Am J Transl Res 2022;14(10):7593-7606
www.ajtr.org /ISSN:1943-8141/AJTR0144094

Brief Communication

Latent class trajectory modelling: impact of changes in model specification

Charlotte Watson[1,2], Nophar Geifman[3], Andrew G Renehan[1,2]

[1]Manchester Cancer Research Centre and NIHR Manchester Biomedical Research Centre, Manchester, UK; [2]Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; [3]School of Health Sciences, Faculty of Health and Medical Sciences, University of Surry, Guildford, Surrey, UK

Received May 12, 2022; Accepted July 13, 2022; Epub October 15, 2022; Published October 30, 2022

Abstract: Latent class trajectory models (LCTMs) are often used to identify subgroups of patients that are clinically meaningful in terms of longitudinal exposure and outcome, e.g. drug response patterns. These models are increasingly applied in medicine and epidemiology. However, in many published studies, it is not clear whether the chosen models, where subgroups of patients are identified, represent real heterogeneity in the population, or whether any associations with clinically meaningful characteristics are accidental. In particular, we note an apparent over-reliance on lowest AIC or BIC values. While these are objective measures of goodness of fit, and can help identify the optimal number of subgroups, they are not sufficient on their own to fully evaluate a given trajectory model. Here we demonstrate how longitudinal latent class models can substantially change by making small modifications in model specification, and the impact of this on the relationship to clinical outcomes. We show that the predicted trajectory patterns and outcome probabilities differ when pre-specified cubic versus linear shapes are tested on the same data. However, both could be interpreted to be the "correct" model. We emphasise that LCTMs, like all unsupervised approaches, are hypotheses generating, and should not be directly implemented in clinical practice without significant testing and validation.

Keywords: Body mass index, cancer, statistical learning, latent variable modelling

Introduction

One method that can be used to identify patterns of change over time in a population is latent class trajectory modelling. This approach, derived from linear mixed models, was originally used in the psychology and criminology literature to identify patterns of behaviour [1, 2], and is now being increasingly utilised in the health sector to model patterns of drug response, disease progression, and other exposure variables over time. Latent class trajectory modelling (LCTM) brings together linear mixed models and latent growth curve modelling (LGCM). LGCM postulates the existence of one average pathway that exists for the population and can be modulated by each included individuals' random effects. LCTMs use group-based random effects instead to simplify a heterogeneous population into more homogeneous clusters or "classes" with respect to an unobserved latent variable [3, 4].

There are several examples of LCTMs in alcohol use and mental health literature, as well as other criminology and psychology studies. For example Cole et al, (2012) demonstrated using LCTM that using premorbid history could be clinically useful for subtyping schizophrenia patients [5] and Stevenson et al, (1996) showed how 72% of underlying behavioural problems in children could be linked to the mother's mental state [6]. Although variations of latent class modelling have been used since the 1960s [7], their use in human epidemiology is only just appearing in the literature. For example, Song et al. (2016) used these models to classify cohorts into latent classes using repeated BMI measures [8]. Other examples include BMI trajectories identified in participants who were later diagnosed with T2D, with three main patterns of change found, each linked to a different phenotypic profile [9]; the identification of three patient trajectory subgroups in Alzheimer's disease progression based on cognitive measures

over time [10]; and in rheumatoid arthritis, where several different trajectory-based phenotypic subgroups were identified in response to biologic therapies [11].

It has been argued that LCTM is well equipped for future forecasting and provides better accuracy for patient generalisation than other models as it can recognise the reoccurrence of patterns already observed in the data [12, 13]. In a more general sense, LCTM has advantages over using only one exposure measurement in that it can inform about at-risk populations and associations between covariates and may offer clinical benefit for identifying earlier adverse trajectories for intervention.

Methods for model selection and performance assessment are well established in prediction modelling, using techniques such as external validation and recalibration to ensure that models are not over-fitted and that they accurately represent the data (and disease). For latent class trajectory models, there is not as yet an established 'good practice'. Often, models are chosen either due to the lowest AIC or BIC, a well-known objective metric detailing how well the model fits the data, but one that does not take into account over-fitting and whether the participants are well assigned to their classes. Several studies have noted that the BIC tends to improve as more classes were added, until the model would no longer converge; this suggests that the BIC (or AIC) are not always the only instrument to use when selecting such models [14].

The interpretation of resulting models is prone to suffer from the subjectivity of the researcher examining the selected model, and the lack of a ground truth. As Ronan et al. summarised in their review of clustering techniques; "Biological systems are complex, so there are likely to be many relevant interactions between different aspects of the system, as well as meaningless relationships due to random chance" [15].

Here, we explore and demonstrate the variability of LCTMs and their interpretations in the context of BMI trajectories and obesity-related cancer incidence.

Methods

Our aim is to demonstrate how longitudinal latent class models can substantially change by making small modifications in model specification, and the impact of this on the relationship to clinical outcomes.

Data and participants

We used the female subset of the National Institutes of Health (NIH) - AARP cohort and derived trajectories for BMI change over time using the R *lcmm* package. Those with extreme BMI values (<15 and >60 kg/m^2) and values taken over the age of 80 were excluded. This cohort has recall BMI values at 18, 35, 50 and a current BMI at baseline. As one of the strengths of LCTMs is their ability to handle missing data well, we included any participant with at least 2 BMI measures after exclusions. The final cohort consisted of 130,979 individuals.

Model specification

Following our previously published framework for deriving these models [16], we ran a scoping model testing a quadratic shape and 5 classes (based on previous hypotheses in the field). By examining the residuals from this model, the random effect structure could be estimated. Here, the results suggested a cubic random-effect structure, which was then used in all tested models.

To ensure that the model had converged appropriately, multiple start points were run per model. If the same log-likelihood was reached for the majority of these runs, it could be assumed that the global maxima had been reached.

Up to 7 classes were tested for a variety of shapes (linear, quadratic, cubic, and cubic splines); the best-fit model (and number of resulting classes) was selected for each shape based on the lowest BIC achieved, up to a point where a decrease in BIC was not thought to add usefulness to the model, i.e. futility was reached.

As using BIC alone could overfit the model, the selected "best-fit" was then examined with other model fit statistics, such as the Average Posterior Probability of Assignment (APPA) and Odds of Correct Classification (OCC). This determined whether the model best represented the data, and whether participants were well assigned to their classes.

Once selected, the models were plotted, and the trajectories examined, to determine whether the resulting shapes were plausible in the context of BMI change over time. Sensitivity analyses were conducted to determine that the model shapes were stable (rerunning trajectories with only participants assigned with >80% probability), and that the random effect structure had taken into account all variability in the data (Elsensohn's envelope of residuals).

For illustrative purposes, we present the 3-class linear model and the 4-class cubic model (as each were selected as the best fitting for all classes tested for that shape).

Trajectory-class characterisation

To determine whether different class specifications had substantial differences in the characteristics of the population assigned to each group, the baseline characteristics for each trajectory assignment were examined. For continuous variables a one-way ANOVA was used to determine whether the difference in assignment per class was significant, and a χ^2 test was used for categorical variables.

