ELSA and Ethnicity

Detailed Roadmap of Ethnicity Data Collection in the English Longitudinal Study of Ageing (ELSA)

Scott Oatley

Introduction

Ethnic inequalities in older populations is a crucially underdeveloped area of quantitative sociological study. This is primarily because of a lack of statistical power – most social surveys simply do not have enough ethnic minority respondents to adequately analyse older (aged 50+) ethnic minority populations. ELSA collects data on individuals' broad categories of ethnic grouping (White vs Non-White)at each wave (0-11). Detailed ethnicity recording in ELSA is restricted to waves 5-7. Due to issues surrounding identification and de-anonymisation, whilst detailed ethnic group data are recorded at the survey level, this is always (in the case of data deposits), across waves deposited in the United Kingdom Data Service (UKDS) collapsed into a binary categorisation of 'White' and 'Non-White'. Detailed ethnicity variables are made available at waves 5, 6, 7, and 11[1], alongside binary categorisations, though these are not available to end users, only to the internal ELSA team. There is some difficulty in producing a single ethnicity variable that is consistent across all waves of ELSA data. This is due to different recording strategies of ethnicity[2], different naming conventions of ethnicity variables, multiple ethnicity variables within each wave, and most importantly, the fact that original ethnicity values were coded forward using the Health Survey England (HSE) data which has recently been restricted for continued use for ELSA in this manner. The purpose of this report is to provide appropriate derived measurements for ethnicity across ELSA waves to first accurately report cross-sectional and longitudinal based statistics on the viability of ethnicity as a measure in analysis. Secondarily, this report develops key findings of the viability of a detailed versus dichotomised ethnicity based variable based on feeding forward and backward related information for repeated observations. This report finds that ethnicity, both as a dichotomous variable and a more detailed version has a healthy overall frequency across most ELSA waves to allow ethnic subgroup analysis. Recording of individuals ethnic grouping is robust to item non-response across all waves of ELSA. Valid responses of ethnic grouping across each wave of ELSA stand at around 95 percent of the sample size for each wave. Regarding the more detailed ethnicity recording, this report finds that it is only at wave 4 of ELSA that ethnic minority categories report over 0.25 percent of valid responses. Before wave 4, using a detailed measure of ethnicity may present statistical power issues. Consequently, cross-sectional

[1] Wave 11 is using internal interim data. The Wave 11 data deposited in the UKDS only has the collapsed binary version of ethnicity.

[2] For the purposes of all recoding strategies, for detailed ethnicity, the primary coding strategy adopts the most recent wave categorisation of separating 'Non-White' individuals into categories of: Mixed, Asian/Asian British, Black/African/Caribbean/Black British, and Any Other Group.

analysis when using waves 0-3 of ELSA ought to use a collapsed version of ethnicity. When utilising longitudinal analysis, a detailed ethnicity variable is only viable across waves 4-11[3].

Measures of Ethnicity in ELSA:

Wave 0 of ELSA uses HSE 1998 fed information on ethnicity collapsed into 'White' and 'Non-White' groupings. Wave 1 also offers HSE fed information though also begins to collect its own dichotomised information that is repeated in waves 1-4 as the sole ethnicity variable. Wave five offers a continuation of this dichotomised variable in addition to a detailed ethnicity breakdown variable. This is repeated from waves 5-7. From waves 8-10 ELSA reverts to a simple dichotomised variable of ethnicity. Current wave 11 interim data is used that utilises both dichotomised ethnicity and a detailed breakdown. One of the key issues in the utilisation of ethnicity across ELSA is the inconsistency in variable collection, naming, and breakdown of ethnicity itself. In Table 1 information is provided on wave, ethnicity variables, and sample size.

