MODEM

modelling outcome and cost impacts
of interventions for dementia

# Population Ageing and Care Simulation model (PACSim)

## BASELINE DATASET AND MODEL CONSTRUCTION

## (VERSION: 241017)

Andrew Kingston, Research statistician
Carol Jagger, AXA Professor of Epidemiology of Ageing

Institute of Health & Society and Newcastle University Institute for Ageing

www.modem-dementia.org.uk

Improving Dementia Care
ESRC·NIHR

# Contents

## Introduction

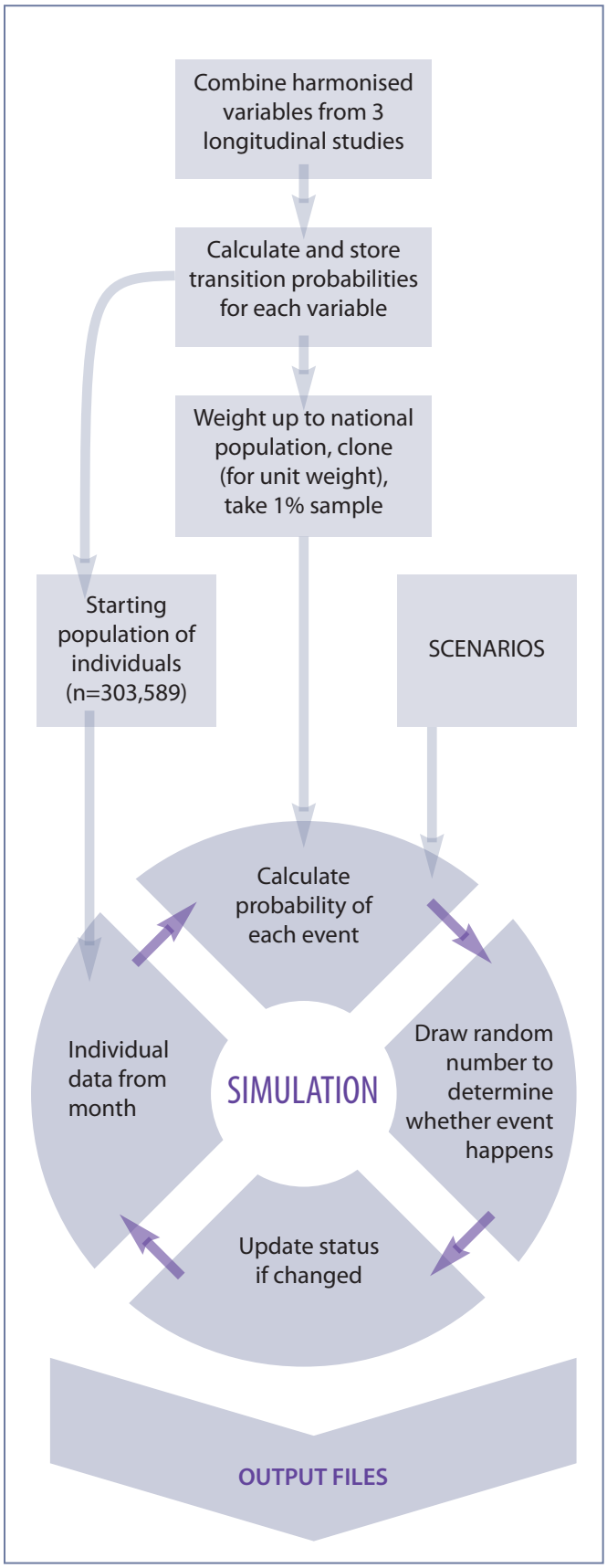
This document describes the construction of the base population and the overall design of the Population Ageing and Care Simulation model (PACSim). PACSim model is an epidemiological microsimulation model of population health and care developed as part of a wider project MODEM (Modelling Outcome and cost interventions for DEMentia) (see www.modem-dementia.org.uk). The aims of PACSim are to produce projections of the health and associated care needs of the English population aged 65 years and over to 2040 and to assess the impact of interventions for risk factor reduction, disease prevention and treatments to slow down progression to disease and disability.

PACSim builds upon (i) the previous epidemiological macrosimulation model SIMPOP (Jagger *et al*. 2009), based on the MRC Cognitive Function and Ageing Study (MRC CFAS) and linking multiple diseases with disability to project future disability burden (and disability-free life expectancy) in people aged 65 years and over; and (ii) DYNOPTAsim, an Australian dynamic microsimulation model (Booth *et al*. 2011), which incorporates multiple longitudinal studies of ageing to model the detailed health and functional status of the Australian population aged 45 years and over as well as the impact of potential interventions for risk reduction. The overall design of PACSim is shown in Figure 1.

