The importance of household composition in epidemiological analyses of sleep: Evidence from the Understanding Society longitudinal panel survey

Helen Fowler, George T. H. Ellison, Eleanor M. Scott, Graham R. Law[*]

Temporal Influences on Metabolic Events (TIME) Research Group, Division of Epidemiology and Biostatistics, Leeds Institute of Genetics, Health and Therapeutics (LIGHT), University of Leeds, Leeds, UK
Email: [*]g.r.law@leeds.ac.uk

ABSTRACT

Aims: To establish the relationship between household composition and sleep, we: 1) used latent class analysis (LCA) to classify households; 2) examined the reliability and stability of household composition classes over time; 3) conducted multinomial logistic regression analyses to determine the relationship between household class and the self-reported sleep duration and quality of adults. Methods: Data were sourced from Waves 1 and 2 of the United Kingdom "Understanding Society" (USoc) longitudinal panel survey. LCA was used to classify household composition as a categorical latent construct using data on the number and ages of household occupants and the number of rooms used for sleeping. The Bayesian Information Criterion assessed model fit and identified the optimum number of latent classes. Multinomial logistic regression was used to investigate cross-sectional relationships between the household classes and self-reported sleep duration and quality amongst adults, after adjustment for confounders. *Results:* Household composition was best defined by 7 latent classes in data from Wave 1 of USoc. This finding was confirmed in Wave 2. Compared to the reference class (households with no children and no overcrowding), there was a higher risk of short sleep (≤5 hours) versus 7 - 8 hours sleep for latent household composition classes that included children (RR: 1.56; 95% CI: 1.29 - 1.89) and for those with both children and overcrowding (RR: 1.57; 95% CI: 1.31 - 1.88). Similarly the risk of "very bad" versus "fairly good" quality sleep was significantly higher in those household classes with overcrowding, particularly those with extended (RR: 1.75; 95% CI: 1.34 - 2.29) and large (RR: 1.51; 95% CI: 1.21 - 1.87) households. **Conclusion:** These analyses of a recent, nationally representative cohort from the UK, demonstrated that latent household composition classes are reliable over time; and that these latent household composition classes are important correlates of self-reported sleep amongst adult occupants. We showed that household composition is an important contextual variable to consider in most epidemiological studies of sleep.

1. INTRODUCTION

What is already known: Previous research has indicated that the number [1] and ages [2] of children in a household are associated with sleep duration amongst adult occupants. However, few epidemiological studies of sleep have operationalised and/or included household composition, and few have considered the role of household composition on both sleep duration and quality.

What this study adds: This research project demonstrates that household composition can be reliably operationalised using data on the age structure of household occupants and overcrowding to reveal the role that: the number of children in the household; large and extended families; and overcrowding play as predictors of sleep duration and sleep quality amongst adult household members.

Associations between sleep and cardiovascular disease, diabetes and mortality have been observed across a range of different studies and contexts [2]. However, fewer studies have documented the many contextual factors that predict or influence sleep, or the role of social structures, such as household composition, on sleep. Previous work examining which household characteristics might predict sleep has primarily considered relationships between the number of household occupants and either the number of children in the home or the ages of any children in the home [1-3]. Typical of these studies was that by Chapman *et al.* [1] who found that the frequency of "insufficient sleep" was highest in households with at least 3 children. However, most such studies have failed to fully explore the inter-relationships between these characteristics, and the role that broader sociodemographic patterns and housing structures might play as predictors of sleep.

Notable exceptions include Burgard *et al.* [4], who adopted a latent class analysis (LCA) approach to categorise study participants according to their "life course stage". Their study, and that by Groeger *et al.* [3], suggested that factors such as partnership status and parental status have a significant association with sleep duration such that childless adults, for example, tend to have longer sleep durations compared than adults living with children. Neither of these studies (and none that we have been able to find) used LCA to consider overcrowding in the home as a potential predictor of sleep alongside the sociodemographic distribution of household occupants. Overcrowding is likely to have a substantial impact on sleep, and on the impact that partnership status, parental status and the number/ages of children in the household have on sleep.

To address this issue, the present study used data from Understanding Society (USoc), the United Kingdom's nationally representative household panel survey, to: define household composition in terms of latent household classes; and examine the relationship between household composition and self-reported sleep duration and quality. The study aimed to extend our understanding of the household characteristics that are associated with sleep and thereby inform future epidemiological research, including the potential development of sleep interventions that are sensitive to variation in household composition.

