For the likelihood-based tests such as the VLMR-LMR test, a p value < .05 was taken to indicate that the model fit has not significantly improved compared to the model with one less class. In cases of disagreement between the test statistics, IC indices (i.e., AIC, BIC, SABIC), followed by VLMR-LRT and model interpretability were used as the primary criteria for selection of the optimal number of classes. Thus, an “elbow”plot using IC indices was used to identify at which step the improvement in model fit plateaued and the models around that point were examined. Entropy and the smallest average latent class posterior probability were also utilised to quantify how well individuals were classified into the latent classes, with values closer to 1 indicating higher classification quality (i.e., good separation between classes).

To assess the sociodemographic characteristics of class membership and assess the differences in health outcomes across the classes, the 3-step BCH approach and the 3-step manual approach for modelling auxiliary variables in LCA were respectively used (Ref). In both these procedures, the latent class model without any auxiliary variables is estimated in the first step. In the second step, the latent class posterior distribution is used to assign participants to the class that they have the highest probability of falling into and this assignment is treated as an imperfect measure containing uncertainty rates. In the final step the auxiliary variables (i.e. sociodemographic variables and health outcomes) are included in the model.

We used the BCH method to estimate sociodemographic characteristics for each class as well as the between-class sociodemographic differences using analyses of variance and chi-square analyses (Aim 2). For assessing differences between the class-specific outcomes while controlling for sociodemographic variables (Aim 3), we conducted binomial logistic regressions with latent class as the independent variable and the health outcomes as the dependent variable and used pairwise Wald test results to check for significant differences in health outcomes across the latent classes. A p-value of <0.05 was considered statistically significant.

**Sensitivity analyses**

To test the robustness of the final solution, the sample was split into 50% random subsamples, and LCA was conducted on both subsamples to determine whether the solution replicated. Reliability was evaluated by comparing the classification results of the split-halves and combined sample.

**Results**

**Sample characteristics**

In Table 2, characteristics are listed for the final, complete case sample (*n* = 3787) with no missing data on sociodemographic variables. Relative to study members that were excluded (*n* = 1642), those included in the present analyses were less likely to be female (55.43%), had higher proportions of individuals with : high parental occupation (41.56%), high level of occupation (40.08%) and upper secondary education (51.33% ; see Table 2 for comparison of socio-demographic characteristics between the excluded and complete case sample). Although the absolute differences between the two samples were not considerable, they attained statistical significance owing to the large study numbers.

**Table 2.**

*Descriptive data and test of representativeness of the complete case sample (i.e. with no missing data on sociodemographic variables)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sociodemographic variables | Complete Case Data  (n=3787) | Excluded Data  (n=1642) | Test estimate | p-value |
| Female | 55.43% | 59.56% | **7.81\*** | **0.005\*** |
| Average Age | 62.73 years | 62.35 years | 0.392 | 0.101 |
| Parental Occupation - *Intermediate* | 31.63% | 31.44% | 0.011 | 0.917 |
| Parental Occupation - *High* | 41.56% | 37.54% | **6.836\*** | **0.009\*** |
| Occupation - *Intermediate* | 25.67% | 26.01% | 0.048 | 0.826 |
| Occupation - *High* | 40.08% | 34.13% | **15.322\*** | **0.000\*** |
| Education – *Upper Secondary* | 51.33% | 44.76% | **7.838\*** | **0.005\*** |
| Education – *Tertiary* | 20.97% | 17.98% | 2.377 | 0.123 |
| Wealth – *Second Tertile* | 33.64% | 33.95% | 0.029 | 0.864 |
| Wealth – *Third Tertile* | 40.01% | 37.34% | 2.854 | 0.091 |

*Note.* The differences for all variables except age were calculated using the two-sample test for equality of proportions. For age, the two-sample Welsh t-test was used to evaluate the differences.**\*p<.05.**

**Identifying latent classes of health behaviour trajectories**

LCA models with increasing number of classes were iteratively fit on four health behaviours indicators measured across five time points and inspected using certain fit indices (See Table 3). We inspected the likelihood-based test—the Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test (VLMR-LRT). In the present case, the test lent support for the three class solution. However, since the VLMR-LRT is considered a useful reference when the sample size is small (i.e., n ≤ 630), with a large sample size caution is recommended. In such cases, BIC and CAIC are preferred (Chen et al., 2017). Thus, we examined fit information criteria (IC) — the BIC, SABIC, CAIC, and AWE — where lower values indicate superior model fit. In the present case, the ICs continued to decrease for each additional class added. Thus, we inspected their plots for an “elbow” point of “diminishing returns” in model fit (e.g., small decreases in the IC for each additional latent class).