## Base population

To enable estimates for the population aged 65+ to be simulated 2040, the base population for PACSim is individuals aged 35 years and over. Three longitudinal studies were used to construct the base population and one follow-up wave (generally around two years later) for calculation of the transition probabilities. The first study, *Understanding Society (US)*, is a longitudinal study of the social and economic circumstances, attitudes, behaviours and health of people aged 16 and over in the UK. We used wave 1 (2009–10) and wave 2 (2010–11). The second, the *English Longitudinal Study of Ageing (ELSA)*, supplied information on the health, social, wellbeing and

**Figure 1: Schematic diagram of the Population Ageing and Care Simulation (PACSim) model**



economic circumstances of the English population aged 50 and older. Currently, there are six waves of data collection covering a period of eleven years from 2002/3. We used wave 5 (2010–11) for the baseline with wave 6 for the transitions. The third dataset is the *Cognitive Function and Ageing Study II (CFAS II)* with the objective of investigating dementia and cognitive decline in a representative sample of more than 18,000 people aged 65 years and over. The first wave of data collection was 2008–11 and there has been one follow-up wave. The second wave of data collection began in 2011 and finished in 2013.

Baseline characteristics of the base population included in PACSim are of four types: socio-demographic; health behaviours; diseases and geriatric conditions; and dependency using a time-based measure that incorporates the intensity of care required (Table 1). All diseases were self-report of a doctor ever diagnosing the disease, apart from depression which was current episode but doctor-diagnosed.

**Table 1: Variables included in CFAS, ELSA and Understanding Society**

| Variable | Categories |
|---|---|
| Sociodemographic | |
| Age | Single years |
| Sex | Male/Female |
| Education (years of) | 0–9/10–11/12+ |
| Marital status | Never married/Married/Widowed/Divorced or separated |
| Occupation (NSSEC3) | Higher managerial, administrative and professional occupations/ Intermediate occupations/ Routine and manual occupations |
| Retirement | Retired/Not retired |
| Lifestyle behaviours | |
| Smoking | Never/Former/Current |
| Physical activity | High/medium/low |
| BMI | Underweight/Normal/Overweight/Obese |
| Disease/impairment | |
| Cognitive Impairment (MMSE) | MMSE 0-9/10-20/21-26//27-30 |
| CHD | Ever/Never |
| Stroke | Ever/Never |
| Hypertension | Ever/Never |
| Diabetes | Ever/Never |
| Arthritis | Ever/Never |
| Incontinence | Severe/Not severe |
| Depression (diagnosed) | Ever/Never |
| Respiratory disease (asthma/bronchitis) | Ever/Never |
| Cancer | Ever/Never |
| Visual impairment | Current yes/No |
| Hearing impairment | Current yes/No |
| Outcome | |
| IADLs/ADLs for interval need/dependency | Independent/low/medium/high dependency |

Full detail on the harmonisation of variables across the studies is given in the Appendix. Cognitive impairment using the Mini-Mental State Examination (MMSE) score was available only in CFAS. However there were other cognitive items comparable between ELSA and CFAS and so the MMSE categories (0–9, 10–20, 21–26, 27–30) for individuals in ELSA are imputed using the chained equation method3 and based on the comparable cognitive items, age, sex and education. Individuals below the age of 50 at baseline are assumed to have normal cognitive function (MMSE 27–0). BMI was not available in CFAS and was similarly imputed based on age, sex and education.

Dependency was measured by the interval of need scale (Isaacs and Neville 2015), categorising people on the basis of ADLs/IADLs on the intensity of care required (24-hour care, regular times daily, less than daily, none). Generally, the interval need measure assigns those with severe cognitive impairment to high dependency (24-hour care), but PACSim reassigns such individuals at the end of the simulation. Further detail on the calculation of the interval need variable is available in Appendix 1.

Dementia status was assigned post simulation probabilistically using the probability of dementia conditional on age group (65–74, 75–84, 85+), care home/community residence, and MMSE category.

## Transition probabilities

All three datasets had two years between the longitudinal waves for estimation of transition probabilities for each characteristic (apart from education and social class which were fixed). For the remaining stochastic characteristics (diseases, health behaviours, disability) transition probabilities are estimated from logistic regression modelling of the longitudinal data from relevant datasets and appropriate predictors. Where missing values are imputed for a whole study (e.g. BMI in CFAS, cognitive impairment in Understanding Society – see later section) we exclude this study from estimation of the transition probabilities.

Most diseases were considered chronic and therefore only incidence (transition to disease) was estimated; the exception was depression where the probability of recovery was also estimated. Similarly, recovery for vision and hearing impairment were allowed, as well as recovery from mild cognitive impairment (MMSE 21–26) to normal cognition (MMSE 27–30). For physical activity, there was insufficient data to model transitions, therefore the transition probabilities were approximated by fitting models to the baseline data only. Transition probabilities for BMI used ELSA waves 4 and 6 as BMI was not collected in wave 5. As BMI (and disability and cognitive function) are ordered categorical data an ordinal logistic model was fitted, or where the proportional odds assumption was violated, a generalised logit model.