2. METHODS

2.1. Study Participants

The first Wave of USoc [5] collected data from January 2009 to January 2011, and involved 50,994 participants from 30,169 households. The second Wave involved 54,597 participants from 30,508 households and took place from January 2010 to April 2012. Individual respondent data were collected from adults (16 years and over) using computer assisted personal interviews. Household data were collected from one adult member of each household acting as a key informant.

2.2. Household Composition Indicator Variables

The household variables used in the latent class analyses were the: number of adults; number of children; number of couples; number of single parents; number of pensioners; and an ordinal variable indicating the extent to which there was overcrowding in the home. This ordinal variable was created using the UK Government's definition of household overcrowding, based on the "Bedroom Standard" measure within the "Household (Overcrowding) Bill" [6]. This involved calculating the ratio of the number of rooms used for sleeping (*i.e.* bedrooms) to the Bedroom Standard as a measure of overcrowding [6]. The Bedroom Standard was calculated using information provided by all household respondents on the total number of residents by age and sex in the household, a single bedroom being assigned for: each adult couple; any single adults aged 21 years and over; all pairs of same sex 10 - 20 year olds; and any pairs of children under 10 years of age, regardless of sex.

2.3. Sleep Variables

The present analyses drew on two sleep-related variables recorded during the first Wave of the study using the "Self Completion Adult Module" questionnaire. This collected self-reports of sleep duration (generating continuous data on hours of sleep) and sleep quality (generating categorical data based on "very good" to "very bad" sleep quality), based on two items in the questionnaire:

- Sleep duration: "How many hours of actual sleep did you usually get at night during the last month? This may be different than the actual number of hours you spent in bed".
- Sleep quality: "During the past month, how would you rate your sleep quality overall?" Very good; Fairly good; Fairly bad; Very bad.

2.4. Descriptive Covariates, Confounders and Competing Exposures

Variables extracted from the Wave 1 USoc questionnaire to further characterize the latent household composition classes, and/or adjust for potential confounding in the polynomial logistic regression analyses of the relationship between household composition class and sleep duration/quality requiring adjustment, comprised: respondent age (calculated to the nearest year using self-reported date of birth and questionnaire completion date); respondent sex (as reported); the number of children

aged 0 - 2, 3 - 4, 5 - 11 and 12 - 15 in the household; adult marital status (single, married/civil partnership, separated, divorced or widowed); highest educational qualification achieved (University degree+, post-school qualifications; A-level or equivalent; GCSE or equivalent, Other school qualification or none); ethnicity (White British, Other White background, Black, Black mixed race, Asian, Asian mixed race, or Other ethnic group); and SF12 single item self-rated health (Excellent, Very good, Good, Fair, Poor).

To establish which of the covariates might operate as potential confounders, mediators or competing exposures in the multinomial logistic regression analyses exploring the relationship between latent household composition class and sleep duration/quality, a causal path diagram was constructed in the form of a Directed Acyclic Graph (DAG), drawing on established and hypothesised functional relationships between the exposure variable (latent household composition class), the outcome variable (sleep duration/quality) and each covariate [7]. This DAG was uploaded to the freeware program DAGitty.net (see: www.dagitty.net; [8]) which uses established algorithms [9] to identify confounders (which cause both exposure and outcome), mediators (which are caused by the exposure and cause the outcome), and competing exposures (which are unrelated to the exposure but cause the outcome). DAGitty.net uses additional algorithms to identify any "minimum sufficient adjustment sets" (*i.e.* groups of available/measured covariates) that need to be included in any multivariable analyses to eliminate confounding (using "back-door paths"; [10]). The benefit of this approach is that it provides an explicit *a priori* model of the postulated relationships between the exposure and outcome variables and each of the available covariates. Such models are invaluable for the specification and subsequent verification of the statistical analyses, to ensure that these avoid the risk of over-adjustment or inappropriate adjustment for mediators (*i.e.* covariates that lie on the causal path between exposure and outcome; [11]). These models also help to establish whether the statistical models used are the most parsimonious (and therefore those with optimum statistical power) rather than simply (over)adjusting for all available covariates (with the resulting reduction in statistical power; [12]).

2.5. Latent Class Analysis (LCA)

Household manifest variables from both Waves 1 and 2 of USoc were analysed using latent class analysis (LCA) to define household composition as a latent categorical construct, using the sociodemographic variables and overcrowding index described earlier, and Latent Gold 4.5 software [13]. The manifest variables were treated as either indicator variables (*i.e.* dependent variables of the latent classes) or covariates (*i.e.* variables used to de-scribe rather than measure or define the latent classes; [14]). Using an exploratory LCA approach, models were produced using indicators and covariates, each model defining the data based on a different number of latent classes. The Bayesian Information Criterion (BIC) was used to evaluate model fit and parsimony. The BIC provides a criterion for model selection amongst a group of models with differing numbers of parameters by taking into the account the sample size, the number of parameters included, and the maximized log-likelihood of each model [15]. The BIC also introduces a penalty term for the number of parameters in a model, and thereby accounts for over-fitting. The model with the lowest BIC value was therefore considered the best fitting model [16]. The best fitting models generated separately for data from Wave 1 and Wave 2 of USoc were then compared to assess the reliability and stability of latent household composition classes over time.