**Table 3.**

*Fit statistics for class enumeration of four health behaviours (i.e. smoking, alcohol consumption, physical activity and fruit and vegetable intake) across time*

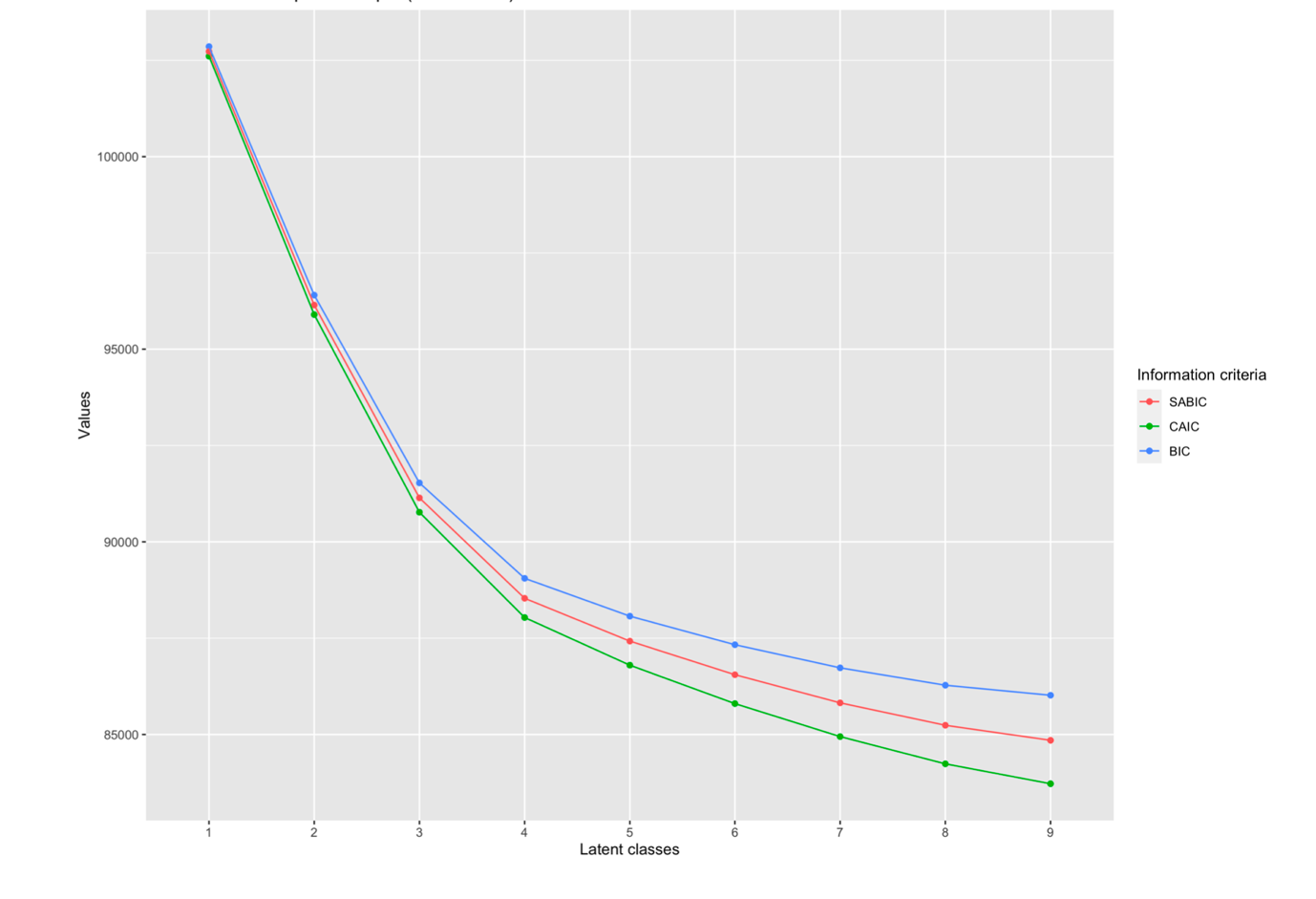
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| K | LL | CAIC | BIC | SABIC | AWE | VLMR LRT  p-value | Entropy | Smallest average latent class posterior probability |
| 1 | -51263.46 | 102606.93 | 102856.50 | 102729.40 | 102916.5 | - | - |  |
| 2 | -47867.12 | 95896.24 | 96401.62 | 96144.24 | 96523.13 | 0.000 | 0.808 | 0.941 |
| 3 | -45261.57 | 90767.14 | 91528.34 | 91140.68 | 91711.34 | 0.000 | 0.878 | 0.942 |
| 4 | -43854.75 | 88035.50 | 89052.51 | 88534.58 | 89297.01 | 0.4069 | 0.864 | 0.897 |
| 5 | -43195.22 | 86798.45 | 88071.27 | 87423.06 | 88377.27 | 0.7609 | 0.848 | 0.874 |
| 6 | -42655.66 | 85801.31 | 87329.95 | 86551.45 | 87697.45 | 0.7722 | 0.837 | 0.856 |
| 7 | -42186.75 | 84845.51 | 86729.96 | 85821.18 | 87158.96 | 0.7648 | 0.826 | 0.823 |
| 8 | -41792.18 | 84238.36 | 86278.62 | 85239.57 | 86769.12 | 07641 | 0.831 | 0.818 |

*Note.* *n*=3787; K = number of classes (the nine-class model failed to converge); LL = model log likelihood; BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE = approximate weight of evidence criterion; VLMR-LRT = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test; p-value significance <0.05;

We found the elbow-point around the 6-class model for the three ICs and examined the models around this point, i.e. five-,six- and seven-class models (see Figure 1). Based on conceptual grounds and considering all model fit statistics, the authors chose a six-class model. Further, the entropy and smallest average latent class posterior probability also fell within the recommended range (value ≥.8) for the six-class model (Nylund-Gibson & Choi, 2018).