To further illustrate how using different models can impact the final outcome, we determined incidence risk of 12 IARC obesity-related cancers (listed in Supplementary Material; multiple myeloma excluded due to variability within ICD code) based on class assignment using Cox regression models from time of cohort entry, and derived hazard ratios (HRs) and 95% confidence intervals (CIs).

Results

Model fit

Two sets of models were run on the same cohort with the same maximum number of classes, each specified with a different shape of trajectory. In both cases, the BIC decreased as the number of classes increased (**Figure 1A** for the linear trajectory model; **Figure 1B** for the cubic trajectory model). Applying the "elbow criterion" to BIC selection, resulted in a 3-class model selected for the linear trajectory specification and a 4-class for the cubic trajectory specification. To determine that these classes were the best fit for the data, other statistical metrics were examined, e.g. the odds of correct classification and the average posterior probability of assignments to each trajectory. These and others are presented in **Table 1**.

When the model convergence was examined over a number of iterations (**Figure 1C** for the linear models, **Figure 1D** for the cubic models), it can be seen that most models had stably converged and reached the same log-likelihood multiple times. This is important for model reproducibility, and accurate presentation of trajectories.

Trajectories and clinical characteristics

The individual trajectory classes differed in terms of change in BMI over time, with one class demonstrating a similar trajectory in both models (**Figure 2A**, **2B**). Depending on the number and shape of classes specified, different patterns were observed. In the linear 3-class model, we observed a "lean-moderate increase" with 63.3% of individuals, "lean-high increase" (31.3%) and "medium-increase" group (5.5%). Whereas, in the cubic 4-class model we observed a "lean-moderate increase" with 42.5% of individuals, "lean-high increase" (40.9%), "medium increase" (15.2%), as well as a "heavy increase" group (1.4%).

When these classes are compared to obesity-related cancer incidence risk, the importance of class membership becomes apparent. In the linear 3-class model, the two heavier classes carry a larger risk of cancer incidence compared to the lowest weight latent class (**Figure 2C**). However, in the cubic 4-class model, the "heavy increase" class had no significant increase in risk, due to the large confidence intervals (**Figure 2D**). The two other classes follow a stepwise progression in increased risk compared to the "lean-moderate increase" class.

The baseline characteristics for each model can be seen in **Tables 2**, **3**. In both models, an expected stepwise increase in body measures can be seen from the leanest to heaviest trajectories. This stepwise increase is also mimicked for diabetics and those with heart disease, with the highest proportion of both being found in the heaviest trajectories. Additionally, a higher number of participants with a high red meat consumption and high NSAID use were also assigned to the heaviest trajectories.

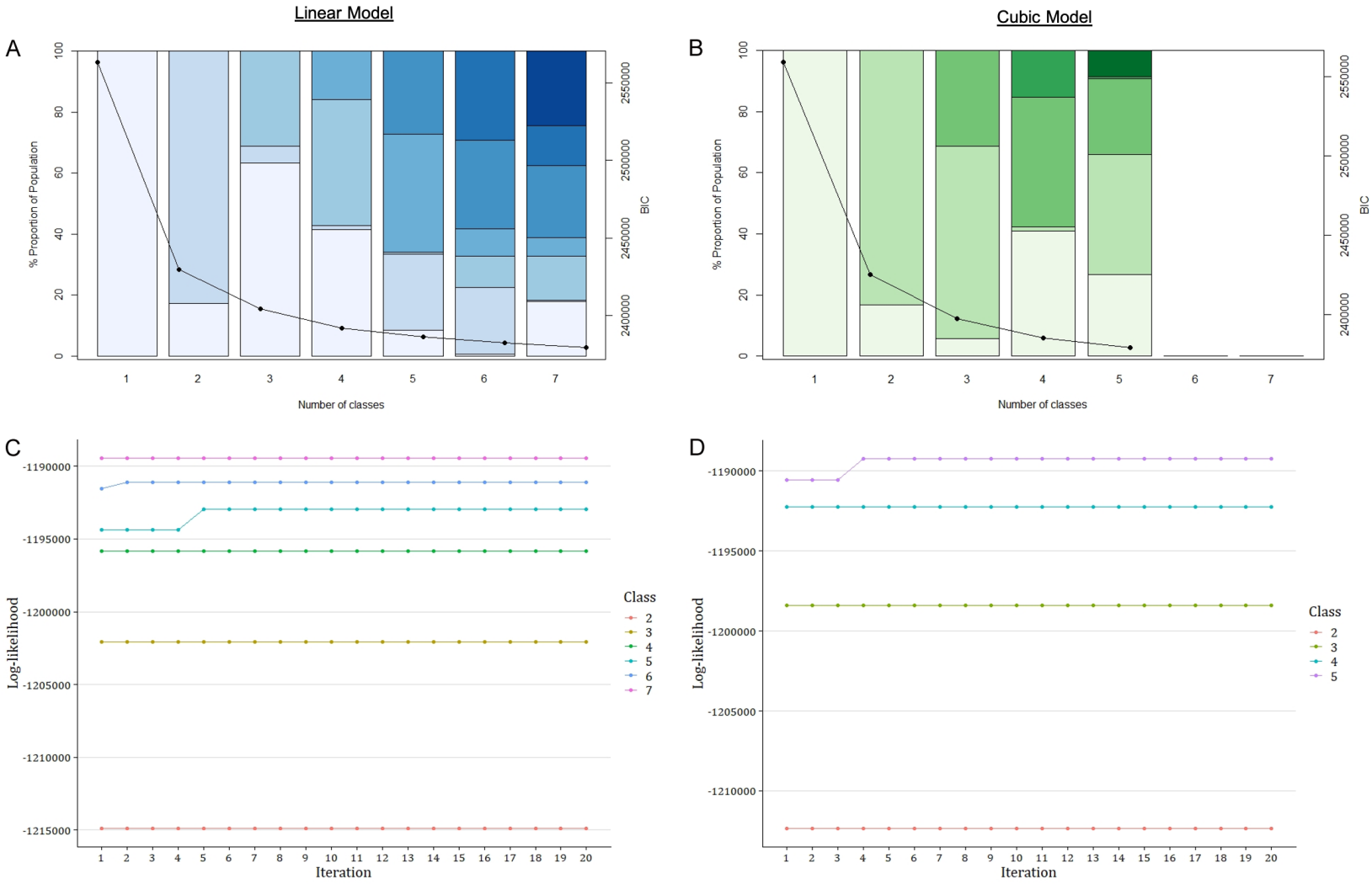


Figure 1. Model specification and selection. The two models are presented, with the linear model on the left and the cubic model on the right. (A and B) Detail model outputs from a linear and cubic specification, with BIC and class proportions illustrated. Best-fit models were selected, for each shape, based on BIC, using the "elbow criterion". The convergence of each of the start points for each tested model are presented in (C and D), where the log-likelihood of each run can be observed. Those that converge upon the same log-likelihood multiple times are likely to be a more stable model fit.