2.6. Multinomial Logistic Regression Analysis

Multinomial logistic regression analysis was undertaken on the Wave 1 data using Stata/IC 12.1 [17], to explore the relationship between latent household composition class membership and both sleep duration and sleep quality, after adjustment for any potential confounders and competing exposures identified using the DAG. These models generated measures of relative risk (RR) and their associated 95% confidence intervals (95%CI) which indicated the likelihood of experiencing any particular sleep outcome as compared to a referent sleep outcome for each of the latent household composition outcomes. For these analyses, self-reported sleep duration was categorized as: 5 hours or less; 5 - 6 hours; 6 - 7 hours; or 10 hours or more, and compared to a sleep duration of 7 - 8 hours as the referent. Self-reported sleep quality was categorized using the questionnaire item's original answer categories as: "Very bad"; "Fairly bad"; or "Very good", and compared to "Fairly good" sleep quality as the referent.

3. RESULTS AND DISCUSSION

3.1. Household Composition

Following an exploratory LCA of Wave 1 household manifest variables and covariates, a comparison of the BIC values for models with between 1 and 8 latent classes resulted in the selection of the model with 7 latent classes as the best-fitting model, based on the lowest log-likelihood BIC value of 267,652 (see **Table 1**). The labeling of each of the seven latent classes in the selected model, and the proportion and number of adult participants falling within each of these classes, has been summarized in **Table 2**. The labels developed to describe each of the seven latent household composition classes

were based on the distribution of the manifest household variables, covariates and individual demographic characteristics, as detailed in **Table 3**.

The "Empty Nesters" class represents the largest latent household composition class (in which 51% of the total USoc sample were living), none of which had children living in the household, some of which comprise pensioners, but most of which comprise working-age adult couples. Members of this latent household composition class were all (100%) living in properties with more bedrooms than the Bedroom Standard and as such did not experience overcrowding.

The "Dwelling Sharing Childless Adults" class had fewer couples than the "Empty Nesters" and tended to represent adults living in smaller properties relative to the "Empty Nesters", the majority (89.9%) having the same number of bedrooms as that prescribed by the Bedroom Standard.

The "Partnered with Child" and "Partnered with Children" classes were living in similar sized dwellings relative to household occupancy, the majority of these households (65.7% and 69.8%, respectively) having more bedrooms that the Bedroom Standard requirement.

However, these classes had different numbers of children, in that "Partnered with Child" had a mean of just 1.3 children in the household whereas "Partnered with Children" had at least one additional child (as reflected in the mean of 2.3 children).

In the "Large Family, Majority with Overcrowding" class the majority of adults (mean = 2.231) were part of a couple with fewer single parents. This latent household composition class had the largest number of children of all classes (mean = 2.9), and the majority (63.2%) of these families were living in dwellings with fewer bedrooms than the Bedroom Standard requirement, the remainder (36.8%) living in homes that match the Bedroom Standard requirement.

The majority of adults living in the "Single Parent Household" class were single parents with a mean number of 1.4 children, and lived in properties that either met (61.5%) or had more bedrooms than (38.5%) the Bedroom Standard.

Finally, the "Extended Family, Majority with Overcrowding" class had the largest rate (at 76.5%) of overcrowding (*i.e.* the greatest proportion of households with fewer bedrooms than the bedroom standard) of all the latent household composition classes, and the largest number of adult occupants (mean = 2.630). The majority of adults were more likely to be part of a couple, less likely to be single parents and, on average, were living with 1 or more children (mean = 1.4).

To assess the reliability/temporal stability of the latent household composition classes identified from the analysis of data from Wave 1, these analyses were repeated with data from Wave 2 of USoc. Comparison of the BIC of models with between 1 and 8 latent classes demonstrated that a very similar model containing 7 similar latent classes was again the model of best fit with the lowest BIC of 268,362. Further similarities were evident from a comparison of the distribution of these classes, and of household manifest variables, covariates and indi-

Table 1. Latent class BIC model fit statistics.

