Understanding Society
Working Paper Series

No. 2026 – 04

March 2026

Effects on Attrition of Ethnic and Immigrant Groups in Understanding Society

Gianfranco Addario
National Centre for Social Research

Understanding Society

Economic and Social Research Council

Non-technical summary

Longitudinal studies are important for policy, social, and academic research, as their data enable in-depth understanding of social issues through the life trajectories of its sample members. Yet, these studies are also vulnerable to attrition, which is the loss of participants from the study sample over time. This represents a significant source of bias in longitudinal studies, because individuals who remain in a study sample tend to differ from those who leave it, affecting the robustness of the inferences that can be drawn from the study's data, as well as the sample sizes for analysis.

Sample members with an ethnic minority or immigrant background are particularly likely to attrit. While we know that individuals in these groups are more likely to experience risk factors related to attrition – such as being renters or living in urban areas – little is known about how such effects might interact to enhance probabilities of attrition. This research paper filled this gap, exploring whether and how interactions between different risk factors contribute to attrition by ethnic or immigration backgrounds of sample members recruited in the first wave of *Understanding Society*.

The first research question of this paper sought to understand whether predictors, or risk factors, of attrition have a different association with the risk of attrition for different ethnic or migration groups. For example, does the effect of the type of housing tenure on attrition vary by ethnic group? Overall, the paper found evidence of some variation, especially for area type (urban or rural) and level of formal education. In other instances, the differences between groups were, on average, very small.

The second and third research questions shifted the attention to how the combination of different predictors shapes the risk of attrition, and how likely sample members with an ethnic minority or immigration background are to experience these specific combinations of risk factors. The analysis focused on six variables (age, type of economic activity, having dependent children, level of formal education, type of housing tenure, and area type) and revealed that different combinations of economic activity, age group, and having dependent children were associated with either an increase or a decrease in the likelihood of attrition. This suggests that there is a further layer of attrition risk that cannot be understood by being young, or in economic activity, or childless, but is linked to having all three characteristics.

The analysis also found that sample members with Bangladeshi, other Asian, mixed, or 2nd Generation backgrounds were more likely to experience combinations of risk factors that would increase the likelihood of attrition, while sample members with white British, Irish, or Caribbean identities were more likely to experience combinations of risk factors that would mitigate it.

This paper demonstrates that attrition among sample members with ethnic minority or immigrant backgrounds is strongly driven not only by the effects of risk factors, but also by their complex interactions. Once we account for these effects, the likelihood of leaving the study sample shows minimal to no variations by ethnic or migration background. Due to these dynamics, fieldwork and design solutions seeking to mitigate attrition in these groups should primarily focus on the effects of risk factors and of their complex interactions.

Effects on Attrition of Ethnic and Immigrant Groups in Understanding Society

Gianfranco Addario, National Centre for Social Research

Abstract: Longitudinal surveys are a cornerstone of social research, shaping key policy decisions. Yet, over time, certain groups are far more likely to drop out, raising concerns about representation and analysis bias. Attrition is especially high among respondents with ethnic minority and immigrant backgrounds – but why? Are they just more exposed to common risk factors, or do specific combinations of demographic and socio-economic characteristics contribute to increasing attrition rates in these groups?

This paper tackles this question using data from *Understanding Society*, the UK's largest longitudinal survey. First, the paper examines whether well-known predictors of attrition – such as type of housing tenure, form of economic activity, and level of education – operate differently across ethnic and immigration groups. While some differences emerge, in many instances, their effect size is small.

Second, the paper employs MAIHDA (Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy) to partition the attrition risk into additive (mean predictors' effects) and multiplicative (effects of predictors' interactions). The findings show strong multiplicative effects, primarily driven by the interaction among age, economic activity, and having dependent children.

The analysis demonstrates that the distribution of these multiplicative effects varies strongly by ethnic and migration backgrounds, disproportionately increasing the likelihood of attrition of second-generation immigrants and sample members from Bangladeshi, mixed, and other Asian backgrounds.