Table 1 Ethnicity Recording in ELSA over Waves 0-11

Wave	Variable	Values	Sample Size	Observation Count[4]	Note
0	ethnic	1 = White 2 = Non-White	26,787	26,729	HSE feed data
1	aethnicr	1 = White 2 = Non-White	12,099	6,742	HSE feed data
1	fqethnr	1 = White 2 = Non-White	12,099	5,275	
2	fqethnr	1 = White 2 = Non-White	9,432	9,421	
3	fqethnr	1 = White 2 = Non-White	9,771	9,765	
4	fqethnr	1 = White 2 = Non-White	11,050	11,035	
5	fqethnr	1 = White 2 = Non-White	10,274	10,265	
5	fffqethn	1 = White 2 = Mixed ethnic group 3 = Black 4 = Black British 5 = Asian 6 = Asian British	10,274	9,981	

[3] This is of course dependent on the complete records of any produced model. It may also be the case that wave 4 becomes non-viable considering its relatively low ethnic minority reporting rates compared to other waves. That is ultimately up to the individual researcher to gauge.

		7 = Any other group		
6	Fqethnr	1 = White 2 = Non-White	10,601	10,597
6	fffqethn	1 = White 2 = Mixed ethnic group 3 = Black 4 = Black British 5 = Asian 6 = Asian British 7 = Any other group	10,601	9,350
7	Fqethnr	1 = White 2 = Non-White	9,666	9,658
7	fffqethn	1 = White 2 = Mixed ethnic group 3 = Black 4 = Black British 5 = Asian 6 = Asian British 7 = Any other group	9,666	9,123
8	fqethnmr	1 = White 2 = Non-White	8,445	8,445
9	fqethnmr	1 = White 2 = Non-White	8,736	8,728
10	fqethnmr	1 = White 2 = Non-White	7,589	7,517
11	fqethnmr	1 = White 2 = Non-White	7,844	2,339
11	fqethnm	1 = White 2 = Mixed ethnic group 3 = Asian/Asian British 4 = Black/African/Caribbean/Black British 5 = Other	7,844	2,339

[1] Sample size includes only viable categories and removes any observations that have refused to answer or responded with 'Don't Know'.

Given that a dichotomous measure of ethnicity is recorded at every wave of ELSA, that appears to be a decent place to begin any investigation. Merging all ethnicity data into long format after data cleaning results in a single ethnicity dichotomous measure with 119,774 observations, with 114,894 observations categorised as 'White' and 4,880 observations as 'Non-White' (See Table 2 for breakdown). This leaves 12,520 observations missing across waves 0-11.

Table 2 Summary Statistics for Dichotomised Ethnicity Variable

Dichotomised Ethnicity	Frequency	Percentage (%)
Non-White	4,880	4.07
White	114,894	95.93
Total	119,774	100.00

The situation becomes much more dire when attempting to construct a more detailed measure of ethnicity. Detailed ethnicity is only recorded at four waves: 5, 6, 7, and 11. Due to this a total figure results in only 30,793 observations out of the 132,294 total. Table 3 provides a breakdown the detailed ethnicity statistics for each grouping.

Table 3 Summary Statistics for Detailed Ethnicity Variable

Detailed Ethnicity	Frequency	Percentage (%)
White	29,465	95.69
Mixed Ethnic Group	102	0.33
Asian/Asian British	404	1.31
Black/African/Caribbean/Black British	645	2.09
Any Other Group	177	0.57
Total	30,793	100.00

Repeated Contacts

Across all waves of ELSA there is a total observation count of 132,294. Within that total observation count individuals have repeated observations across waves that can be identified by their unique identification code 'uniqid'. Table 4 documents the iterations of repeated instances that individuals have across waves within ELSA. Only 10.57 percent of the ELSA sample has an individual recorded at only one wave of data. This means that just shy of 90% of individuals in ELSA have viable repeated contact information. This information can be used to boost the n of fixed variables such as ethnicity[5]. The dichotomised variable uses last observation carried forward (LOCF) and detailed version of ethnicity uses LOCF, and last observation carried backward (LOCB). These mechanisms work by imputing an individual's ethnicity if it is recorded as missing at a particular wave by using information from an earlier, or later wave where that information is recorded for that individual.

[5] The fixed nature of ethnicity is somewhat debatable. This is something the report delves into further down. For now, this is taken as an assumption of ethnicity reporting.