Table 1. Model selection criteria

k	APPA							OCC							Mismatch							Entropy	Relative Entropy	DoSk
Linear																								
1	-							-							-							-	-	-
2	0.913	0.955						43.3	5.1						-0.022	0.022						18933.2	0.791	3.372
3	0.870	0.894	0.840					4.7	123.6	9.8					0.047	-0.009	-0.037					45483.8	0.684	2.534
4	0.804	0.897	0.790	0.866				6.7	570.5	5.2	28.7				0.033	-0.002	-0.006	-0.026				60476.2	0.667	2.185
5	0.855	0.775	0.902	0.721	0.770			50.1	11.2	1218.9	4.6	8.0			-0.019	0.013	-0.001	0.029	-0.023			75782.7	0.641	2.094
6	0.899	0.753	0.604	0.856	0.680	0.774		1180.5	12.3	12.4	48.3	5.8	7.7		-0.001	0.02	-0.006	-0.020	0.024	-0.018		87324.6	0.628	2.162
7	0.627	0.902	0.638	0.851	0.674	0.754	0.764	8.3	1673.4	11.3	66.5	7.0	21.6	9.2	0.011	-0.001	0.008	-0.017	0.007	0.007	-0.016	96866.6	0.620	2.141
Quadratic																								
1	-							-							-							-	-	-
2	0.914	0.957						45.1	5.2						-0.021	0.021						18232.5	0.799	3.405
3	0.895	0.869	0.843					121.8	4.8	9.8					-0.009	0.046	-0.037					45063.7	0.687	2.586
4	0.807	0.898	0.791	0.865				6.6	544.1	5.3	29.6				0.034	-0.002	-0.007	-0.025				60179.2	0.669	2.239
5	0.905	0.724	0.774	0.774	0.858			1204.5	4.6	11.2	8.3	52.2			-0.001	0.029	0.012	-0.022	-0.018			75130.0	0.644	2.152
6	0.602	0.859	0.686	0.901	0.753	0.776		13.7	51.3	5.7	1181.0	11.9	7.9		-0.008	-0.019	0.025	-0.001	0.021	-0.017		86385.2	0.632	2.210
7	*							*							*							*	*	*
Cubic																								
1	-							-							-							-	-	-
2	0.913	0.958						45.3	5.2						-0.021	0.021						18130.0	0.800	3.432
3	0.894	0.871	0.843					120.1	4.8	9.9					-0.008	0.045	-0.037					44940.4	0.688	2.605
4	0.790	0.897	0.809	0.864				5.3	524.5	6.6	29.4				-0.006	-0.002	0.034	-0.026				60145.7	0.669	2.261
5	0.774	0.725	0.776	0.904	0.857			8.4	4.6	11.1	1186.6	52.4			-0.022	0.029	0.012	-0.001	-0.018			75059.3	0.644	2.165
6	*							*							*							*	*	*
7	*							*							*							*	*	*
Natural Splines																								
1	-							-							-							-	-	-
2	0.913	0.958						45.3	5.2						-0.021	0.021						18144.0	0.800	3.426
3	0.870	0.894	0.843					4.8	120.4	9.9					0.045	-0.008	-0.037					44932.8	0.688	2.603
4	0.808	0.791	0.899	0.864				6.6	5.3	533.7	29.3				0.034	-0.006	-0.002	-0.026				60164.0	0.669	2.259
5	0.774	0.903	0.775	0.725	0.857			8.4	1170.4	11.1	4.6	52.5			-0.022	-0.001	0.012	0.029	-0.018			75036.4	0.644	2.162
6	0.731	0.775	0.769	0.924	0.835	0.817		4.4	10.2	8.9	2842.5	106.1	104.5		0.022	-0.029	0.020	0.000	-0.007	-0.006		76197.2	0.675	2.974
7	*							*							*							*	*	*

*indicates that the model did not converge, models highlighted were selected for further analysis. APPA, Average posterior probability of assignment; OCC, Odds of correct classification; DoSk, Degrees of freedom. A model is considered to have a good fit with APPA ≥0.7, OCC ≥5, Mismatch close to 0, Relative Entropy close to 1.

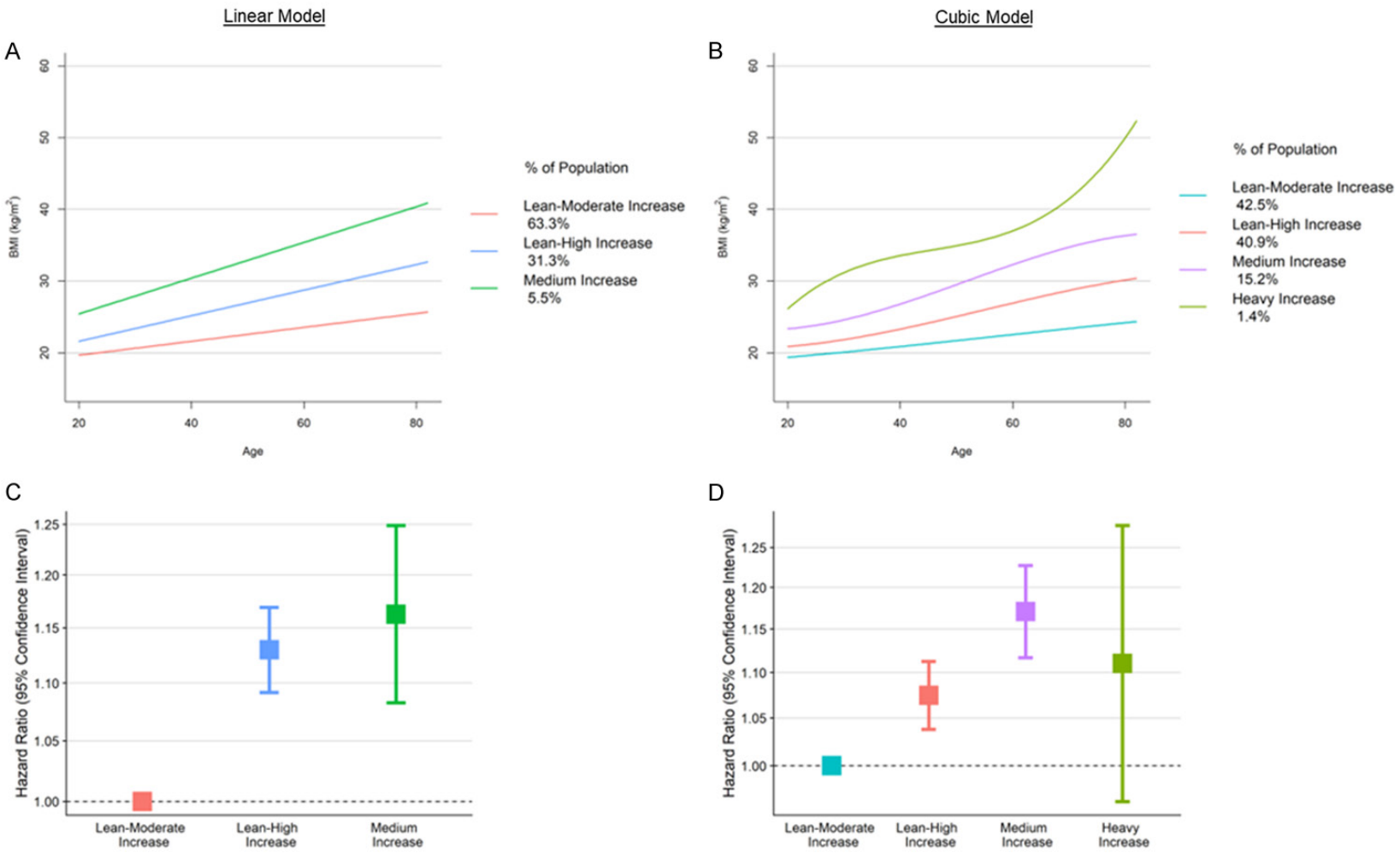


Figure 2. Selected trajectory models and class specific ORC risk. A and B. Show the trajectories within the selected model from each shape specification. C and D. Show the hazard ratios with a 95% confidence interval, for the relative obesity related cancer risk compared to the leanest class within each model. For the time-to-event analysis, both models were adjusted for smoking status (current, former, never) and stratified by age category at baseline (5-year age groups).