**Figure 1.**

*Elbow plot of information criteria for latent class models*

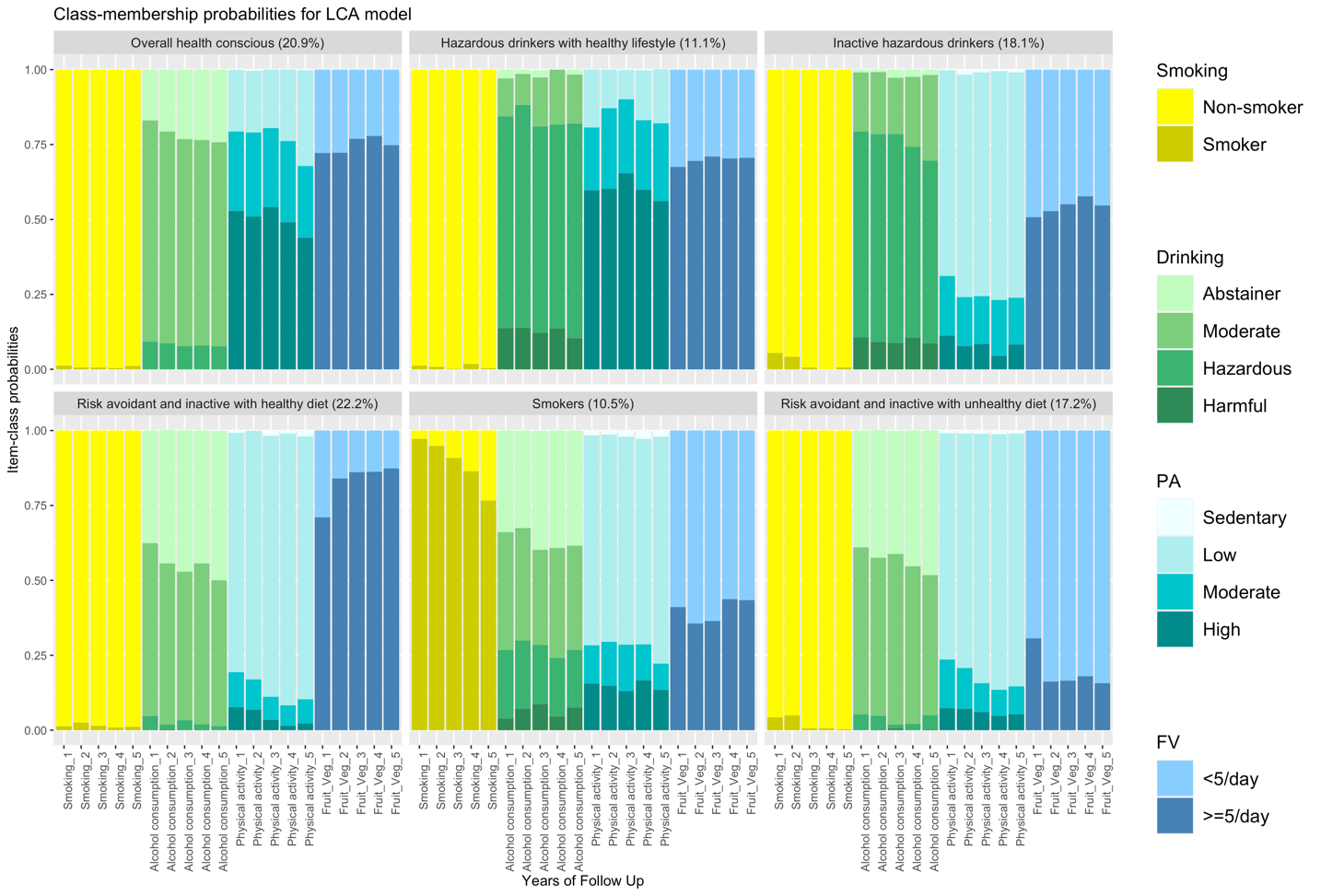
****

*Note.* BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion

Figure 2 represents the selected six-class model and graphically illustrates the characteristics of the six classes based on responses to the health behaviour indicators. A fifth of all participants (~21%) were in the *Overall health conscious* class. This class had a consistently high probability of drinking at moderate levels and maintained high levels of physical activity and adequate (>5 per day) fruit and vegetable intake across time. A tenth of all participants (11%) were in the *Hazardous drinkers with a healthy lifestyle* class since they exhibited a high probability of drinking at hazardous levels despite maintaining high levels of physical activity and adequate (>5 per day) fruit and vegetable intake across the time points considered. Conversely, a larger percentage of the sample (18%) fell in *Inactive hazardous drinkers* category on account of a high probability od drinking at hazardous levels while displaying consistently low levels of physical activity. However, the largest percentage of participants (22%) belonged to the *Risk avoidant and inactive with a healthy diet,* as they had a low probability of engaging in any risk behaviour (i.e. smoking or alcohol consumption) and displayed low levels of physical activity while consuming the recommended portions of fruit and vegetable across time. A fifth of the sample (10.5%) in the *Smokers* class were characterised by a steadily declining but high probability of smoking. Finally, the *Risk avoidant and inactive with unhealthy diet* class represented 17% of all participants and had a similar health profile to the *Risk avoidant and inactive with a healthy diet,* and differed only account of having a consistently high probability of inadequate fruit and vegetable intake and a slightly higher probability of smoking.