Number of classes	Log–likelihood BIC	Classification error
1	329,715	-
2	273,117	0.0002
3	270,637	0.0005
4	269,903	0.0193
5	269,313	0.1973
6	268,415	0.1265
7	267,652	0.0267
8	267,753	0.0272

Table 2. Descriptive labels chosen for each of the latent household composition classes identified through LCA of Wave 1 USoc data (see **Table 1**).

LCA class	Household category label	Modal posterior probability	
		Frequency	Sample proportion
1	"Empty Nesters"	15,326	0.509
2	"Dwelling Sharing Childless Adults"	4908	0.163
3	"Partnered with Child"	3643	0.121
4	"Partnered with Children"	1897	0.063
5	"Large Family, Majority with Overcrowding"	1746	0.058
6	"Single Parent Household"	1626	0.054
7	"Extended Family, Majority with Overcrowding"	1024	0.034

Table 3. Household manifest variables, covariates and individual characteristics of the seven latent household composition classes identified through LCA of Wave 1 USoc data (see **Table 1**) and labeled in **Table 2**.

	Household Latent Class						
	1	2	3	4	5	6	7
Estimated population proportion	0.509	0.162	0.123	0.058	0.059	0.061	0.027
Household manifest variables							
Age and marital status of household occupants							
Mean number of adults	1.830	2.083	2.345	2.098	2.231	1.265	2.630
Mean number of children	0.000	0.000	1.290	2.332	2.932	1.359	1.382
Mean number of couples	0.585	0.319	0.987	0.971	0.759	0.000	0.695
Mean number of single parents	0.000	0.000	0.023	0.048	0.304	0.997	0.388
Mean number of pensioners	0.651	0.331	0.033	0.025	0.042	0.009	0.109
Bedroom Standard (BS) Occupancy							
Proportion with < Bedrooms than BS	0.000	0.099	0.000	0.000	0.632	0.000	0.765
Proportion with Bedrooms ≡ BS	0.000	0.899	0.343	0.302	0.368	0.615	0.235
Proportion with > Bedrooms than BS	1.000	0.001	0.657	0.698	0.000	0.385	0.000
Covariates							
Proportion with children aged 0 - 2 yrs	0.000	0.000	0.254	0.469	0.435	0.180	0.297
Proportion with children aged 3 - 4 yrs	0.000	0.000	0.086	0.501	0.355	0.154	0.066
Proportion with children aged 5 - 11 yrs	0.000	0.000	0.331	0.797	0.861	0.465	0.448
Proportion with children aged 12 - 15 yrs	0.000	0.000	0.324	0.295	0.636	0.357	0.184
Individual characteristics							
Sex							
Male	0.472	0.502	0.473	0.458	0.412	0.128	0.409
Female	0.528	0.498	0.527	0.542	0.588	0.872	0.591
Adult age in years							
Mean	53.7	42.1	37.4	36.3	35.5	33.8	35.0
Standard deviation	18.3	19.5	11.2	7.9	10.4	11.1	13.5

vidual characteristics amongst these (compare **Tables 2** and **4**). On this basis, the labels used to describe each of the latent household composition classes in **Table 3** would also apply to the best fitting LCA model of Wave 2 data, the only notable difference being the increased size of the so-called "Empty Nesters" class (Wave 1: 50.9%; Wave 2: 54.1%) and the modest reduction in the size of the "Dwelling Sharing Childless Adults" class (Wave 1: 16.2%;; Wave 2: 14.4%). This is likely to have been explained, at least in part, by the 11.2% increase in household participation between Waves 1 and 2 [5].

3.2. Household Composition and Sleep

The DAG summarising the hypothesised functional rela-

tionships that informed the design of the multivariate statistical analyses was uploaded to DAGitty.net to identify any minimum sufficient adjustment set(s) required to address potential confounding. Based on this model, a single minimum sufficient adjustment set was identified for addressing potential confounding in the relationship between latent household composition class and sleep duration/quality. This set comprised: age; sex; marital status; educational attainment; and ethnicity. However, as a result of uncertainty regarding the directionality of likely causal relationships between self-reported health and household class, and between self-reported health and sleep duration/quality within the USoc study population, it was decided to add self-reported health to the

Table 4. Household manifest variables, covariates and individual characteristics of the seven latent household composition classes identified through LCA of Wave 2 USoc data (see **Table 1**) and labeled in **Table 2**.