Table 4 Total Number of Repeated Contacts Within ELSA

Copies	Observations
1	13,983
2	8,760
3	7,812
4	7,100
5	7,805
6	8,880
7	9,030
8	9,744
9	14,796
10	15,480
11	9,548
12	19,356

For the dichotomised variable of ethnicity this results in a total n of 132,128. This results in only 166 missing cases. Results are displayed by wave as seen in Table 5. The detailed breakdown of ethnicity also sees a large boost in size, with a new total n of 66,101. This is still far short of the total population. The observation count can be boosted by using the dichotomised ethnicity variable, 'White' is a category both variables share, as such, if values are missing for the detailed ethnicity variable but are available for the dichotomised variable, the values are imputed. This takes the n for the detailed variable from 66,101 to 129,404 as seen in Table 6.

Table 5 Summary Statistics by Wave for Dichotomised Ethnicity

	ELSA wave												
	0	1	2	3	4	5	6	7	8	9	10	11	Total
N	26,787	12,099	9,432	9,771	11,050	10,274	10,601	9,666	8,445	8,736	7,589	7,844	132,294
	(20.2%)	(9.1%)	(7.1%)	(7.4%)	(8.4%)	(7.8%)	(8.0%)	(7.3%)	(6.4%)	(6.6%)	(5.7%)	(5.9%)	(100.0%)
Ethnicity													
Non-White	1,131	364	229	305	380	353	414	385	317	430	462	611	5,381
	(4.2%)	(3.0%)	(2.4%)	(3.1%)	(3.4%)	(3.4%)	(3.9%)	(4.0%)	(3.8%)	(4.9%)	(6.1%)	(7.8%)	(4.1%)
White	25,598	11,718	9,198	9,463	10,662	9,916	10,184	9,275	8,128	8,298	7,098	7,209	126,747
	(95.8%)	(97.0%)	(97.6%)	(96.9%)	(96.6%)	(96.6%)	(96.1%)	(96.0%)	(96.2%)	(95.1%)	(93.9%)	(92.2%)	(95.9%)

Table 6 Summary Statistics by Wave for Detailed Ethnicity

	ELSA wave												
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Total**
N	26,787 (20.2%)	12,099 (9.1%)	9,432 (7.1%)	9,771 (7.4%)	11,050 (8.4%)	10,274 (7.8%)	10,601 (8.0%)	9,666 (7.3%)	8,445 (6.4%)	8,736 (6.6%)	7,589 (5.7%)	7,844 (5.9%)	132,294 (100.0%)
Ethnicity													
White	25,598 (100.0%)	11,718 (99.9%)	9,198 (99.9%)	9,463 (99.7%)	10,667 (97.1%)	9,916 (96.6%)	10,184 (96.4%)	9,275 (96.4%)	8,128 (96.6%)	8,298 (96.8%)	7,113 (98.0%)	7,217 (94.3%)	126,775 (98.0%)
Mixed Ethnic Group	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (0.0%)	23 (0.2%)	25 (0.2%)	27 (0.3%)	25 (0.3%)	22 (0.3%)	21 (0.2%)	11 (0.2%)	37 (0.5%)	194 (0.1%)
Asian/Asian British	0 (0.0%)	6 (0.1%)	4 (0.0%)	4 (0.0%)	83 (0.8%)	88 (0.9%)	102 (1.0%)	92 (1.0%)	74 (0.9%)	71 (0.8%)	29 (0.4%)	165 (2.2%)	718 (0.6%)
Black/African/Caribbean/Black British	1 (0.0%)	2 (0.0%)	2 (0.0%)	19 (0.2%)	168 (1.5%)	186 (1.8%)	204 (1.9%)	187 (1.9%)	157 (1.9%)	148 (1.7%)	88 (1.2%)	178 (2.3%)	1,340 (1.0%)
Any Other Group	1 (0.0%)	0 (0.0%)	2 (0.0%)	3 (0.0%)	48 (0.4%)	49 (0.5%)	51 (0.5%)	46 (0.5%)	37 (0.4%)	35 (0.4%)	20 (0.3%)	54 (0.7%)	346 (0.3%)

Ethnicity Switching Across Waves – Robustness Check

The above recoding is based upon the assumption that ethnicity is fixed. Fixed in the sense that someone's ethnicity cannot change over time. This is used to justify using someone's answer from an earlier, or indeed later wave, and using that to impute missing values for the same person across their life course. Of course, if ethnicity is not fixed, and people do fluidly change their ethnicity through recorded answers in social surveys, this makes the above approach a difficult one to justify. To investigate this possibility, 'ethnic identity switchers' were identified and assessed for both the dichotomised and detailed ethnicity breakdown.