**Figure 2.**

*Latent class prevalences and item-class probabilities in the selected six-class latent model of health behaviours across time*

**

Note. The *x*-axis lists each of the four behaviours – smoking, alcohol consumption, physical activity, and fruit and vegetable intake - across five time points. The *y*-axis provides the average probability for each of the indicators (i.e. four health behaviours) conditional on membership in a given class.

**Sensitivity analysis**

Reliability of the six class model was evaluated by comparing the classification results of the 50% random split-halves of the sample and combined sample (see Supplementary Material). The structure (i.e. latent class prevalences and indicator values) of the six-model was the most stable across the three samples and replicated consistently.

**Latent classes and sociodemographic characteristics**

The sociodemographic profile for each latent class is displayed in Table 4. Chi-square and ANOVA tests showed statistically significant differences (*p* ≤ 0.01) across the six latent classes for all of the sociodemographic variables except intermediate level of occupation, secondary education level and the second tertile of wealth.

Briefly, in the *Overall health conscious class*, the majority of assigned participants tended to fall within the highest tiers of parental occupation, self-occupation and wealth; however, the sex distribution was fairly even in this class. The *Hazardous drinkers with a healthy lifestyle* class had an even larger majorityin the highest tiers of parental occupation, self-occupation and wealth; however, this class had a male to female ratio of 70:30 – the highest in any class. A similar sociodemographic profile was also observed in the *Inactive, hazardous drinkers* class, albeit with a smaller proportion falling in the highest tiers of parental occupation, self-occupation and wealth as compared to *Hazardous drinkers with a healthy lifestyle* class*.*

Importantly*,* compared to the first three classes*,* the last three classes reveal a trend reversal in that the majority of participants in these classes fall in the lowest tier of parental occupation, self-occupation, education and wealth. Notably, although the *Smokers* and *Risk avoidant and inactive with a healthy diet* classes had a fairly even distribution of sexes, the *Risk avoidant and inactive with an unhealthy diet* had a male to female ratio of 30:70 – the lowest of any class.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Socio-demographic characteristics | Categories | 1  Overall health conscious | 2  Hazardous drinkers with a healthy lifestyle | 3  Inactive hazardous drinkers | 4  Risk avoidant and inactive with a healthy diet | 5  Smokers | 6  Risk avoidant and inactive with a healthy diet | p-values |
|  |  | (n= 20.9 %) | (n=11.1%) | (n=18.1%) | (n=22.2%) | (n=10.5%) | (n=17.2%) |  |
|  |  |  |  |  |  |  |  |  |
| Age | *(Cont.)* | 61.881 | 60.095 | 63.370 | 66.532 | 60.769 | 66.579 | **<0.01\*** |
| (in years) |  |  |  |  |  |  |  |  |
| Sex | Male | 47.300% | 70.200% | 64.600% | 29.200% | 45.300% | 42.300% | **<0.01\*** |
|  | Female | 52.70% | 29.80% | 35.40% | 70.80% | 54.70% | 57.70% | **<0.01\*** |
| Paternal | Low | 22.900% | 21.300% | 24.200% | 34.200% | 34.800% | 35.900% | **<0.01\*** |
| Occupation | Intermediate | 33.5% | 28.3% | 29.5% | 30.6% | 40.7% | 36.3% | **0.01\*** |
|  | High | 43.6% | 50.4% | 46.3% | 35.2% | 24.5% | 27.8% | **<0.01\*** |
| Occupation | Low | 30.5% | 17.3% | 34.9% | 51.8% | 51.9% | 51.4% | **<0.01\*** |
| - Self | Intermediate | 28.6% | 24.0% | 23.8% | 21.1% | 22.3% | 26.7% | 0.122 |
|  | High | 40.9% | 58.7% | 41.3% | 27.1% | 25.8% | 21.9% | **<0.01\*** |
| Education | Low | 25.5% | 14.2% | 25.9% | 45.9% | 49.2% | 51.1% | **<0.01\*** |
| Level | Secondary | 50.1% | 49.1% | 51.1% | 45.6% | 43.0% | 44.7% | 0.160 |
|  | High | 24.4% | 36.7% | 23.0% | 8.5% | 7.8% | 4.2% | **<0.01\*** |
| Wealth | First tertile | 15.2% | 11.8% | 21.1% | 39.7% | 52.1% | 47.0% | **<0.01\*** |
|  | Second tertile | 37.1% | 27.8% | 33.3% | 36.4% | 27.8% | 33.2% | 0.018 |
|  | Third tertile | 47.7% | 60.4% | 45.6% | 23.9% | 20.1% | 19.8% | **<0.01\*** |

**Table 4.**

*Baseline sociodemographic characteristics of the 3787 ELSA participants, stratified by latent class assignment*

*Note.* Age is a continuous variables and the p-value column indicates the results of an ANOVA for the row. All other socio-demographic characteristics are categorical and the p-value column indicates the results of a Chi square test for each row. **(\*=p<.05)**

**Latent classes and health outcomes**

Table 5 shows the results for the binomial logistic regressions of health outcomes at Wave 9 on latent classes. More specifically, it shows the pairwise comparisons for health outcomes at Wave 9 across latent classes controlling for sociodemographic variables (for unadjusted results, see Appendix C). It should be noted that the regression models for two outcomes: mental and behavioural issues and nervous systems disorders, were not reliable on account of too much missing data and were excluded.