The modelled transition relationships between variables are shown in Table 2 with √ denoting a significant association. If variables (apart from age and sex) were not significant in the model, the coefficient was replaced by zero (and shown in Table 2 by X).

## Weighting to 2014 England population, cloning and sampling

Each study had study-specific weights which accounted for sampling and response. These were recalibrated using inverse probability weighting, to have the same age (in single years) and sex distribution to the 2014 England population which forms the basis of the 2014 population projections (ONS 2015). The final age group was closed at 91+ since there were few individuals above age 90 and, in ELSA, exact ages above age 90 were not available.

Weighting up to the 2014 England population resulted in minimum weights of 9. Since large weights can produce uneven characteristics and therefore less smooth population-level dynamics, we cloned the population to produce multiple individuals with weight 1. This resulted in a very large dataset of over 30m individuals so, to ease computing power and time, a 1% random sample

**Table 2: Modelled transition relationships between variables**

| | Predictors | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Age | Sex | Education | Marital status | Occupation | Smoking | BMI | Physical activity | CHD | Stroke | Hypertension | Diabetes | Arthritis | Depression | Respiratory | Cancer | Vision | Hearing | Cognition | Dependency |
| Marital status | √ | √ | | | | | | | | | | | | | | | | | | |
| Smoking | √ | √ | X | X | | | | | | | | | | | | | | | | |
| BMI | √ | √ | √ | | | | | | | | | | | | | | | | | |
| Physical activity | √ | √ | √ | | | | √ | | | | | | | | | | | | | |
| CHD | √ | √ | | | | X | √ | √ | | | X | X | | | | | | | | |
| Stroke | √ | √ | | | | √ | √ | | | | √ | | | | | | | | | |
| Hypertension | √ | √ | √ | | | X | √ | X | | | | | | | | | | | | |
| Diabetes | √ | √ | √ | | | X | √ | √ | | | √ | | | | | | | | | |
| Arthritis | √ | √ | | | | | √ | | | | | | | | | | | | | |
| Depression | √ | √ | | √ | | | | √ | | | | | | | | | | | | |
| Respiratory disease | √ | √ | X | | X | √ | √ | | | | | | | | | | | | | |
| Cancer | √ | √ | | | | X | X | | | | | | | | | | | | | |
| Vision | √ | √ | √ | | | √ | | | | | | | | | | | | | | |
| Hearing | √ | √ | √ | | | √ | | | | | | | | | | | | | | |
| Cognition | √ | √ | √ | | | √ | √ | √ | | √ | √ | √ | | √ | | | √ | √ | | |
| Dependency | √ | √ | √ | | | | | | √ | √ | √ | √ | √ | √ | √ | √ | √ | X | √ | |
| Retirement | √ | √ | X | | X | | | | √ | | √ | | √ | √ | X | | | | | |
| Institution | √ | √ | √ | √ | | | | | | | | | | | | | | | √ | √ |

√ = significant predictor, X = entered into transition model but not significant

was selected as the base population for the simulation. The final base population comprised 303,588 individuals.

## Imputation of missing values

As explained earlier, BMI (not collected in CFAS) and disability and cognitive function which were less comparable across datasets were imputed in the base population stage prior to weighting and cloning. Missing values for remaining characteristics were imputed after the 1% sample had been selected using the chained equations method (Royston and White 2011) and all available variables. A single imputation was used as the dataset was large.

## Simulation

PACSim adopts a discrete time approach and uses monthly transition probabilities for each stochastic characteristic. A monthly transition step minimises biases arising from competing events (Van Imhoff and Post 1998) better accounts for non-linearity over a longer period, but also allows events to happen more closely and multiply within a year (for example, stroke and cognitive impairment) which is likely to occur at older ages. Migration and emigration are not included and the only exit is death. For variables with multiple ordered categories it is assumed that transitions can only take place to the next category above or below.

If $p$ is the probability of a transition between the two waves and $m$ is the mean interval in months between the waves then the monthly probability of a transition was estimated as .

$$p^* = 1 - (1-p)^{1/m}$$

At each simulation step all of the model variables are updated by calculating the relevant monthly transition probability and comparing it to a randomly generated uniformly distributed variable. The order of simulation of the model variables is given in Table 3.