Table 2. Baseline characteristics per class for 3-class Linear model

3-class Linear	Lean-Moderate Increase	Lean-Heavy Increase	Medium Increase	*P*-value
No of participants	82907	40932	7140	
Mean entry age, years	62.6 (5.29)	61.46 (5.34)	59.49 (5.04)	<0.001
Body Measures				
Mean BMI at baseline, kg/m^2	23.97 (3.4)	30.64 (5.17)	36.31 (9)	<0.001
Mean BMI at 18, kg/m^2	19.69 (2.03)	22.11 (3.05)	26.06 (6.15)	<0.001
Mean BMI at 35, kg/m^2	21.07 (2.27)	24.53 (3.57)	30.15 (8.21)	<0.001
Mean BMI at 50, kg/m^2	22.52 (2.84)	27.92 (4.54)	34.37 (8.93)	<0.001
Mean maximum BMI, kg/m^2	25.21 (3.7)	32.73 (5.49)	41.22 (9.01)	<0.001
Mean minimum BMI, kg/m^2	19.27 (2.25)	21.4 (3.16)	23.61 (5)	<0.001
Mean absolute difference of maximum and minimum BMI, kg/m^2	6.07 (3.75)	11.45 (5.62)	17.87 (8.53)	<0.001
Mean waist circumference, cm	80.12 (10.03)	93.47 (13.52)	101.56 (18.27)	<0.001
Mean hip circumference, cm	99.95 (8.03)	111.4 (12.17)	121.08 (18.69)	<0.001
Mean waist to hip ratio	0.80 (0.08)	0.84 (0.09)	0.84 (0.09)	<0.001
Comorbidities				
Diabetes	2665 (3%)	4765 (12%)	1577 (22%)	<0.001
Heart disease	5876 (7%)	4189 (10%)	922 (13%)	<0.001
Smoking Status				
Never	36460 (44%)	18150 (44%)	2914 (41%)	<0.001
Former	32446 (39%)	16575 (40%)	3107 (44%)	
Current	11589 (14%)	4966 (12%)	874 (12%)	
Missing	2412 (3%)	1241 (3%)	245 (3%)	
Ethnicity				
Non-Hispanic White	76986 (93%)	36570 (89%)	6340 (89%)	<0.001
Black	2643 (3%)	2705 (7%)	525 (7%)	
Hispanic	1242 (1%)	691 (2%)	102 (1%)	
Asian	896 (1%)	220 (1%)	17 (<1%)	
Other	276 (<1%)	171 (<1%)	31 (<1%)	
Missing	864 (1%)	575 (1%)	125 (2%)	
Highest Educational level achieved, %				
<8 years	3400 (4%)	2553 (6%)	460 (6%)	<0.001
8-11 years	19259 (23%)	10545 (26%)	1749 (24%)	
Completed high school	8612 (10%)	4605 (11%)	861 (12%)	
Post high school or college	21216 (26%)	10118 (25%)	1864 (26%)	
College/Postgraduate	28325 (34%)	11876 (29%)	1979 (28%)	
Missing	2095 (3%)	1235 (3%)	227 (3%)	
Nutrition and Diet				
Mean alcohol consumption, g/day	7.09 (17.99)	4.79 (16.8)	3.82 (20.59)	<0.001
Mean red meat consumption, g/day	43.24 (37.99)	52.76 (50.28)	60.4 (62.55)	<0.001
Mean calorie consumption, kcal/day	1551.63 (671.36)	1623.62 (828.06)	1739.32 (982.23)	<0.001
At least 3 fruit servings a day, %	64382 (78%)	30900 (75%)	5302 (74%)	<0.001
At least 3 vegetable servings a day, %	57702 (70%)	27230 (67%)	4786 (67%)	<0.001
Mean total number of fruit servings a day	2.98 (2.37)	2.94 (2.49)	3.11 (2.91)	0.020
Mean total number of vegetable servings a day	3.8 (2.51)	3.91 (2.77)	4.24 (3.12)	<0.001
Drug Use				
Ibuprofen use 3+ times a week	14698 (18%)	10228 (25%)	2192 (31%)	<0.001
Missing	34800 (42%)	15595 (38%)	2608 (37%)	
Aspirin use 3+ times a week	21085 (25%)	11155 (27%)	2072 (29%)	<0.001
Missing	28049 (34%)	14983 (37%)	2873 (40%)	
HRT use ever, %	48505 (59%)	20312 (50%)	3140 (44%)	<0.001
Never	34402 (41%)	20620 (50%)	4000 (56%)	
Outcomes				
Number of cancer diagnoses, %	17446 (21%)	8808 (22%)	1471 (21%)	0.076
Number of obesity related cancer diagnoses, %	9698 (12%)	5177 (13%)	858 (12%)	<0.001
Number of non-obesity related cancer diagnoses, %	7748 (9%)	3631 (9%)	613 (9%)	0.006
Mean age of cancer diagnosis, y	70.29 (6.6)	69.33 (6.58)	66.95 (6.37)	<0.001
Mean age of obesity related cancer diagnosis, y	69.72 (6.56)	68.74 (6.5)	66.53 (6.44)	<0.001
Mean age of non-obesity related cancer diagnosis, y	70.99 (6.59)	70.17 (6.61)	67.54 (6.22)	<0.001

P value <0.05 considered statistically significant and were calculated using a one-way ANOVA for continous variables and χ^2 test for categorical. Data are given as mean ± s.d unless stated otherwise.