	Household Latent Class						
Proportions/Means	1	2	3	4	5	6	7
Estimated population proportion	0541	0.144	0.126	0.059	0.055	0.051	0.024
Household manifest variables							
Age and marital status of household occupants							
Mean number of adults	1.840	2.132	2.334	2.098	2.239	1.299	2.696
Mean number of children	0.000	0.000	1.312	2.342	2.848	1.331	1.402
Mean number of couples	0.589	0.343	0.990	0.967	0.763	0.000	0.729
Mean number of single parents	0.000	0.000	0.017	0.046	0.295	0.996	0.350
Mean number of pensioners	0.695	0.376	0.033	0.016	0.044	0.014	0.113
Bedroom Standard (BS) Occupancy							
Proportion with < Bedrooms than BS	0.000	0.099	0.000	0.000	0.616	0.000	0.764
Proportion with Bedrooms ≡ BS	0.000	0.899	0.330	0.288	0.385	0.596	0.236
Proportion with > Bedrooms than BS	1.000	0.001	0.670	0.712	0.000	0.404	0.000
Covariates							
Proportion with children aged 0 - 2 years	0.000	0.000	0.238	0.455	0.392	0.141	0.237
Proportion with children aged 3 - 4 years	0.000	0.000	0.083	0.534	0.316	0.144	0.080
Proportion with children aged 5 - 11 years	0.000	0.000	0.354	0.792	0.869	0.437	0.474
Proportion with children aged 12 - 15 years	0.000	0.000	0.325	0.293	0.687	0.411	0.208
Individual characteristics							
Sex							
Male	0.471	0.486	0.486	0.476	0.410	0.144	0.428
Female	0.529	0.514	0.514	0.524	0.590	0.856	0.572
Adult age in years							
Mean	54.5	43.3	37.5	36.3	35.7	33.9	35.2
Standard deviation	18.3	20.0	11.3	8.0	10.7	11.9	13.9

minimum sufficient adjustment set.

As such this is the minimum sufficient adjustment set required for estimating the direct (as opposed to the total) effect of latent household composition class on sleep duration/quality. The largest latent household composition class ("Empty Nesters") was then used as the reference group for comparison with each of the remaining household classes, since this class did not contain two ostensibly important household characteristics: neither children nor overcrowding. The results of these analyses have been summarized in **Tables 5** and **6**.

In terms of sleep duration (**Table 5**), these results indicate that, compared to adults living in households in the "Empty Nesters" class, adults in all of the remaining household classes had an increased risk of very short (*i.e.* <5 hrs) as opposed to 7 - 8 hrs sleep (the referent category). Adults in latent household composition classes with the highest risk of very short sleep duration included those living in classes labeled "Partnered with Children" (RR: 1.56; 95% CI: 1.29, 1.89) and "Large Family, Majority with Overcrowding" (RR: 1.57; 95% CI: 1.31, 1.88), adults living in the latter also having an increased risk (RR: 2.12; 95% CI: 1.54, 2.90) of very long (≥10 hrs) sleep duration.

In terms of sleep quality (**Table 6**), these results indicate that, compared to adults living in households in the "Empty Nesters" class, adults in all other household classes had an increased risk of "very bad" compared to

Table 5. Multinomial logistic regression exploring the relationship between latent household composition class and sleep duration, after adjustment for potential confounders[1].

	≤5 hrs vs. 7 - 8 hrs			5 - 6 hrs vs. 7 - 8 hrs			6 - 7 hrs vs. 7 - 8 hrs			≥10 hrs vs. 7 - 8 hrs		
	RR	95% CI		RR	95% CI		RR	95% CI		RR	95% CI	
Household classes												
Empty Nesters (referent)	1.00			1.00			1.00			1.00		
Dwelling Sharing Adults	1.28	1.15	1.43	0.99	0.90	1.08	1.00	0.92	1.08	1.25	0.98	1.61
Partnered with Child	1.32	1.15	1.51	1.21	1.10	1.34	1.17	1.07	1.27	1.11	0.82	1.52
Partnered with Children	1.56	1.29	1.89	1.49	1.30	1.71	1.37	1.22	1.53	0.60	0.34	1.09
Large Family, Majority with Overcrowding	1.57	1.31	1.88	1.42	1.23	1.63	1.15	1.02	1.31	2.12	1.54	2.90
Single Parent Household	1.24	1.02	1.51	1.17	0.99	1.37	1.05	0.91	1.21	0.94	0.63	1.40
Extended Family, Majority with Overcrowding	1.47	1.17	1.84	1.07	0.89	1.30	1.00	0.85	1.18	1.09	0.69	1.72

[1]Adjusted for: sex; age; marital status; educational attainment; ethnicity; and self-reported health.

Table 6. Multinomial logistic regression exploring the relationship between latent household composition class and sleep quality, after adjustment for potential confounders[1].