For transitions to death we use assumed future mortality underlying the most recent (2014–based) UK Principal Population Projection

**Table 3: Order of full simulation**

| | |
|---|---|
| 1. | Survival |
| 2. | Marital status |
| 3. | Retirement |
| 4. | Smoking |
| 5. | Institutionalisation |
| 6. | Physical activity |
| 7. | BMI |
| 8. | Hypertension |
| 9. | Stroke |
| 10. | Diabetes |
| 11. | Cognitive impairment |
| 12. | CHD |
| 13. | Arthritis |
| 14. | Depression |
| 15. | Respiratory disease |
| 16. | Cancer |
| 17. | Visual impairment |
| 18. | Hearing impairment |
| 19. | Disability/care needs |
| 20. | Age |

(ONS 2015). These are in the form of annual probabilities of dying between exact ages x and x+1, denoted $q_x$. The Office for National Statistics (ONS) publishes $q_x$ for 1981–2062 for males and females by single years of age (x = 0 to 100); $q_{95}$ was used for the final age group 91+ years, an approximation that produced negligible differences in life expectancies. These annual probabilities were converted to monthly probabilities in the same way as the estimated transitions between states for characteristics. At present, any differential effects of disease on mortality have not been included though later versions of PACSim will include a differential effect for dementia/cognitive impairment. Life expectancies for each year in the simulation were calculated using standard abridged life table techniques with observed deaths in each year and mid-year population.

## Outputs

Outputs from the model for each calendar year are (i) numbers with and prevalence of characteristics by age group and sex; (ii) numbers with and prevalence of multi-morbidity; (iii) years lived with individual diseases, multi-morbidity and dependency calculated by Sullivan's method (Sullivan 1971).

## Validation

The number of individuals by five-year age group (65–69, 70–74, 75–79, 80–84, 85–89, 90+) and year (2014-2041) from PACSim and the ONS 2014 population projections are shown in Figure 2. The largest percentage difference was in the oldest age groups (90+) which reached over 2.5% between 2022 and 2025; for ages 65–84 the percentage difference in numbers of individuals between PACSim and ONS was under 1% and mostly below 0.5% throughout. The ONS Principal Projection includes assumptions about changing migration over time which are not included in PACSim, though migration will have a greater impact on the younger rather than the older population.

We compared the age and sex specific prevalence of stroke and diabetes and current smoking and overweight and obesity for the non-institutionalised population of PACSim in 2014 with the Health Survey for England 2014 (Table 4) (HSCIC 2015). Generally, there was good agreement apart from the prevalence of obesity where PACSim prevalences were lower by around 8 percentage points for men aged 35–64 and for women of all ages.

**Figure 2: Comparison of number of individuals by age group and year from PACSim with ONS 2014–based population projections**