Table 3. Baseline characteristics per class for 4-class cubic model

4-class Cubic	Lean-Moderate Increase	Lean-High Increase	Medium Increase	Heavy Increase	*P*-value
No of participants	55645	53522	19932	1880	
Mean entry age, years	62.71 (5.27)	62.1 (5.34)	60.63 (5.26)	57.84 (4.48)	<0.001
Body Measures					
Mean BMI at baseline, kg/m^2	22.68 (2.93)	28.04 (3.67)	33.57 (7.25)	36.4 (9.9)	<0.001
Mean BMI at 18, kg/m^2	19.28 (1.82)	21.03 (2.4)	23.83 (4.43)	26.36 (7.61)	<0.001
Mean BMI at 35, kg/m^2	20.47 (1.99)	23.01 (2.62)	26.69 (5.34)	33.7 (10.46)	<0.001
Mean BMI at 50, kg/m^2	21.54 (2.38)	25.64 (3.26)	30.95 (6.61)	35.71 (10.53)	<0.001
Mean maximum BMI, kg/m^2	23.88 (3.07)	29.57 (4.19)	36.67 (7.31)	43.48 (10.29)	<0.001
Mean minimum BMI, kg/m^2	18.83 (2.11)	20.56 (2.58)	22.42 (3.99)	24.07 (5.82)	0.295
Mean absolute difference of maximum and minimum BMI, kg/m^2	5.14 (3.34)	9.06 (4.54)	14.49 (6.93)	19.76 (9.84)	<0.001
Mean waist circumference, cm	77.42 (8.83)	88.96 (11.26)	97.77 (16.42)	101.31 (18.7)	<0.001
Mean hip circumference, cm	97.8 (7.04)	107.15 (9.47)	116.27 (15.82)	121.34 (19.48)	<0.001
Mean waist to hip ratio	0.79 (0.08)	0.83 (0.08)	0.84 (0.1)	0.84 (0.1)	<0.001
Comorbidities					
Diabetes	1290 (2%)	3841 (7%)	3437 (17%)	439 (23%)	<0.001
Heart disease	3675 (7%)	4750 (9%)	2318 (12%)	244 (13%)	<0.001
Smoking Status					
Never	24343 (44%)	23904 (45%)	8540 (43%)	737 (39%)	<0.001
Former	21456 (39%)	21577 (40%)	8262 (41%)	833 (44%)	
Current	8249 (15%)	6431 (12%)	2512 (13%)	237 (13%)	
Missing	1597 (3%)	1610 (3%)	618 (3%)	73 (4%)	
Ethnicity					
Non-Hispanic White	52126 (94%)	48410 (90%)	17677 (89%)	1683 (90%)	<0.001
Black	1381 (2%)	2882 (5%)	1476 (7%)	134 (7%)	
Hispanic	725 (1%)	980 (2%)	303 (2%)	27 (1%)	
Asian	704 (1%)	361 (1%)	67 (<1%)	1 (<1%)	
Other	173 (<1%)	203 (<1%)	99 (<1%)	3 (<1%)	
Missing	536 (1%)	686 (1%)	310 (2%)	32 (2%)	
Highest Educational level achieved, %					
<8 years	2075 (4%)	2909 (5%)	1318 (7%)	111 (6%)	<0.001
8-11 years	12365 (22%)	13689 (26%)	5073 (25%)	426 (23%)	
Completed high school	5681 (10%)	5857 (11%)	2325 (12%)	215 (11%)	
Post high school or college	14423 (26%)	13242 (25%)	4998 (25%)	535 (28%)	
College/Postgraduate	19748 (35%)	16298 (30%)	5610 (28%)	524 (28%)	
Missing	1353 (2%)	1527 (3%)	608 (3%)	69 (4%)	

Nutrition and Diet					
Mean alcohol consumption, g/day	7.56 (18.35)	5.64 (17.15)	4.11 (17.55)	3.99 (20.19)	0.003
Mean red meat consumption, g/day	41.61 (36.74)	48.92 (44.53)	56.46 (55.06)	61.79 (70.1)	0.087
Mean calorie consumption, kcal/day	1548.08 (665.71)	1581.32 (748.22)	1672.89 (883.73)	1806.17 (1065.58)	<0.001
At least 3 fruit servings a day, %	43324 (78%)	40984 (77%)	14865 (75%)	1411 (75%)	<0.001
At least 3 vegetable servings a day, %	38916 (70%)	36377 (68%)	13153 (66%)	1272 (68%)	<0.001
Mean total number of fruit servings a day	3 (2.36)	2.92 (2.4)	2.99 (2.67)	3.25 (3.16)	<0.001
Mean total number of vegetable servings a day	3.8 (2.51)	3.84 (2.6)	4.01 (2.88)	4.62 (3.81)	<0.001
Drug Use					
Ibuprofen use 3+ times a week	9101 (16%)	11755 (22%)	5660 (28%)	602 (32%)	<0.001
Missing	23979 (43%)	20971 (39%)	7378 (37%)	675 (36%)	
Aspirin use 3+ times a week	13776 (25%)	14387 (27%)	5604 (28%)	545 (29%)	<0.001
Missing	18763 (34%)	18690 (35%)	7672 (38%)	780 (41%)	
HRT use ever, %	33461 (60%)	28489 (53%)	9148 (46%)	859 (46%)	<0.001
Never	22184 (40%)	25033 (47%)	10784 (54%)	1021 (54%)	
Outcomes					
Number of cancer diagnoses, %	11742 (21%)	11385 (21%)	4246 (21%)	352 (19%)	0.059
Number of obesity related cancer diagnoses, %	6467 (12%)	6559 (12%)	2502 (13%)	205 (11%)	<0.001
Number of non-obesity related cancer diagnoses, %	5275 (9%)	4826 (9%)	1744 (9%)	147 (8%)	0.001
Mean age of cancer diagnosis, y	70.37 (6.59)	69.91 (6.62)	68.31 (6.48)	65.68 (6.14)	<0.001
Mean age of obesity related cancer diagnosis, y	69.78 (6.55)	69.33 (6.59)	67.81 (6.38)	65.5 (6.21)	<0.001
Mean age of non-obesity related cancer diagnosis, y	71.09 (6.58)	70.68 (6.58)	69.03 (6.56)	65.93 (6.06)	<0.001

P value <0.05 considered statistically significant and were calculated using a one-way ANOVA for continous variables and χ^2 test for categorical. Data are given as mean ± s.d unless stated otherwise.

However, the proportion of overall cancers, and obesity-related cancers, were similar across all classes.

As the sample size is quite large, most of the differences in characteristics between classes appear significant. However, differences can be seen in the reported significance of some characteristics depending on model assignment. For example, in the 3-class linear model, red meat consumption is considered statistically significant, however in the 4-class cubic model, this characteristic is not reported as significantly different. This could impact further analyses depending on model choice, as a researcher using the 4-class cubic model would discount this variable from further analyses, whereas it would be included if using the 3-class linear model.

Overlap in resulting classes

Mapping the overlap in class membership between the two models demonstrates that the "lean-moderate increase" group in the linear model directly maps to and forms the entirety of the "lean moderate" group from the cubic model (**Figure 3**). The "lean-high increase" group in the cubic model seems to be made up of a split of participants from the "lean-moderate increase" and "lean-high increase" groups in the linear model. Finally, the "high increase" group in the cubic model appear to be a small subset of the "medium increase" group in the linear model and the "medium increase" group from the cubic model is mostly made up of the "lean-high increase" group and a small percentage of the "medium increase" group from the linear model. These results show that, moving from a 3-class model to a 4-class model, doesn't just add a new class (by splitting a previous group of individuals), but rather generates a number of very different classes in terms of individuals' composition.