Compared to the *Overall health conscious* class, all classes except the *Hazardous drinkers with a healthy lifestyle* class have a significantly higher proportion of participants withmultimorbidity and complex multimorbidity. Similarly, compared to the *Hazardous drinkers with a healthy lifestyle,* all classes except the *Overall health conscious* class have a significantly higher proportion of participants with complex multimorbidity. However, both the *Risk avoidant* classes have a higher proportion of participants with multimorbidity than the *Hazardous drinkers with a healthy lifestyle class.* Yet, between these two *Risk avoidant* classes, *the Risk avoidant and inactive with a healthy diet* class has a higher proportion of complex multimorbidity.

The proportion of respiratory disorders was significantly higher in the *Smokers* class compared to all other classes. For both musculoskeletal and connective tissue disorders, as well as circulatory disorders, the *Overall health conscious* class had a significantly smaller proportion of participants with these conditions than the *Inactive hazardous drinkers* and *Risk avoidant and inactive with a healthy diet* classes. Similarly, the *Hazardous drinkers with a healthy lifestyle* class had a smaller proportion of participants with circulatory disorders than *Risk avoidant and inactive with a healthy diet class*.

In the case of endocrine, nutritional and metabolic disorders, the *Risk avoidant and inactive with an unhealthy diet* and the *Risk avoidant and inactive with healthy diet* classes had a significantly higher proportion of participants with these disorders than all other classes, except each other (for which the difference was not significant). Conversely, the Hazardous drinkers with a healthy lifestyle class had a significantly smaller proportion of participants with endocrine, nutritional and metabolic disorders than *Smokers, Inactive hazardous drinkers* and *the Overall health conscious* class.

**Table 5.**

Pairwise comparisons across latent classes for health outcomes (adjusted for sociodemographic variables)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Health outcomes | Latent classes | Class proportions | 1  Overall health conscious | 2  Hazardous drinkers with a healthy lifestyle | 3  Inactive hazardous drinkers | 4  Risk avoidant and inactive with a healthy diet | 5  Smokers | 6  Risk avoidant and inactive with an unhealthy diet |
|  |  |  | *n*= 20.9 % | *n*=11.1% | *n*=18.1% | *n*=22.2% | *n*=10.5% | *n*=17.2% |
| Multimorbidity | 1 | 0.503 |  |  |  |  |  |  |
|  | 2 | 0.442 | -0.061 |  |  |  |  |  |
|  | 3 | 0.58 | **0.077\*** | 0.138 |  |  |  |  |
|  | 4 | 0.708 | **0.205\*\*** | **0.266\*\*** | 0.128 |  |  |  |
|  | 5 | 0.599 | **0.096\*** | 0.157 | 0.019 | -0.109 |  |  |
|  | 6 | 0.667 | **0.164\*** | **0.225\*** | 0.087 | -0.041 | 0.068 |  |
| Complex | 1 | 0.195 |  |  |  |  |  |  |
| Multimorbidity | 2 | 0.131 | -0.064 |  |  |  |  |  |
|  | 3 | 0.265 | **0.070\*** | **0.134\*\*** |  |  |  |  |
|  | 4 | 0.392 | **0.197\*\*** | **0.261\*\*** | 0.127 |  |  |  |
|  | 5 | 0.292 | **0.097\*** | **0.161\*\*** | 0.027 | -0.1 |  |  |
|  | 6 | 0.323 | **0.128\*** | **0.192\*** | 0.058 | **-0.069\*** | 0.031 |  |
| Respiratory | 1 | 0.104 |  |  |  |  |  |  |
| disorders | 2 | 0.115 | 0.011 |  |  |  |  |  |
|  | 3 | 0.126 | 0.022 | 0.011 |  |  |  |  |
|  | 4 | 0.164 | 0.06 | 0.049 | 0.038 |  |  |  |
|  | 5 | 0.242 | **0.138\*\*** | **0.127\*\*** | **0.116\*\*** | **0.078\*\*** |  |  |
|  | 6 | 0.13 | 0.026 | 0.015 | 0.004 | -0.034 | **-0.112\*\*** |  |
| Eye Disorders | 1 | 0.363 |  |  |  |  |  |  |
|  | 2 | 0.316 | -0.047 |  |  |  |  |  |
|  | 3 | 0.396 | 0.033 | 0.08 |  |  |  |  |
|  | 4 | 0.449 | 0.086 | 0.133 | 0.053 |  |  |  |
|  | 5 | 0.317 | -0.046 | 0.001 | -0.079 | -0.132 |  |  |
|  | 6 | 0.445 | 0.082 | 0.129 | 0.049 | -0.004 | 0.128 |  |
| Musculoskeletal | 1 | 0.415 |  |  |  |  |  |  |
| and Connective | 2 | 0.362 | -0.053 |  |  |  |  |  |
| tissue disorders | 3 | 0.468 | **0.053\*** | 0.106 |  |  |  |  |
|  | 4 | 0.553 | **0.138\*** | 0.191 | 0.085 |  |  |  |
|  | 5 | 0.507 | 0.092 | 0.145 | 0.039 | -0.046 |  |  |
|  | 6 | 0.51 | 0.095 | 0.148 | 0.042 | -0.043 | 0.003 |  |
| Neoplasms | 1 | 0.047 |  |  |  |  |  |  |
|  | 2 | 0.052 | 0.005 |  |  |  |  |  |
|  | 3 | 0.056 | 0.009 | 0.004 |  |  |  |  |
|  | 4 | 0.056 | 0.009 | 0.004 | 0.000 |  |  |  |
|  | 5 | 0.044 | -0.003 | -0.008 | -0.012 | -0.012 |  |  |
|  | 6 | 0.06 | 0.013 | 0.008 | 0.004 | 0.004 | 0.016 |  |
| Circulatory | 1 | 0.451 |  |  |  |  |  |  |
| Disorders | 2 | 0.416 | -0.035 |  |  |  |  |  |
|  | 3 | 0.528 | **0.077\*** | 0.112 |  |  |  |  |
|  | 4 | 0.616 | **0.165\*\*** | **0.200\*\*** | 0.088 |  |  |  |
|  | 5 | 0.542 | 0.091 | 0.126 | 0.014 | -0.074 |  |  |
|  | 6 | 0.556 | 0.105 | 0.14 | 0.028 | -0.06 | 0.014 |  |
| Endocrine, | 1 | 0.088 |  |  |  |  |  |  |
| nutritional and | 2 | 0.044 | **-0.044\*** |  |  |  |  |  |
| metabolic | 3 | 0.107 | 0.019 | **0.063\*** |  |  |  |  |
| disorders | 4 | 0.236 | **0.148\*\*** | **0.192\*\*** | **0.129\*\*** |  |  |  |
|  | 5 | 0.137 | 0.049 | **0.093\*\*** | 0.030 | **-0.099\*\*** |  |  |
|  | 6 | 0.193 | **0.105\*\*** | **0.149\*\*** | **0.086\*\*** | -0.043 | **0.056\*** |  |