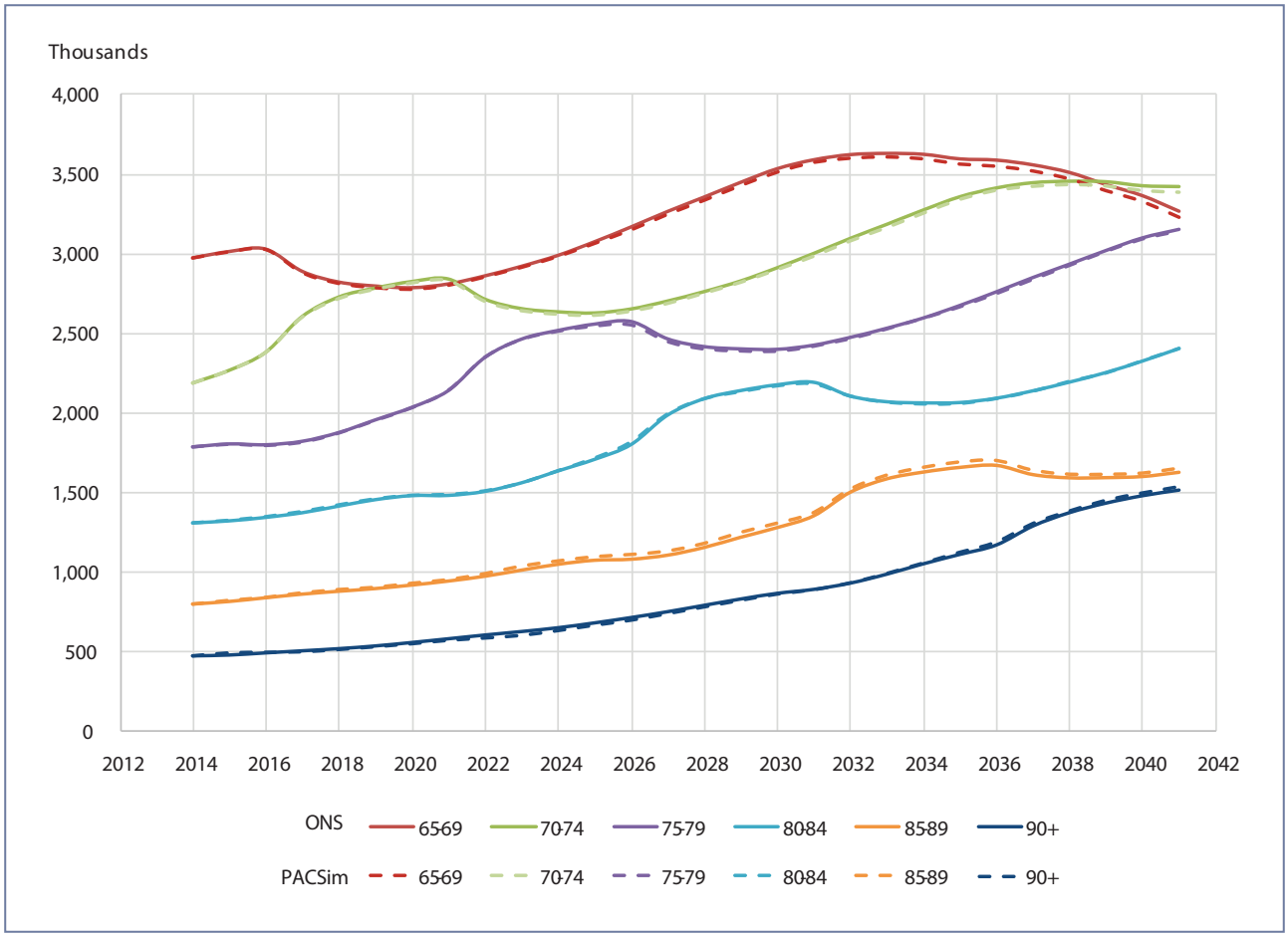


**Table 4: Comparison of disease and risk factor prevalences from PACSim and from Health Survey for England 2014, by age group and sex**

| | HSE 2014 | | | PACSim (non-institutionalised) | | |
|---|---|---|---|---|---|---|
| | Men (%) | Women (%) | All (%) | Men (%) | Women (%) | All (%) |
| Stroke | | | | | | |
| 35–44 | 0.5 | 0.1 | 0.3 | 0.4 | 0.5 | 0.5 |
| 45–54 | 1.5 | 1.3 | 1.4 | 1.0 | 1.0 | 1.0 |
| 55–64 | 4.1 | 2.4 | 3.2 | 2.7 | 1.8 | 2.2 |
| 65–74 | 7.2 | 4.4 | 5.7 | 6.0 | 4.2 | 5.1 |
| 75+ | 12.2 | 9.4 | 10.6 | 10.8 | 8.9 | 9.7 |
| Diabetes | | | | | | |
| 35–44 | 2.9 | 2.8 | 2.8 | 2.1 | 2.4 | 2.3 |
| 45–54 | 8.8 | 4.6 | 6.7 | 5.9 | 4.2 | 5.1 |
| 55–64 | 12.7 | 7.2 | 9.9 | 11.2 | 6.9 | 9.0 |
| 65–74 | 15.1 | 11.1 | 13.0 | 16.7 | 10.0 | 13.2 |
| 75+ | 17.2 | 13.8 | 15.3 | 16.7 | 13.3 | 14.8 |
| Smoking | | | | | | |
| Current | | | | | | |
| 35–44 | 25.0 | 20.4 | 22.7 | 25.8 | 23.1 | 24.5 |
| 45–54 | 21.0 | 16.2 | 18.5 | 24.1 | 21.6 | 22.8 |
| 55–64 | 17.3 | 19.4 | 18.4 | 19.5 | 18.9 | 19.2 |
| 65–74 | 13.5 | 11.7 | 12.5 | 13.8 | 12.8 | 13.3 |
| 75+ | 5.1 | 5.5 | 5.3 | 8.0 | 7.1 | 7.5 |
| BMI | | | | | | |
| Overweight | | | | | | |
| 35–44 | 43.0 | 33.8 | 38.6 | 46.0 | 28.6 | 37.2 |
| 45–54 | 43.6 | 31.6 | 40.7 | 45.8 | 31.3 | 38.4 |
| 55–64 | 44.6 | 34.7 | 39.1 | 45.1 | 34.8 | 39.8 |
| 65–74 | 52.2 | 35.2 | 39.6 | 44.5 | 39.1 | 41.7 |
| 75+ | 55.5 | 39.3 | 45.3 | 44.1 | 36.3 | 39.6 |
| Obese | | | | | | |
| 35–44 | 28.0 | 28.5 | 28.7 | 20.8 | 20.4 | 20.6 |
| 45–54 | 33.8 | 35.9 | 34.2 | 25.6 | 22.8 | 24.2 |
| 55–64 | 37.9 | 39.6 | 40.3 | 26.9 | 26.5 | 26.7 |
| 65–74 | 26.6 | 33.5 | 34.9 | 26.4 | 25.8 | 26.1 |
| 75+ | 25.9 | 30.5 | 27.2 | 23.4 | 24.2 | 23.9 |

## References

Brown L, Nepal B, Booth H, Pennec S, Anstey K, Harding A (2011) *Dynamic Modelling of Ageing and Health: The Dynopta Microsimulation Model,* NATSEM Working Paper 11/14, University of Canberra, Canberra.