Discussion

Summary

LCTMs are increasing in popularity as a tool to explore subgroups of patients with differing patterns of change in some disease measure over time. However, the extent to which any resulting model truly represents a population's heterogeneity is sometimes over-relied upon. Here we show that depending on the model specification different patterns and proportions of the trajectory assigned to each trajectory can change. For example, patients that form the "heavy increase" curve in the cubic model were included in the "medium increase" group in the linear model. When related to cancer incidence risk, only the linear "medium increase" group shows a significant effect of increased weight.

When applying the results of latent class modelling to a biological exposure, a clinically relevant explanation can often be applied to any pattern that occurs because of the enormous heterogeneity in a given disease's population. The results presented here demonstrate that care is needed when interpreting such models in a clinical setting. This could have clinical implications when implementing intervention strategies and cut-off values for intervention. To counter this, if using these models in a health setting, we advise clear presentation of results to include the caveats of using a model based on probability assignment and showing alternative models that have been tested before deciding on the "best" one.

Strengths and limitations

The strengths of the model development process shown above were the comprehensive selection and validation steps taken to ensure that the model best represented the data and presented plausible changes in BMI over time. This built on the current framework, developed previously within our research team [16], by the incorporation of multiple start points to ensure a reproducible final model and the extension of current code to include missing data at different time points for the Elsensohn's residual plots (Supplementary Figure 1).

There was one key limitation. Due to computational power (using a high performing computer cluster, these models generally took 7 days to run), a maximum of 7 classes only could be fitted to the data before convergence failed. However, as the models selected above appeared to have reached the best fitting model at 3/4 classes, this is unlikely to have overly impacted the results presented.

Context

The creation of the GRoLTs publishing checklist highlights the lack of reporting of model selec-

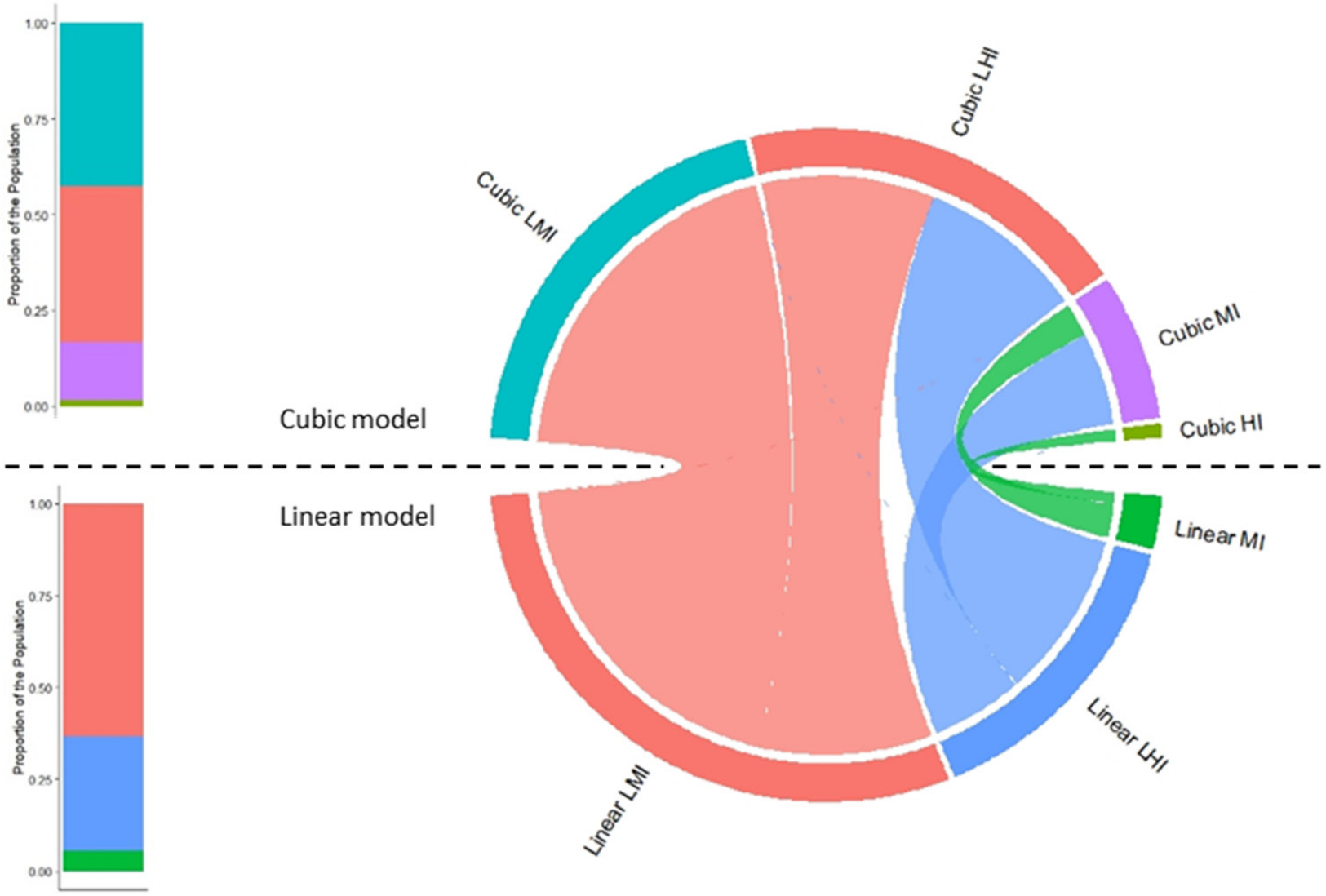


Figure 3. Chord plot showing population mapping between the linear and cubic models. The width of the bands between models denotes the proportion of individuals that are shared between the two corresponding classes. The bands in the top half relate to the 4 classes in the cubic model, while the bottom bands relate to the three classes in the linear model.

tion in current latent class publications. Although a useful guide for improving transparency in the development of these models, this serves merely as a reporting guideline, rather than a good practice handbook. This 16-item checklist focuses on the importance of including model selection, although more detail should be included to aid researchers [17]. In the guidance, BIC is suggested but there is no mention of other metrics such as the APPA, OCC or mismatch, as additional model checks, as has been suggested previously [18]. APPA measures how well individuals are assigned to their respective classes, and the overall average probability of being assigned to a class, whereas OCC examines the usefulness of the fitted model by comparing the odds ratio of the class assignment compared to a random assignment. Finally, mismatch describes the difference between originally estimated class proportions and the final model class proportions, therefore indicating how "certain" the assignment of individuals to each class are. All of this is important to consider when trying to generalise the outcome of these models onto a larger population.

BIC selection using the "elbow criterion" is equally subjective as it is up to the researcher to decide when the "bend" is and when a decrease in BIC does not necessarily mean a significantly better model. There is currently an R package (available at https://github.com/hlennon/LCTMtools) which calculates these statistics and compare the outputs of different models developed through the *lcmm* software. This gives a better indication of how well individuals are assigned to their respective classes and how well the overall model performs, in conjunction with the BIC statistic.