	Very bad vs. fairly good			Fairly bad vs. fairly good			Very good vs. fairly good		
	RR	95% CI		RR	95% CI		RR	95% CI	
Household classes									
Empty Nesters (referent)	1.00			1.00			1.00		
Dwelling Sharing Adults	1.38	1.20	1.59	1.00	0.92	1.08	0.98	0.91	1.06
Partnered with Child	1.30	1.09	1.55	1.11	1.01	1.21	0.88	0.81	0.95
Partnered with Children	1.18	0.91	1.54	1.24	1.10	1.4	0.85	0.76	0.95
Large Family, Majority with Overcrowding	1.51	1.21	1.87	1.19	1.05	1.34	0.99	0.88	1.11
Single Parent Household	1.17	0.93	1.48	1.24	1.08	1.41	0.93	0.81	1.07
Extended Family, Majority with Overcrowding	1.75	1.34	2.29	1.22	1.04	1.43	0.87	0.75	1.01

[1]Adjusted for: sex; age; marital status; educational attainment; ethnicity; and self-reported health.

"fairly good" sleep. Adults in the latent household composition classes with the highest risk of having "very bad" sleep quality were those living in classes with overcrowding: "Extended Family, Majority with Overcrowding" (RR: 1.75; 95% CI: 1.34, 2.29) and "Large Family, Majority with Overcrowding" (RR: 1.51; 95% CI: 1.21, 1.87). Adults living in household classes with overcrowding also had an increased risk of "fairly bad" (versus "fairly good") sleep quality, although those living in the "Single Parent Household" class (RR: 1.24; 95% CI: 1.08, 1.41) and the "Partnered with Children" class (RR: 1.24; 95% CI: 1.10, 1.4) had the highest relative risks in this category. Adults living in both the "Partnered with Children" class and the "Partnered with Child" class also had a reduced risk of "very good" versus "fairly good" sleep quality (RR: 0.85; 95% CI: 0.76, 0.95, and RR:0.88;

95% CI: 0.81, 0.95, respectively), indicating the importance of children in the household as a predictor of better than "fairly good sleep".

3.3. Limitations

The data and analytical techniques used in the present study are subject to four key limitations, namely: a high prevalence of missing data for some of the key variables; variation in the numbers of respondents in different latent household composition classes; the somewhat subjective and limited acuity of latent class analysis (LCA); and incomplete adjustment for potential confounders in the relationship between latent household composition class and sleep duration/quality.

Despite the gift vouchers offered to USoc participants (and the separate risk this entails of participation bias), a

substantial proportion of data were missing for many of variables included in the analyses. This was most pronounced for self-reported sleep duration and sleep quality, and as such the multinomial logistic regression analyses exploring the relationship between latent household composition class and sleep duration/quality (**Tables 5** and **6**) were conducted on a small subset of those respondents for whom data on sociodemographic and overcrowding data were available for the LCA summarized in **Tables 1-3**. Thus, while we are confident that the latent household composition classes identified from our analysis of the nationally representative USoc data are likely to reflect the distribution of these classes for the data set (and therefore the UK) as a whole, any tendency for differential reporting of sleep duration/quality by respondents from each of these household composition classes may undermine the strength and validity of their relationships with sleep (as summarized in **Tables 5** and **6**).

Indeed, the relatively small numbers of respondents in some of the latent household composition classes (particularly those containing children), will have inevitably reduced the statistical power of any relationship between these classes and self-reported sleep duration/quality. However, without conducting additional data collection to create boosted samples of respondents of all but those in the so-called "Empty Nesters" class, there was still more than a thousand (n = 1024; see **Table 2**), adult respondents in the class with the least respondents (the so-called "Extended Family, Majority with Overcrowding" class); and this sample size should have been more than sufficient to reveal any statistical association between household class and sleep duration/quality, even after the loss of respondents with missing data and the reduction in power incurred by adjustment for six potential confounders.

The choice of latent variable analysis used is therefore, perhaps, a more important potential limitation facing the present study. Although LCA is ostensibly an appropriate choice for classifying household composition, the categorization involved is essentially subjective. In particular, it was necessary to interpret means of count data (such as "1.84 adults") to derive conceptually plausible labels for each of the latent classes—a process that is inherently subjective.

Using the Bayesian Information Criterion to establish which of the LCA models had the best statistical fit ensured that perhaps the most important aspects of this approach (that is, the number of latent household composition classes recognized as distinct) was data driven and determined by an objective statistical criterion. And since the results of the LCA models conducted on Wave 1 USoc data were very similar to those generated by LCA of Wave 2 data (see **Table 4**), we are confident that the approach we adopted successfully identified the (seven) most discrete, and therefore most important, latent household composition classes that could be described with the sociodemographic and overcrowding variables used. As for the labeling of these classes (e.g. as "Empty Nesters" or "Single Parent Household"), we offer these labels less as specific conceptual constructs per se and more as descriptive "handles" or monicker to facilitate the presentation, discussion and interpretation of the analyses.