*Note.* The estimates are the absolute differences in proportions of participants having the health outcomes in the class (in row) minus the class (in column). \*p<.05 and \*\*p<.01 indicate the p-values for pairwise Wald’s test for differences in proportion for the class (in row) minus the class (in column)

**References**

Banks, J., Phelps, A., Oskala, A., Steptoe, A., Blake, M., Oldfield, Z., Marmot, M., Clemens, S., Rogers, N., Nazroo, J. (2021). English Longitudinal Study of Ageing: Waves 0-9, 1998-2019. ( [data collection]. 36th Edition.UK Data Service. SN: 5050). <https://doi.org/http://doi.org/10.5255/UKDA-SN-5050-23>

Chen, Q., Luo, W., Palardy, G. J., Glaman, R., & McEnturff, A. (2017). The efficacy of common fit indices for enumerating classes in growth mixture models when nested data structure is ignored: A Monte Carlo study. *Sage Open*, *7*(1), 2158244017700459.

Dhalwani, N. N., O’Donovan, G., Zaccardi, F., Hamer, M., Yates, T., Davies, M., & Khunti, K. (2016). Long terms trends of multimorbidity and association with physical activity in older English population. *International Journal of Behavioral Nutrition and Physical Activity*, *13*(1), 1-9.

Elias, P., & McKnight, A. (2003). Earnings, unemployment and the NS-SEC. *A Researcher’s Guide to the National Statistics Socio-economic Classification. London: Sage*, 151-172.

Hackett, R. A., Moore, C., Steptoe, A., & Lassale, C. (2018). Health behaviour changes after type 2 diabetes diagnosis: Findings from the English Longitudinal Study of Ageing. *Scientific reports*, *8*(1), 1-8.

Hamer, M., de Oliveira, C., & Demakakos, P. (2014). Non-exercise physical activity and survival: English longitudinal study of ageing. *American journal of preventive medicine*, *47*(4), 452-460.

Harrison, C., Britt, H., Miller, G., & Henderson, J. (2014). Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice. *BMJ open*, *4*(7), e004694.

Health, D. o. (2016). Alcohol Guidelines Review – Report from the Guidelines Development Group to the UK Chief Medical Officer. . <https://asssets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/545739/GDG_reportJan2016.pdf>

Kojima, G., Iliffe, S., Jivraj, S., & Walters, K. (2020). Fruit and Vegetable Consumption and Incident Prefrailty and Frailty in Community-Dwelling Older People: The English Longitudinal Study of Ageing. *Nutrients*, *12*(12), 3882.

Muthén, B., & Muthén, L. (2017). Mplus. In *Handbook of item response theory* (pp. 507-518). Chapman and Hall/CRC.

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440.

Phillips, D., Lin, Y.-C., Wight, J., Chien, S., & Lee, J. (2014). Harmonized ELSA documentation. *Version C*.