Despite the publication of these guidelines, latent class reporting still has a long way to go. For example, many journals have published studies on latent classes that simply choose the "best" model using BIC alone without extra statistical tests [19]. When used alone this could lead to overfitting the model to the dataset used and creating a model which is too specific to the cohort and does not generalise. Similar to overfitting in prediction models, this has a knock-on effect when class membership is applied to time to event data and produces a larger positive effect than would be observed in a general population.

The output of a latent class model is extremely dependent on the cohort it is derived from. More research is needed to make these models generalisable to the larger population. Therefore, reported results might only illustrate changes in modelling assumptions rather than true population changes. An example of this is a latent class analysis study by Ferreira et al. (2018) that clearly states the results are hypothesis generating only and need validation to assess the clinical implication.

Validation and verification of subgroup discovery is equally important. One way to validate any results would be to mimic the approach commonly taken in prediction modelling and use a new cohort to "test" the final model "learnt" from original cohort. As yet, there is no specific framework for conducting such validation in LCTMs, but it would give clear results as to whether the classes are "true" to a population rather than being cohort specific. Similarly to this idea, Seymour et al. (2019) derived latent classes in one cohort and then re-derived their groups in an external cohort and assessed the similarity.

Despite some limitations, latent classes can be a very useful tool for identifying subgroups that have not been previously categorised. Many studies have showed the added benefit of this analysis. For example, through clustering analysis, Sørlie et al. (2001) were able to differentiate different clinical outcomes between tumour subclasses and Geifman et al. (2018) identified 3 distinct subgroups of Alzheimer's patients with varying rates of cognitive decline and disease progression which could inform better patient selection for clinical trials.

Conclusions

Here, we showed that the predicted trajectory patterns and outcome probabilities differ when pre-specified cubic versus linear shapes are tested on the same data. Furthermore, such model specification differences could lead to differences in the time-to-event outcome hazard ratios if using class membership as a variable in a Cox model. Inevitably, this can have an impact on the reported relationship between variables and outcomes, as well as the significance of particular groupings; hence, great care should be taken when it comes to model selection.

Overall we show that there is evidence to suggest that there is a need for more robust methods to be implemented when fitting and reporting latent class trajectory models. We recommend that stringent guidelines should be enforced for the reporting of model development and testing, to ensure that results are reproducible and transparent.

Acknowledgements

This research was funded by the NIHR Manchester Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This work was also enabled by the Medical Research Council and the Engineering and Physical Sciences Research Council grant MR/N00583X/1 "Manchester Molecular Pathology Innovation Centre (MMPathIC): bridging the gap between biomarker discovery and health and wealth" (to NG).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Nophar Geifman, School of Health Sciences, Faculty of Health and Medical Sciences, University of Surry, Guildford, Surrey, GU2 7YH, UK. Tel: +44-1483682568; E-mail: n.geifman@surrey.ac.uk

References

[1] Lemoine M, Gmel G, Foster S, Marmet S and Studer J. Multiple trajectories of alcohol use and the development of alcohol use disorder: do Swiss men mature-out of problematic alcohol use during emerging adulthood? PLoS One 2020; 15: e0220232.

[2] Bennett KM, Morselli D, Spahni S and Perrig-Chiello P. Trajectories of resilience among widows: a latent transition model. Aging Ment Health 2020; 24: 2014-2021.

[3] Jung T and Wickrama K. An introduction to latent class growth analysis and growth mixture modelling. Soc Personal Psychol Compass 2008; 2: 302-317.

[4] Notelaers G, Einarsen S, De Witte H and Vermunt J. Measuring exposure to bullying at work: the validity and advantages of the latent class cluster approach. Work & Stress 2006; 20: 289-302.

[5] Cole VT, Apud JA, Weinberger DR and Dickinson D. Using latent class growth analysis to form trajectories of premorbid adjustment in schizophrenia. J Abnorm Psychol 2012; 121: 388-395.

[6] Stevenson J, Thompson MJ and Sonuga-Barke E. Mental health of preschool children and their mothers in a mixed urban/rural population. III. Latent variable models. Br J Psychiatry 1996; 168: 26-32.

[7] Gibson W. 3 multivariate models - factor analysis, latent structure-analysis and latent profile analysis. Psychometrika 1959; 24: 229-252.

[8] Song M, Willett WC, Hu FB, Spiegelman D, Must A, Wu K, Chan AT and Giovannucci EL. Trajectory of body shape across the lifespan and cancer risk. Int J Cancer 2016; 138: 2383-2395.

[9] Vistisen D, Witte DR, Tabák AG, Herder C, Brunner EJ, Kivim äki M and Færch K. Patterns of obesity development before the diagnosis of type 2 diabetes: the Whitehall II cohort study. PLoS Med 2014; 11: e1001602.

[10] Geifman N, Kennedy RE, Schneider LS, Buchan I and Brinton RD. Data-driven identification of endophenotypes of Alzheimer's disease progression: implications for clinical trials and therapeutic interventions. Alzheimers Res Ther 2018; 10: 4.

[11] Dagliati A, Plant D, Nair N, Jani M, Amico B, Peek N, Morgan AW, Isaacs J, Wilson AG, Hyrich KL, Geifman N and Barton A; BRAGGSS Study Group. Latent class trajectory modeling of 2-component disease activity score in 28 joints identifies multiple rheumatoid arthritis phenotypes of response to biologic disease-modifying antirheumatic drugs. Arthritis Rheumatol 2020; 72: 1632-1642.

[12] Qian T and Masino AJ. Latent patient cluster discovery for robust future forecasting and new-patient generalization. PLoS One 2016; 11: e0162812.

[13] Proust-Lima C, Séne M, Taylor JM and Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. Stat Methods Med Res 2014; 23: 74-90.

[14] D'Unger AV, Land KC and McCall PL. Sex differences in age patterns of delinquent/criminal careers: results from poisson latent class analyses of the philadelphia cohort study. J Quant Criminol 2002; 18: 349-375.

[15] Ronan T, Qi Z and Naegle KM. Avoiding common pitfalls when clustering biological data. Sci Signal 2016; 9: re6.

[16] Lennon H, Kelly S, Sperrin M, Buchan I, Cross AJ, Leitzmann M, Cook MB and Renehan AG. Framework to construct and interpret latent class trajectory modelling. BMJ Open 2018; 8: e020683.

[17] Van de Schoot R SM, Winter SD, Depaoli S and Vermunt JK. The GRoLTS-checklist: guidelines for reporting on latent trajectory studies. Struct Equ Modeling-a Multidisciplinary Journal 2016; 24: 451-467.
[18] Nagin DS. Group-based modeling of development. Harvard University Press 2005.
[19] Dhana K, van Rosmalen J, Vistisen D, Ikram MA, Hofman A, Franco OH and Kavousi M. Trajectories of body mass index before the diagnosis of cardiovascular disease: a latent class trajectory analysis. Eur J Epidemiol 2016; 31: 583-592.