Finally, the potential for under adjustment in the multinomial logistic regression analyses of the relationship between latent household composition class and sleep duration/quality (*i.e.* **Tables 5** and **6**) is a problem faced by any analysis of pre-existing observational data. This is because only those variables collected by the study (in this instance Wave 1 of USoc) are available for consideration/inclusion as potential confounders. Indeed, even when data are collected prospectively a decision will always have to make regarding which variables can and should be collected and which can/should not.

The importance of age, sex and social class as likely confounders in any relationship between living conditions (in this instance, household composition) and health (in this instance, sleep duration/quality) is what drove the selection of these variables (with highest educational attainment and ethnicity as proxies for the consequences/determinants of social class) as key potential confounders in these analyses. As such, we are confident that these analyses will have captured a substantial amount of the potential confounding linking latent household composition class and sleep duration/quality. Moreover, adjustment for two additional variables in these analyses (marital status and self-reported health) considered to be independent of household composition yet determinants of sleep duration/quality (*i.e.* considered "competing exposures" for sleep duration/quality), is likely to have improved the precision of any observed relationship between latent household composition class and the residual variation in sleep duration/quality. For these reasons, we are confident that the analyses described in **Tables 5** and **6** provide a robust statistical assessment of these relationships.

3.4. Implications

Notwithstanding the limitations discussed above, the findings of the present study demonstrate the importance of considering potential inter-relationships between sociodemographic characteristics and overcrowding to operationalise the full complexity of household composition in large-scale epidemiological studies, such as USoc. Moreover, the study demonstrates that latent class analysis is capable of capturing a substantial proportion of this underlying complexity, generating distinct classes of household composition that are conceptually meaningful

and statistically reliable/stable over time.

This is not only an important finding in its own right, but also has bearing on epidemiological analyses of outcomes that are likely to be household context-specific, such as a whole range of what might be considered "household lifestyle variables" including diet, activity and sleep. This is evident from the subtle yet important differences in self-reported adult sleep duration/quality observed amongst different latent household composition classes. These differences suggest that adults from household classes that are overcrowded or have children present report/experience shorter sleep durations and worse (or less good) sleep quality than those without overcrowding or children. Our study was also found some evidence that adults from household classes with both overcrowding and children (particularly two or more) reported the shortest sleep duration and worst (or least good) sleep quality—evidence that suggests the possibility of an interactive and/or summative effect of overcrowding and children on adult sleep patterns within households.

The results therefore indicate that adult sleep duration and quality are influenced by living in overcrowded homes and living with children, and suggest that not only the presence but also the number of children in the household plays a key role in the self-reported sleep of adults. For example, the "Partnered with Children" class had a higher risk of short sleep and "very bad" sleep quality than the "Partnered with Child" class–the main difference between these classes being that the former had at least one more additional child. This concurs with Chapman *et al.*'s [2] finding that the frequency of "insufficient sleep" was highest in households with 3 children. However, Groeger *et al.* [3] found that partnered adults with children had worse sleep duration and sleep quality outcomes than those adults in their study who belonged to the "Single Parent Household" class, the majority of whom (87.2%) were female.

This somewhat counterintuitive finding is perplexing, not least because women tend to report shorter and worse quality sleep than men (something we also observed, albeit incidentally, in our analysis of USoc data). Notwithstanding substantial differences in study design and context, one more straightforward possibility is that, in contrast to the sleep of unpartnered adults with children (which can only be influenced by their children and not by a non-existent adult partner), the sleep of partnered adults with children can be influenced not only by their children but also by their adult partner (and so on, by the impact of their children on their adult partner's sleep).

Clearly, further research is warranted to quantify the relative importance of overcrowding, children and the number of children on adult sleep duration and quality, not least because overcrowded households are commonly those with one or more children. This is particularly important given the growing interest in and use of household occupancy measures (such as the UK's Bedroom Standard) to determine overcrowding and welfare entitlements; and evidence linking appropriate levels of sleep duration and sleep quality as important prerequisites for mental and physical health.