Of those that have recorded observations for the dichotomised ethnicity variable, out of the 132,128 observations, there are 566 observations that are recorded as 'identity switchers'. This amounts to 74 individuals across all ELSA waves that have instances whereby they change their ethnicity either from 'White' to 'Non-White' or vice versa. With regards to the detailed breakdown of ethnicity, out of 129,373 observations, there are 268 observations that are recorded as 'identity switchers', of those, 108 observations are counted as switching from a 'White' to 'Non-white' or vice versa category. The remaining 160 observations are individuals that switched their ethnicity from within 'Non-White' categories, for example, from Any Other Group to Black/African/Caribbean/Black British. This 268 observation count amounts to 43 individuals over the course of all ELSA waves changing their ethnicity in the detailed categorisation of ethnicity.

Whilst not strictly zero, the level of cases comparative to the total sample size across ELSA waves provides some robust evidence that those in the ELSA population consider their ethnicity to be a fixed concept that very rarely changes across waves of data.

Ethnicity Recording by Dichotomised and Detailed Measurements

Two figures are produced for each measure of ethnicity. The first is based on the dichotomised variable of ethnicity that includes valid observations across waves 1-11 of ELSA[6]. For the dichotomised variable, a dual-y axis line graph is provided which includes percentage of valid observations for each waves total N and a count line. A similar figure is produced for the detailed ethnicity variable, though the number of categories for the detailed breakdown of ethnicity prevents a combined graph, this results in three graphs produced – one dual-y axis graph for the 'White' category and then a separate count based and percentage based graph for the ethnic minority categories.

Figure 1 covers the trends of the dichotomised ethnicity variable within ELSA. The 'White' and 'Non-White' categories are relatively stable across ELSA waves, though there is a noticeable uptick in the percentage of 'Non-White' compared to 'White' population in ELSA from waves 8 onwards, especially with regards to the wave 10-11 reported difference in ethnic minority participation, both in terms of observation count and percentage. This is most likely a product of concerted efforts on behalf of the ELSA team to boost ethnic minority numbers within the social survey. Wave 10 saw a 'refreshment' cohort aged 50-93 introduced into ELSA that represented younger age groups as well as those from ethnic minority

[6] Wave zero is not included in this figure because it has over 25,000 respondents as it is based from HSE data, and not primarily ELSA data. This would result in a rather incomprehensible looking graph.

backgrounds[7] (Lloyd *et al.*, 2024). This sample boosted also came with a revised ethnicity question that matched current UK census questions on ethnicity. At wave 11, whilst still only making up around just shy of 10 percent of the sample, it is the highest level that ELSA has ever recorded.