Supplementary material

Obesity-related cancers

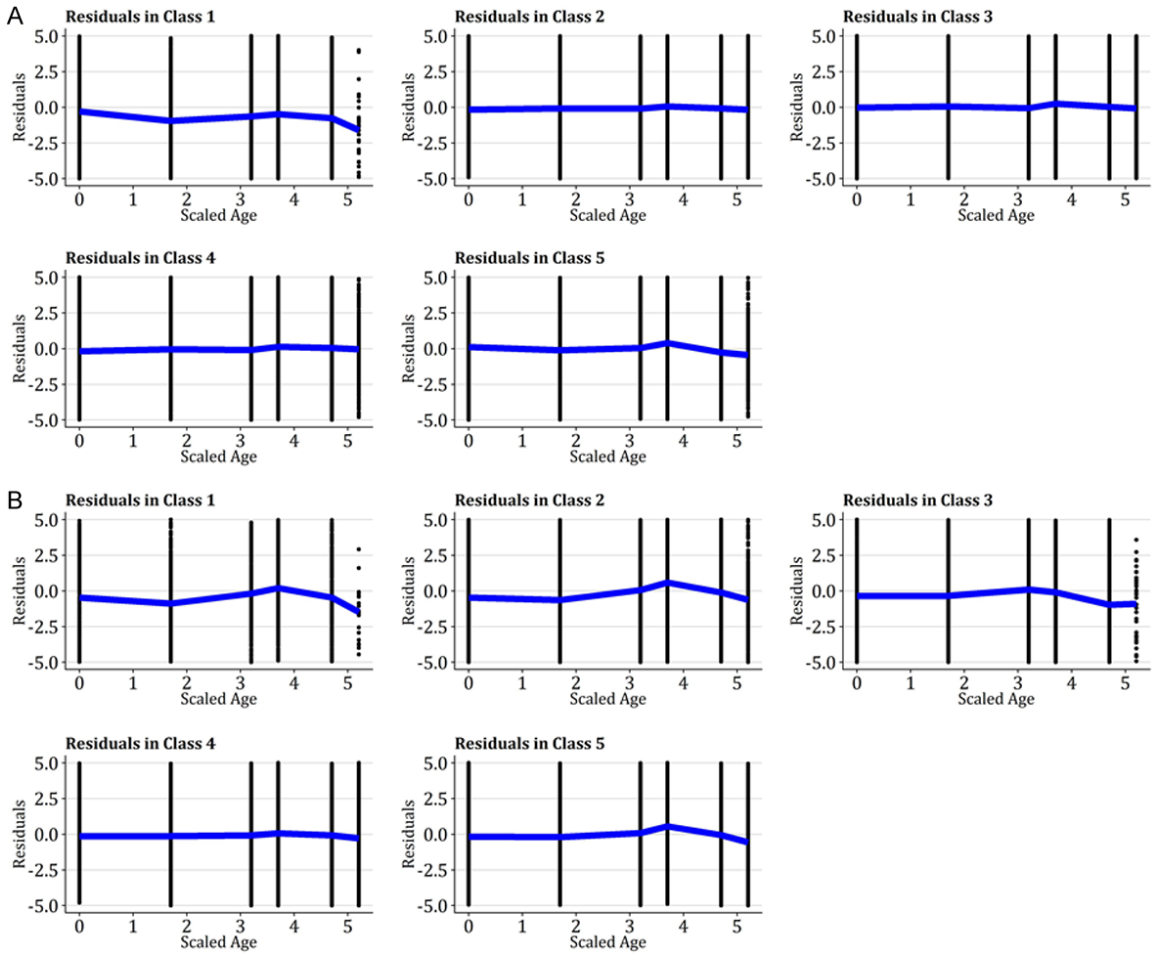
The International Agency for Research on Cancer (IARC) identified 13 obesity-related cancers - these are (with ICD-10 codes) as follows: Oesophagus - lower third (C15.5, C15.8); Colorectal (C18.0-18.9, C19.9, C20.0); Liver (C22.0); Gallbladder (C23.9); Pancreas (C25.0-25.9); Breast (C50.0-50.9); Corpus Uteri/Endometrial (C54.0-54.9, C55.9); Ovary (C56.9); Kidney (C64.9); Gastric cardia (C16.0); Malignant meningioma (C70.0, C70.1, C70.9); Thyroid (C73.0, C73.9); and Multiple myeloma (C90.0). We did not stratify breast cancer by menopausal status.

Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer-Viewpoint of the IARC Working Group. Colditz G, Anderson AS, Herbert RA, Kaaks R, Thompson HJ, Baker JL, Breda J, Byers T, Cleary MP, DiCesare M, Gapstur SM, Gunter M, Hursting SD, Leitzmann M, Ligibel J, Renehan A, Romieu I, Shimokawa I, Ulrich CM, Wade K, Weiderpass E. N Engl J Med 2016; 375: 794-8.

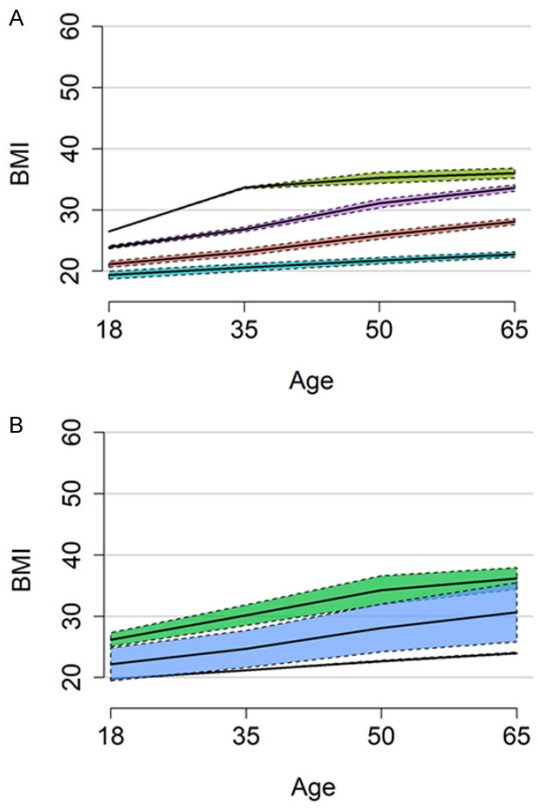
Random effect selection

To determine the random effects of the model, a 5-class quadratic model with no random effects was run, in line with other results in the literature. If the residuals were a horizontal line, then a random intercept could be considered, but no extra random effects needed to be incorporated in the model. If a diagonal line was observed then a linear random effect structure was suggested, and if a curve was observed then a quadratic random effect structure could be assumed. If an S-shaped curve was observed, then a cubic structure was chosen.

The figure below indicates the random effects for the AARP cohort in both men and women. In both panels, in at least 2 of the classes an ‘S-shaped’ curve is observed, therefore, a cubic random effect structure was chosen.



(A) shows the residuals for men, (B) for women. The age at entry was used, categorised into 5-year groups.



Supplementary Figure 1. Elsensohn's residual plots for each model. Residuals shown as a coloured "envelope" around the mean of each trajectory. Dashed lines indicate the border of each residual envelope. (A) shows the residuals the cubic 4 class model, (B) for the linear 3 class model. Similar-sized envelopes and parallel boundaries indicate a better fit.