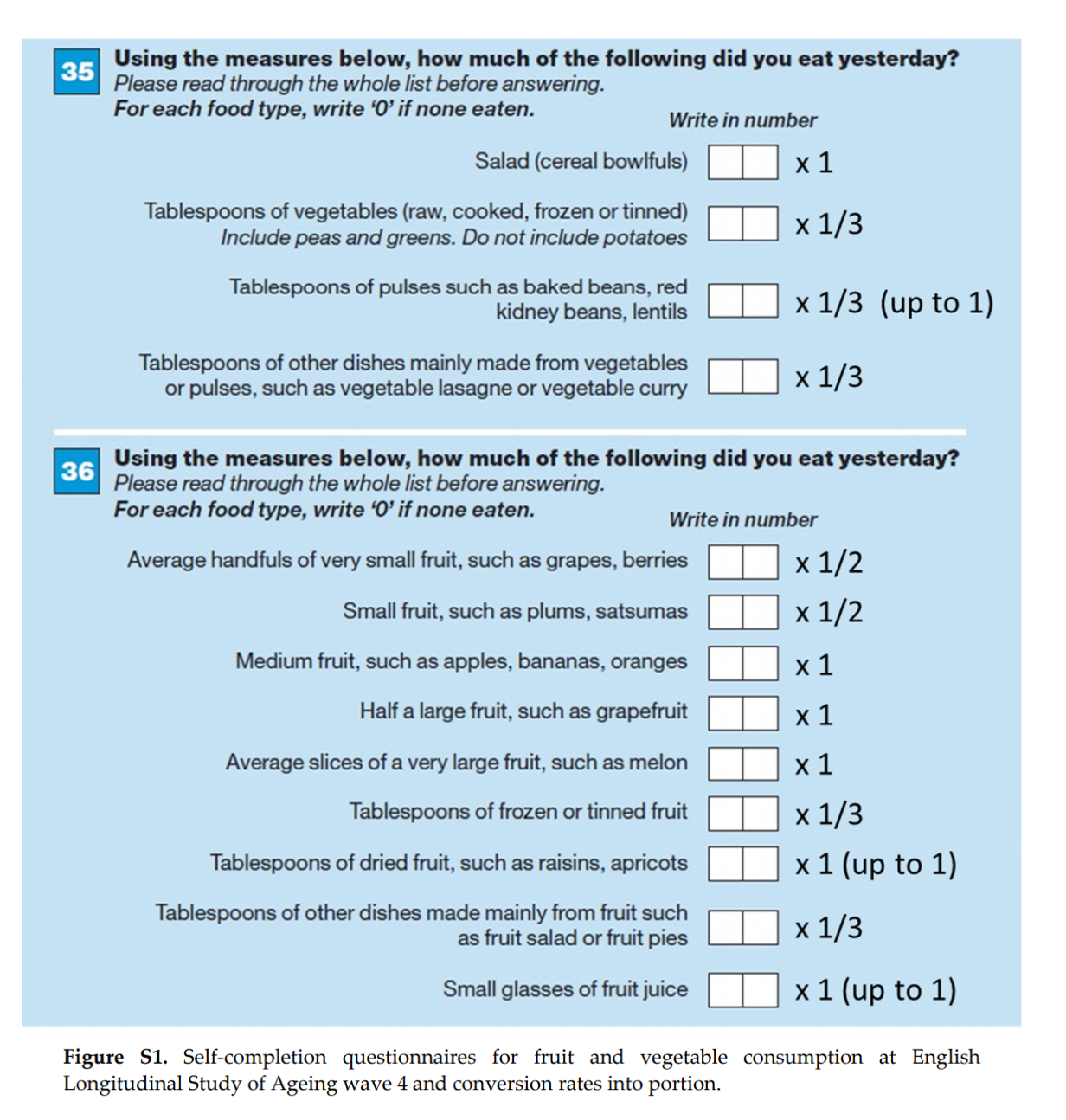
Schreiber, J. B. (2017). Latent class analysis: an example for reporting results. *Research in social and administrative pharmacy*, *13*(6), 1196-1201.

Singer, L., Green, M., Rowe, F., Ben-Shlomo, Y., Kulu, H., & Morrissey, K. (2019). Trends in multimorbidity, complex multimorbidity and multiple functional limitations in the ageing population of England, 2002–2015. *Journal of comorbidity*, *9*, 2235042X19872030.

Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort profile: the English longitudinal study of ageing. *International journal of epidemiology*, *42*(6), 1640-1648.

## Appendix A

## Appendix B

**Conversion rates for nutrition in Wave 4**

***Figure.*** “The amount of consumed fruit and vegetables were converted into portions (1 portion = 80 g for both vegetables and fruit) in accordance with the Welsh Health Survey methodology and the “5 A Day” campaign portion size from the National Health Service (NHS).” Kojima et. al (2020, p.3)