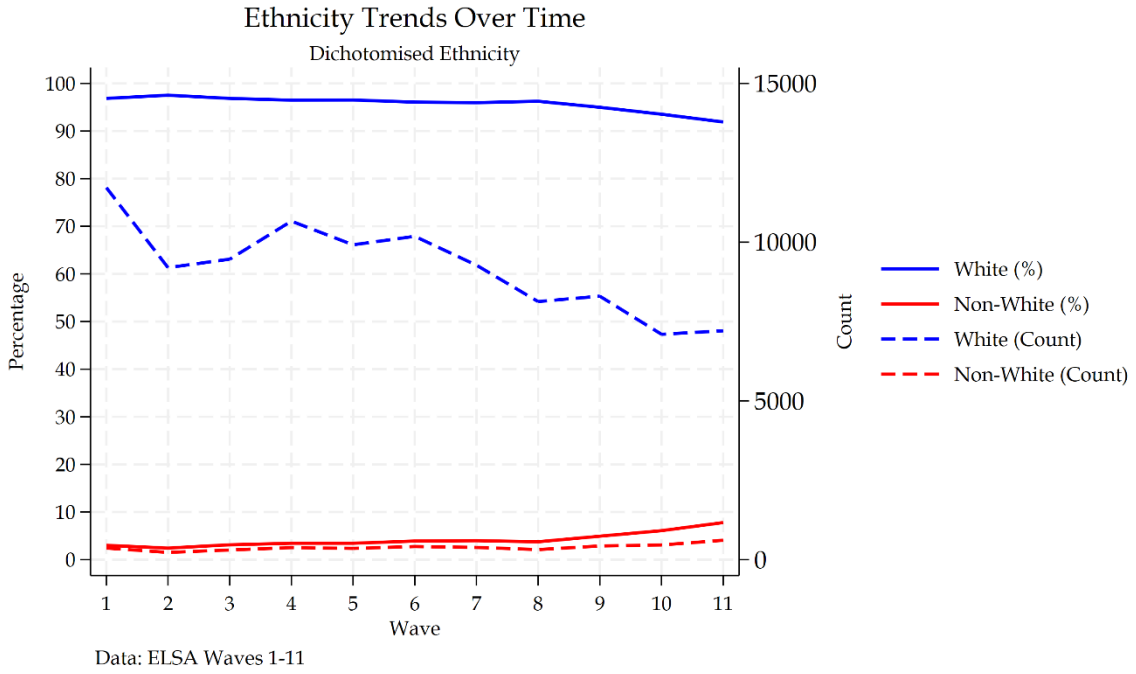

Figure 1 Breakdown of Dichotomised Ethnicity by ELSA Wave

Figure 2 presents a dual-y axis graph of the percentage and count of each waves valid 'White' respondents. There are several peaks and troughs that indicate several attempts within ELSA's timeframe of boosting recruitment and including refresher samples. Whilst the 'White' population within ELSA has appeared to decline along a similar pattern to the overall attrition line, the 'Non-White' population has remained relatively stable across waves, and has even increased as a total percentage of the overall ethnicity population. The result is that much of the attrition across ELSA waves is based on 'White' participant non-response.

[7] Wave 10 requirements were based on individuals being born between 28 February 1971 and 28 February 1967 and any ethnicity or born before 28 February 1957 and non-White