Health and Social Care Information Centre (2015) Health Survey for England 2014 Adult Trend Tables, Health and Social Care Information Centre, London.

Isaacs B, Neville Y (1976) The needs of old people. The 'interval' as a method of measurement, *British Journal of Preventive & Social Medicine*, 30, 2, 79–85.

Jagger C, Matthews R, Lindesay J, Robinson T, Croft P, Brayne C (2009) The effect of dementia trends and treatments on longevity and disability, *Age and Ageing,* 38: 319–325.

Office for National Statistics (2015) 2014-based National Population Projections, 20 October 2015 ed, Office for National Statistics.

Royston P, White IR (2011) Multiple imputation by chained equations (MICE): implementation in Stata, *Journal of Statistical Software*, 45, 4, 1–20.

Stochl J, Jones PB, Croudace TJ (2012) Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers, *BMC Medical Research Methodology*, 12, 1, 74.

Sullivan DF (1971) A single index of mortality and morbidity, *Health Services Mental Health Administration Health Reports*, 86, 347–354.

Van Imhoff E, Post W (1998) Microsimulation methods for population projection, *Population: An English Selection*, 10, 1, 97–138.

## Acknowledgements

Understanding Society and ELSA data were made available through the UK Data Archive. Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. ELSA was developed by a team of researchers based at the NatCen Social Research, University College London and the Institute for Fiscal Studies. The data were collected by NatCen Social Research. The funding is provided by the National Institute of Aging in the United States, and a consortium of UK government departments co-ordinated by the Office for National Statistics. CFAS data was requested through the CFAS website (www.cfas.ac.uk). CFAS is a collaboration between the Universities of Cambridge, Newcastle, Nottingham, Oxford, Gwynedd and Liverpool, funded by major awards from the Medical Research Council, the Economic and Social Research Council and Alzheimer's Society.

The developers and funders of Understanding Society, ELSA, CFAS and the Archive do not bear any responsibility for the analyses or interpretations presented here.

# APPENDIX: Variable construction

## Education

Education is measured by the number of years of formal education and then categorised as 0–9 years, 10–11 years and 12+ years. This is a fixed variable created from the baseline waves of the respective studies. For CFAS this is wave one CFAS II (from the derived variables file). For ELSA, this is taken from wave 5 and is determined from the age that education ended (fed forward from previous waves – question: fffqend – wave 5 dataset). For US this is taken from wave one and is determined by the school leaving age (question: a_scend).

## Marital Status

Marital status was taken from wave 1 of the CFAS II study (question: married_w1) and for wave two (question: v5_w2). From ELSA wave 5 the question was dimar and for wave 6 marstat. The US question in wave 1 was a_pmarstat and wave 2 b_marstat.

## Occupation (NSSEC-3)

The National Statistics Socio-economic classification (NSSEC) was assumed to be fixed over time. For CFAS participants this was taken from the derived variables dataset (ex2_cv3). For ELSA, this was provided as part of the dataset from the UK Data Service in their derived variables file (wave 5: w5nssec3; wave 6: NSSEC). Wave 1 and 2 of US included NSSEC as part of their dataset (questions: a_jbnssec3_dv and b_jbnssec3_dv respectively).

## Institutionalisation

Institutional status was taken from CFAS waves one and two (question: v9_w1 and v1-W2 respectively). Participants recruited to ELSA did not include those in institutions, however, they did follow people who later moved to an institution, and therefore institutional status was taken from question: inst of the derived variables file for both wave 5 and 6. Data on institutionalisation for US were not available as they did not include those in institutions, therefore, all people were deemed not to be institutionalised in the base population.

## Retirement

Retirement status for CFAS was taken from the variables v43_w1 and v40_w2 respectively (question: have you retired?). For ELSA, the questions: wpdes and WpDes respectively (question: what best describes your current employment situation) were used. For US the questions a_jbstat and b_jbstat (wave 1 and 2 respectively) regarding current economic activity had a potential response of 'retired'.

## Smoking Status

Smoking status is categorised as never smokers, current smokers and ex-smokers. Data from CFAS was taken from the derived variables file (smoking_w1) and categorised accordingly. Status from wave 2 was taken from question v474_w2. This question was of binary format and so smokers who had stopped since the previous wave were fed forward from the wave 1 data. Smoking status from ELSA was based on two questions:
(i) whether a participant has ever smoked and
(ii) whether a participant smokes at all nowadays.
The cross-tabulation of this question allowed the construction of the smoking variable in the desired format at both waves. Smoking status was not available for US at wave 1 but was at wave 2. Considering this, the status that was determined at wave 2 was assumed to hold at wave 1.

## Physical activity

ELSA has 3 variables (HEACTA, HEACTB, HEACTC) which indicate the frequency of vigorous, moderately energetic and mildly energetic activities respectively. The set of activities for each question is shown on Card C18. The coding for each variable is:

1 = more than once a week

2=once a week

3=one to three times a week

4=hardly ever, or never.