**Appendix C**

**Table 6.**

Pairwise health outcome Pairwise comparisons across latent classes for health outcomes (unadjusted for sociodemographic variables)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Health outcomes | Latent Classes | Class proportions | | 1  Overall health conscious | 2  Hazardous drinkers with a healthy lifestyle | 3  Inactive hazardous drinkers | 4  Risk avoidant and inactive with a healthy diet | 5  Smokers | 6  Risk avoidant and inactive with a healthy diet |
|  |  | |  | n= 20.9 % | n=11.1% | n=18.1% | n=22.2% | n=10.5% | n=17.2% |
| Multimorbidity | 1 | | 0.49 |  |  |  |  |  |  |
|  | 2 | | 0.424 | -0.066 |  |  |  |  |  |
|  | 3 | | 0.576 | **0.086\*\*** | **0.152\*\*** |  |  |  |  |
|  | 4 | | 0.723 | **0.233\*\*** | **0.299\*\*** | **0.147\*\*** |  |  |  |
|  | 5 | | 0.597 | **0.107\*\*** | **0.173\*\*** | 0.021 | **-0.126\*\*** |  |  |
|  | 6 | | 0.678 | **0.188\*\*** | **0.254\*\*** | **0.102\*\*** | -0.045 | **0.081\*** |  |
| Complex | 1 | | 0.184 |  |  |  |  |  |  |
| multimorbidity | 2 | | 0.121 | **-0.063\*** |  |  |  |  |  |
|  | 3 | | 0.258 | **0.074\*\*** | **0.137\*\*** |  |  |  |  |
|  | 4 | | 0.405 | **0.221\*\*** | **0.284\*\*** | **0.147\*\*** |  |  |  |
|  | 5 | | 0.291 | **0.107\*\*** | **0.170\*\*** | 0.033 | **-0.114\*\*** |  |  |
|  | 6 | | 0.334 | 0.150 | **0.213\*\*** | **0.076\*** | -0.071 | 0.043 |  |
| Respiratory | 1 | | 0.102 |  |  |  |  |  |  |
| disorders | 2 | | 0.112 | 0.010 |  |  |  |  |  |
|  | 3 | | 0.125 | 0.023 | 0.013 |  |  |  |  |
|  | 4 | | 0.166 | **0.064\*\*** | **0.054\*** | 0.041 |  |  |  |
|  | 5 | | 0.242 | **0.140\*\*** | **0.130\*\*** | **0.117\*\*** | **0.076\*** |  |  |
|  | 6 | | 0.132 | 0.030 | 0.020 | 0.007 | -0.034 | **-0.110\*\*** |  |
| Eye Disorders | 1 | | 0.408 |  |  |  |  |  |  |
|  | 2 | | 0.348 | -0.06 |  |  |  |  |  |
|  | 3 | | 0.457 | 0.049 | **0.109\*** |  |  |  |  |
|  | 4 | | 0.569 | **0.161\*\*** | **0.221\*\*** | **0.112\*** |  |  |  |
|  | 5 | | 0.507 | 0.099 | 0.159 | **0.050\*** | **-0.062\*\*** |  |  |
|  | 6 | | 0.518 | **0.110\*\*** | **0.17\*\*** | 0.061 | -0.051 | **0.011\*\*** |  |
| Musculoskeletal | 1 | | 0.392 |  |  |  |  |  |  |
| and Connective | 2 | | 0.32 | -0.072 |  |  |  |  |  |
| tissue disorders | 3 | | 0.405 | 0.013 | **0.085\*\*** |  |  |  |  |
|  | 4 | | 0.444 | **0.052\*\*** | **0.124\*\*** | **0.039\*\*** |  |  |  |
|  | 5 | | 0.496 | **0.104\*\*** | **0.176\*\*** | 0.091 | 0.052 |  |  |
|  | 6 | | 0.518 | **0.126\*\*** | **0.198\*\*** | 0.113 | 0.074 | 0.022 |  |
| Neoplasms | 1 | | 0.045 |  |  |  |  |  |  |
|  | 2 | | 0.05 | 0.005 |  |  |  |  |  |
|  | 3 | | 0.056 | 0.011 | 0.006 |  |  |  |  |
|  | 4 | | 0.057 | 0.012 | 0.007 | 0.001 |  |  |  |
|  | 5 | | 0.044 | -0.001 | -0.006 | -0.012 | -0.013 |  |  |
|  | 6 | | 0.063 | 0.018 | 0.013 | 0.007 | 0.006 | 0.019 |  |
| Circulatory | 1 | | 0.442 |  |  |  |  |  |  |
| Disorders | 2 | | 0.407 | -0.035 |  |  |  |  |  |
|  | 3 | | 0.526 | **0.084\*** | **0.119\*\*** |  |  |  |  |
|  | 4 | | 0.624 | **0.182\*\*** | **0.217\*\*** | **0.098\*\*** |  |  |  |
|  | 5 | | 0.541 | **0.099\*\*** | **0.134\*\*** | 0.015 | **-0.083\*** |  |  |
|  | 6 | | 0.566 | **0.124\*\*** | **0.159\*\*** | 0.04 | -0.058 | 0.025 |  |
| Endocrine, | 1 | | 0.086 |  |  |  |  |  |  |
| nutritional and | 2 | | 0.044 | **-0.042\*** |  |  |  |  |  |
| metabolic | 3 | | 0.108 | 0.022 | **0.064\*\*** |  |  |  |  |
| disorders | 4 | | 0.234 | **0.148\*\*** | **0.19\*\*** | **0.126\*\*** |  |  |  |
|  | 5 | | 0.137 | **0.051\*** | **0.093\*\*** | 0.029 | **-0.097\*\*** |  |  |
|  | 6 | | 0.193 | **0.107\*\*** | **0.149\*\*** | **0.085\*\*** | -0.041 | 0.056 |  |

*Note.* The estimates are the absolute differences in proportions of participants having the health outcomes in the class (in row) minus the class (in column). \*p<.05 and \*\*p<.01 indicate the p-values for pairwise Wald’s test for differences in proportion for the class (in row) minus the class (in column)