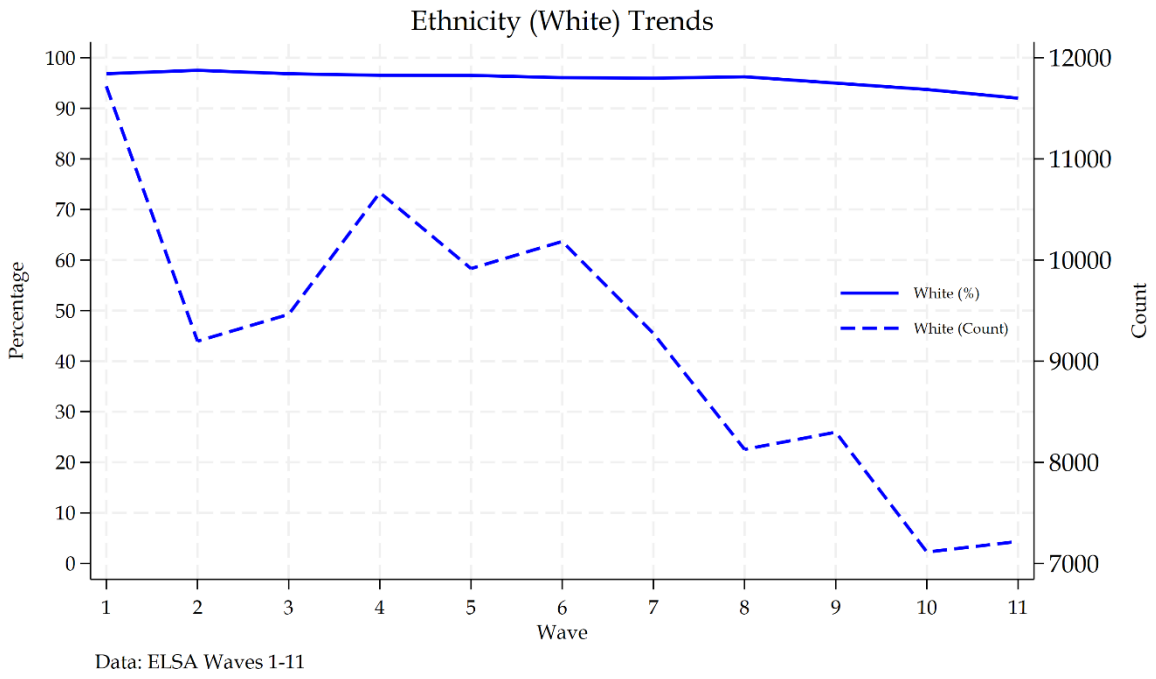

Figure 2 Breakdown of Detailed Ethnicity (White) by ELSA Wave

Figure 3 presents the same data in terms of percentage and count valid responses but this time for ethnic minority respondents. This figure echoes earlier reports that a detailed measure of ethnicity is not possible to utilise before wave four of the survey. The observations as a percentage of responses to the detailed ethnicity question using this derived approach are less than 0.25 percent for all ethnic minority categories until wave four. Even at wave four it may be difficult to conduct some analysis with specific ethnic minority categories such as 'Mixed' that still only has around 0.25 percent. Both 'Asian/Asian British' and 'Black/African/Caribbean/Black British' category groups continue to grow from wave four to wave six and then enter a decline, with all ethnic minority groups seeing a sharp decline from wave 9 to wave 10.