CFAS has approximately the same questions (Q375 – Q395) but the individual activities are asked about. The response codes are the same as ELSA. The same variables can therefore be calculated as follows:

First make sure missing values for Q375 – Q395 are set to missing.

If Q375=0 then HEACTA = 4

else if Q375=1 then HEACTA = min (Q376 – Q381) NB do not use Q382 as this is not included in the set of ELSA activities.

If Q383=0 then HEACTB=4

else if Q383=1 then HEACTB = min (Q384 – Q389) NB do not use Q390 as this is not included in the set of ELSA activities.

If Q391=0 then HEACTC=4

else if Q391=1 then HEACTC = min (Q393 – Q395) NB do not use Q392 as this is not included in the set of ELSA activities.

For ELSA and CFAS the physical activity score is then calculated as:

3*(4-HEACTA)+2*(4-HEACTB)+(4-HEACTC)

This score ranges from 0 to 18 and is categorised into: low physical activity (score 0–1), moderate physical activity (score 2–6), and high physical activity (score 7–18). We assume that if a respondent does not do one of the three sets of activities then the relevant variable is set at 0.

Assessing level of physical activity in US was not possible at baseline as the data were not routinely collected. Wave 2 physical activity was based on the frequency of taking part in vigorous sporting activity, moderate sporting activity, and mild sporting activity (b_sportsact). It was assumed that activity levels at wave 2 were the same at wave 1. We could not find levels of sporting activity that were consistent with levels determined through ELSA or CFAS. We therefore chose not to include physical activity from US, but imputed values based on age, sex and education.

## Body Mass Index

Body mass index was available for ELSA at wave 4 and wave 6. It was assumed that wave 4 values would endure at wave 5. For US, BMI was only available at wave 1. As data were not available for CFAS, these were imputed using multiple imputation with chained equations based on age, sex and education. This imputation was also used to populate missing values from both US and ELSA.

## Diseases and Impairments

Disease status was ascertained for the following conditions: coronary heart disease, stroke, hypertension, diabetes, arthritis, depression, respiratory disease and cancer. Diseases were taken from self-reported doctor diagnoses. Vision, hearing and cognitive impairment were generally taken from questions about difficulty with these conditions. Data on disease and impairments was available at both waves of respective studies.

## Coronary Heart Disease

For CFAS, participants were asked whether they had ever suffered an MI or angina. Affirmative responses to either of these categories was taken as presence of coronary heart disease. For ELSA previous diagnoses of MI or angina were fed forward from wave 1–4 and combined with newly reported cases at wave 5 and 6. In US the question asked about diagnoses of MI or angina at both waves 1 and 2.

## Stroke

Participants in all studies were asked whether they had ever suffered a stroke.

## Hypertension

For CFAS, hypertension was determined through self-report from the question, "Firstly have you ever been diagnosed /told you have any of the following health conditions", which included hypertension. For ELSA previous wave self-reported diagnoses were fed forward and added to subsequent diagnoses for waves 5 and follow-up wave 6. For US, the diagnosed health conditions section was used to determine hypertension cases at wave 1 then additional diagnoses added at wave 2.

## Diabetes

For CFAS, participants were asked whether they suffer type II diabetes in both waves 1 and 2. For ELSA previous diagnoses of diabetes were fed

forward from wave 1–4 and combined with newly reported cases at waves 5 and 6. In US the question asked about diagnoses of diabetes at both waves 1 and 2.

### Arthritis

For CFAS, participants were asked whether they had ever suffered arthritis. For ELSA previous diagnoses of asthma were fed forward from wave 1–4 and combined with newly reported cases at waves 5 and 6. For US diagnoses of arthritis were reported at waves 1 and 2.

### Depression

Depression in CFAS was determined through the question v104_w1 which details information on self-reported depressive symptomology. The question was: Is the depression/crying/feeling like crying there most of the time? If a participant responded in the affirmative then they were deemed depressed. Depression from participants in ELSA was determined through self-report after a doctor had informed them of a diagnosis. Had they not been informed of a diagnosis, they were deemed not depressed. US depression was determined by a question pertaining to clinical depression with a yes/no response. Like ELSA, this was deemed to be depression only if a doctor had confirmed a diagnosis.

### Respiratory Disease

For CFAS and US, respiratory disease was a composite variable made up of responses to questions on bronchitis and asthma. Participants were asked whether they suffer from bronchitis or asthma and affirmative responses to either defined respiratory disease. The exception to this was ELSA. ELSA did not have additional diagnoses to wave 5 and so wave 5 diagnoses were assumed to endure at wave 6.

### Cancer

For all studies participants were asked whether they suffer from cancer, excluding non-melanoma skin cancer.