ACKNOWLEDGEMENTS

REFERENCES

[1] Krueger, P.M. and Friedman, E.M. (2009) Sleep duration in the United States: A cross-sectional population-based study. *American Journal of Epidemiology*, **169**, 1052-1063. http://dx.doi.org/10.1093/aje/kwp023

[2] Chapman, D.P., Wheaton, A.G., Perry, G.S., Sturgis, S.L., Strine, T.W. and Croft, J.B. (2012) Household demographics and perceived insufficient sleep among US adults. *Journal of Community Health*, **37**, 344-349. http://dx.doi.org/10.1007/s10900-011-9451-x

[3] Groeger, J.A., Zijlstra, F.R.H. and Dijk, D.J. (2004) Sleep quantity, sleep difficulties and their perceived consequences in a representative sample of some 2000 British adults. *Journal of Sleep Research*, **13**, 359-371. http://dx.doi.org/10.1111/j.1365-2869.2004.00418.x

[4] Burgard, S.A. and Ailshire, J.A. (2013) Gender and time for sleep among US adults. *American Sociological Review*, **78**, 51-69. http://dx.doi.org/10.1177/0003122412472048

[5] University of Essex, Institute for Social and Economic Research and National Centre for Social Research (2012) Understanding society: Waves 1-2, 2009-2011 (UKDA-6614-stata12). 4th Edition, UK Data Archive SN, Colchester, 6614. http://discover.ukdataservice.ac.uk/catalogue/?sn=6614&type=Data%20catalogue

[6] Department for Communities and Local Government (2003) Housing (overcrowding) bill. HMSO, London. http://www.publications.parliament.uk/pa/cm200203/cmbills/046/2003046.pdf

[7] Law, G.R., Head, R.F. and Ellison, G.T.H. (2012) Confounding and causal path diagrams. In: Tu, Y.K. and Greenwood, D., Eds., *Modern Methods for Epidemiology*, Springer, London, Chapter 1, 1-14. http://dx.doi.org/10.1007/978-94-007-3024-3_1

[8] Textor, J., Hardt, J. and Knuppel, S. (2011) DAGitty: A graphical tool for analysing causal diagrams. *Epidemiology*, **22**, 745. http://dx.doi.org/10.1097/EDE.0b013e318225c2be

[9] Greenland, S., Pearl, J. and Robins, J.M. (1999) Causal diagrams for epidemiologic research. *Epidemiology*, **10**,

37-48.
http://dx.doi.org/10.1097/00001648-199901000-00008

[10] Pearl, J. (1993) Comment: graphical models, causality and intervention. *Statistical Science*, **8**, 266-2669.
http://dx.doi.org/10.1214/ss/1177010894

[11] Tu, Y.K., West, R., Ellison, G.T.H. and Gilthorpe, M.S. (2005) Why evidence for the fetal origins of adult disease might be a statistical artefact: The "reversal paradox" for the relation between birth weight and blood pressure in later life. *American Journal of Epidemiology*, **161**, 27-32.
http://dx.doi.org/10.1093/aje/kwi002

[12] Law, G.R. and Pascoe, S.W. (2013) *Statistical Epidemiology.* CABI, Oxfordshire.

[13] Statistical Innovations Inc. (2003) Latent gold software: Release 4.5. Belmont.

[14] Vermut, J.K. and Magidson, J. (2005) Latent GOLD 4.0 user's guide. Statistical Innovations Inc., Belmont.
http://www.statisticalinnovations.com/technicalsupport/UG.html

[15] Schwartz, G. (1978) Estimating the dimenstions of a model. *Annals of Statistics*, **6**, 461-464.
http://dx.doi.org/10.1214/aos/1176344136

[16] Vermut, J.K. and Magidson, J. (2005b) Technical guide for latent GOLD 4.0. Statistical Innovations Inc., Belmont.
http://www.statisticalinnovations.com/technicalsupport/UG.html

[17] STATACORP (2012) Stata statistical software: Release 12.1. Stata Press, College Station.

APPENDIX

DAGitty.net code to generate the DAG used to inform the analyses conducted in the present study. To view the DAG and its related minimum sufficient dataset, the text below can be pasted into the "model text data" window of www.dagitty.net:

Age 1 @-1.037, 1.268
Education 1 @-0.192, 1.236
Ethnicity 1 @-1.571, -2.098
Health 1 @1.238, 0.648
Household Class E @0.595, -0.431
Marital Status 1 @0.053, -2.079
Sex 1 @-1.554, -0.632

Sleep Duration %2F Quality O @1.420, -0.825
Age Household Class Sleep Duration %2F Quality Marital Status Education Health
Education Household Class Sleep Duration %2F Quality Marital Status Health
Ethnicity Household Class Sleep Duration %2F Quality Age Marital Status Education Health
Health SleepDuration%2FQuality
Household Class Sleep Duration %2F Quality Health
Marital Status Household Class Sleep Duration %2F Quality Health
Sex Household Class Sleep Duration %2F Quality Education Health