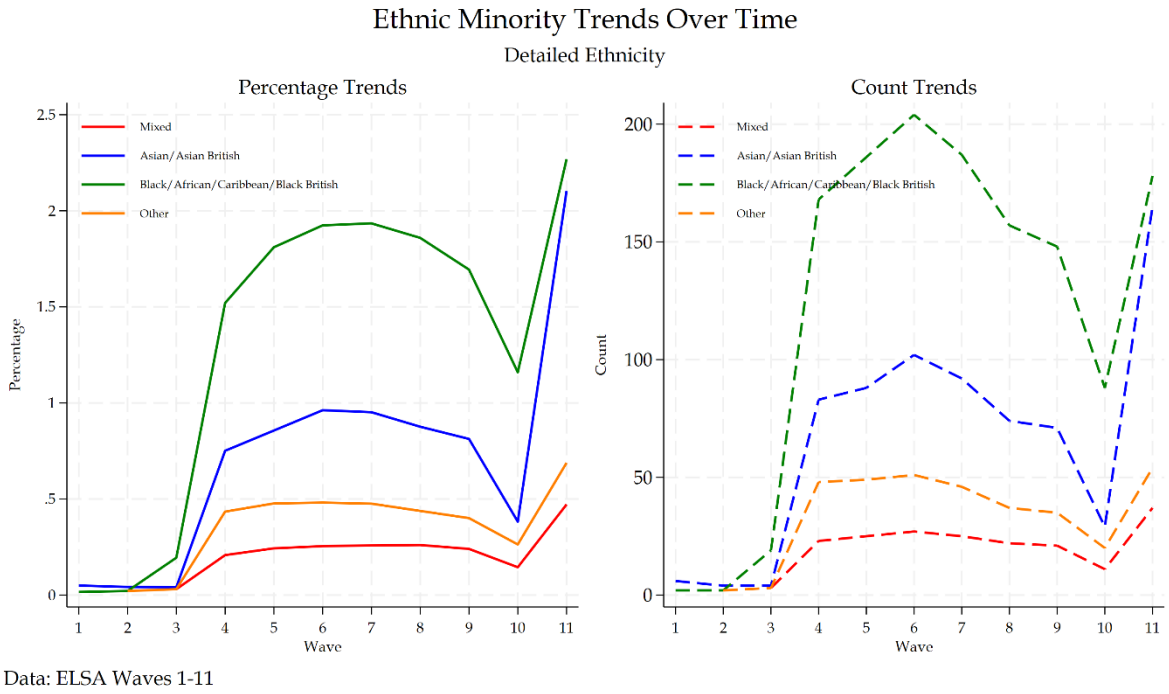

Figure 3 Breakdown of Detailed Ethnic Minorities by ELSA Wave

The wave 10-11 breakdown demonstrates a remarkable increase in ethnic minority observations within ELSA with 'Mixed' jumping from below 0.25 percent to just below 0.5 percent, 'Asian/Asian British' jumping from below 0.5 percent to just below 2 percent, 'Black/African/Caribbean/Black British' moving from around 1.25 percent to around 2.25 percent, and 'Other' moving from around 0.25 percent to just below 0.75 percent. Comparing this to the England and Wales 2021 National Census for over 65 year olds 3.8 percent identified in the Asian, Asian British, Asian Welsh ethnic group, and 1.4% in the Black, Black British, Black Welsh, Caribbean, or African ethnic group. The remaining 1.2% of people aged 65 years and over identified in the Mixed or Multiple ethnic groups and Other ethnic groups (Wood, Standeven and Gwynn, 2023).

Wave 11 of ELSA is the single largest proportion of ethnic minority respondents in ELSA's history. Though importantly, when looking at raw count data, it is not the single largest wave for ethnic minorities responses, that would be wave six of ELSA. Waves six or 11 are prime waves for any cross-sectional analysis of detailed ethnic minority populations.

Conclusions

ELSA records ethnicity at every single wave of its survey but there have traditionally been issues and questions surrounding the utility of such a measure both in a cross-sectional and longitudinal aspect. This report has documented a process of derived ethnicity in the form of two measures: a dichotomised 'White' and 'Non-White' variable, and a more granular, detailed ethnicity variable. Both variables have been derived from multiple ethnicity waves of data and given the assumption that ethnicity is a fixed characteristic have used methods such as LOCF and LOCB to impute missing values within individuals that have recorded values at one point across ELSA. This has accomplished the task of presenting ethnicity within ELSA

across the entire wave-design of the survey. Through this, the viability of ethnicity as a measurement of analysis has been assessed.

Whilst in most cases a more detailed ethnicity variable would be desirable, that is only viable from around wave four onwards, where ethnic minority populations are all recorded above 0.25 percent of the total recorded ethnicity population within each wave. This means that from a longitudinal analysis standpoint, if a researcher required a detailed version of ethnicity, their analysis would have to be restricted to at a maximum of using waves 4-11. If the researcher was intending to use cross-sectional data and the time of collection was of little importance, wave 11 would be our suggestion to use. Wave 11 has the single largest amount of recorded ethnic minority population compared to the total observable ethnicity recorded out of all ELSA waves. If the desire is to conduct a longitudinal piece of research and there is ability to compromise on ethnicity detail, this report has also demonstrated that the dichotomised ethnicity variable that has been derived, has near 100 percent response rates as a form of total observations divided by total cohort sample for each wave.

In relation to attrition over time, both derived variables offer total observation counts that are near 100 percent of sample wave. Importantly, attrition appears to only affect the 'White' population within ELSA, with the 'Non-White' population seeing an albeit slow, but steady increase as total percentage of ethnicity recorded. A similar, albeit slightly more complex story is presented for the detailed ethnicity measure.

Overall, this report has presented two derived measures of ethnicity: one simplified, another detailed for researchers use and discretion. Both have relatively high sample sizes compared to the wave total. Whilst the dichotomised measure can be used for cross-sectional or longitudinal analysis across any wave(s), the detailed measure is restricted to analysis within waves 4-11 of ELSA. Robustness checks were conducted to assess the viability in the assumption that ethnicity is a fixed construct in individuals life courses. Such robustness checks appear to have confirmed this assumption – with a handful of individual cases that do appear to 'switch' ethnicities across their life course. The extent to which these cases are a matter of measurement error or true 'switching' will unfortunately never truly be known, though it could equally be the case that the way we record ethnicity for a specific sub-group of individuals, may be inadequate in expressing their ethnic identity. Given the relatively tiny number of 'switchers', whilst this may be an interesting avenue of research in the future, for ELSA it appears that ethnicity can be treated as fixed which justifies our derivation practices.

Reference List

Lloyd, L., *et al.* (2024) ‘The dynamics of ageing: The 2021/2022 English Longitudinal Study of Ageing (Wave 10) Technical Report’, *National Centre for Social Research* [Preprint].

Wood, S., Standeven, C. and Gwynn, E. (2023) ‘Ethnic group by age and sex, England and Wales’, *ONS* [Preprint].