### Visual Impairment

For CFAS, visual impairment was interviewer observed from the question: Did the subject have poor/no eyesight that interfered with reading, writing or drawing? Participants were deemed visually impaired at any level of impairment. Determination of visual problems from ELSA was by questions pertaining to whether a person reported that their ability to recognise a friend across the street, or read a newspaper was poor. Visual problems in US were determined through a question pertaining to whether person has difficulty with vision beyond the use of standard spectacles.

### Hearing Impairment

For CFAS, hearing impairment was interviewer observed from the question: Did the subject have hearing problems that interfered with the questioning? Participants were deemed hearing impaired at any level of impairment. Hearing problems in ELSA participants were determined via those who reported difficulty with hearing in in the presence of hearing aid. The same was true of those participants from US.

### Cognitive Impairment

To incorporate severe cognitive impairment into the interval of need variables (see below) and to test the effect of interventions, cognitive impairment had to be based on the Mini-Mental State Examination (MMSE), a score from 0–30. We used the NICE categories of 0–9 (severe cognitive impairment), 10–20 (moderate cognitive impairment), 21–26 (mild cognitive impairment, and 27–30 (no impairment). MMSE was only available in CFAS (waves 1 and 2).

ELSA used a cognitive battery to examine performance and some of these were also available in CFAS: orientation in time, and instant and delayed word recall, as well as IADL questions that examined cognitive ability: management of medications and the ability to make telephone calls. MMSE category was therefore derived for ELSA from these items in addition to age, sex and education, by multiple imputation using chained equations. US had no cognitive items and therefore MMSE category was imputed based on age, sex and education for participants aged 50 and over. MMSE scores for people under age 50 were assumed normal (27–30).

## Dependency/Interval of Need

The main outcome measure for PACSim was dependency which was measured by interval of need classification (IoN)4) which uses basic and instrumental activities of daily living (ADLs, cognitive impairment and incontinence to categorise individuals into the intensity of care provided: high dependency (needs 24-hour care), medium dependency (needs help at regular times daily), low dependency (needs help less than daily), independent (remainder). IoN had previously been calculated for CFAS I but the ADL items in CFAS and ELSA were not identical in two key areas.

First, in CFAS requiring help could be directly determined from the ADL item whilst for ELSA the response for CFAS related to needing help whereas for ELSA there was an initial screen asking about difficulty experienced with tasks and then if the participant had difficulty or did not do they were asked whether anyone helped them.

Second, and more importantly, the same ADL items were not available in ELSA as CFAS (e.g. heavy and light housework). To overcome this, we investigated all ADL items in each study and, through Mokken Scaling (Stochl *et al*. 2012), we ascertained the hierarchy of loss in ability (requiring help). We then identified similar items across both studies, in terms of their location in the hierarchy and prevalence, and then used these, with cognitive impairment and incontinence, to create the study specific IoN categories. As we required cognitive impairment to be simulated separately, individuals who had a MMSE score 0–9 were assigned to the high dependency category at each step of the simulation.

The items (and questions) used from each study are detailed below. US was dealt with in a different way (outlined below).

### CFAS:

High dependency was defined as: needs help for toilet (v550_w1) or chairfast or bedfast permanently (v559_w1) (unable to get out of a chair/bed without help), or incontinence (v551_w1) and needs help with shoes and socks (v542_w1) (proxy for dressing), or from HAS (proxy) interview needs help feeding (v56_h1), or often incontinent and needs help dressing (v53_h1 & v551_w1). During the simulation those with MMSE score 0-9 were reassigned to high dependency.

Medium dependency was defined as: needs help every, or most days, to put on shoes and socks (v542_w1) (proxy for dressing), or cook a hot meal every, or most days, (v539_w1), or from HAS (proxy) interview unable to dress without help (v53_h1).

Low dependency was defined as: needs help with wash all over/bath (v533_w1), or cutting toenails (v532_w1), or heavy housework (v390_w1), or shopping or light housework (v536_w1), or from HAS (proxy) interview considerable difficulty with household tasks (v51_h1).

A participant was defined as Independent they were not otherwise classified (and had no items missing from the other categories).

### ELSA:

High dependency was defined as: needs help using the toilet (headlwc), or chairfast (hemobch), or bedfast (headlbe), or has problems with continence (headlwc) and needs help putting on shoes and socks (headldr)).

Medium dependency was defined as: needs help putting on shoes and socks (headldr), or needs help to prepare a hot meal (headlpr).

Low dependency was defined as: needs help with bathing/showering (headlba), or difficulty pulling/pushing large objects (hemobpu), or difficulty doing work around house and garden (headlho).

A participant was defined as Independent they were not otherwise classified (and had no items missing from the other categories).

### US:

There were no items to define dependency in US that were comparable with ELSA or CFAS. IoN was therefore imputed by multiple imputation using chained equations based on age, sex and education.

Follow-up waves of IoN were calculated in the same way for each study.