These findings carry two significant contributions for survey research. First, they demonstrate that the associations between ethnic minority and immigration backgrounds and attrition are largely explained by the effects of nonresponse predictors and by their complex interactions. Second, it proposes MAIHDA as a method for the study of survey nonresponse as it can offer important insights for statisticians and survey researchers seeking to understand the structure and the role of multiplicative effects on attrition in their studies.

Keywords: Ethnic minority groups, attrition, nonresponse

Acknowledgements: I want to thank Andrew Bell and George Leckie for their suggestions and feedback on the analysis, and especially on the application of the MAIHDA method. I also want to express my gratitude to Martin Wood at the National Centre for Social Research and to the Understanding Society team, and especially to Alita Nandi, Pablo Cabrera Alvarez, Peter Lynn, Olena Kaminska, and Lucinda Platt, for providing constructive comments and suggestions that helped me shape this paper.

Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by the National Centre for Social Research (NatCen) and Verian . The research data are distributed by the UK Data Service.

Corresponding author: Gianfranco Addario, Gianfranco.Addario@citystgeorges.ac.uk

Introduction

Over the past 50 years, survey unit nonresponse – the decision of a sampled individual not to take part in a study after being invited – has gradually increased (Atrostic et al., 2001; De Leeuw et al., 2002; Luiten et al., 2020). The consequent decline in response rates poses a significant challenge to studies based on random probability sampling. The statistical theories behind these studies assume full sample cooperation; the higher the nonresponse, the more we part away from these statistical assumptions, resulting in data with unknown and unknowable forms of bias.

Survey unit nonresponse introduces bias into survey estimates because it typically does not occur at random: individuals who choose not to participate tend to be different from those who do. While this bias can be tolerated with relatively low levels of nonresponse, it becomes a significant issue when the nonresponse is high. With high nonresponse, there is likely and systematic non-participation from certain groups, resulting in study findings that no longer represent the target population.

Besides survey unit nonresponse, longitudinal studies, such as *Understanding Society*, are also vulnerable to sample members withdrawing from, or leaving, the study. This phenomenon is known as attrition. These two types of nonresponse are very similar in their dynamics and tend to be associated with the same sample members' demographic and socioeconomic characteristics. Indeed, survey unit nonresponse is often a key predictor of attrition in later waves (Loosveldt et al., 2002).

Consequently, they also have a similar biasing effect on the survey, introducing bias in a study's findings. However, the biasing effect of attrition is considerably greater than that of survey unit nonresponse. This because it represents a form of nonresponse that accumulates over time (Watson and Wooden, 2009), as a consequence of multiple one-off decisions to withdraw from the study sample after the initial recruitment, resulting in greater volumes of survey nonresponse. While data collected at the initial recruitment can be used to mitigate bias in a study's estimates, the main effect of attrition is the reduction of the analysis sample, leading to a loss of analytical power, and therefore larger standard errors and greater uncertainty around the point estimates.

The scale of attrition on the sample integrity of Understanding Society can be seen in the findings of Cabrera and Lynn (2023). Their results indicate that only 35.5% of the 43,673 members from the General Population Sample (GPS) recruited in Wave 1 participated in Wave 13. The authors found this proportion varied dramatically among subgroups of the population. Most notably, it differed by ethnic background: 36.9% of the GPS sample members with a white background participated in Wave 13, compared to 17.1% of those with a black background and 18.3% of those with a Bangladeshi background.

This high likelihood of leaving the study for immigrants and sample members with an ethnic minority background has been the object of concern for people engaging in academic and policy research using survey data (see, for example, Pyper, 2021; Font and Mendez, 2013; Kappelhof, 2017). The initial sample size at recruitment for these

groups – often already small as they represent a minority of the population – tends to decrease at a faster rate than it does for other groups. Consequently, the surviving sample members tend to be very different from those who left the study sample and from the target population they are meant to represent. The effect on analysis can be particularly severe: findings may not correctly represent a subgroup of the population with high attrition, leading to potentially biased evidence for policy action, and the analysis may fail to capture with sufficient confidence some effects, due to the loss of sample size.

The aim of this research paper is to enhance the understanding of attrition for immigrants and people with an ethnic minority background in the UK, using data from *Understanding Society*. The overarching research objective stemmed from two considerations. First, while we know these groups tend to have a higher likelihood of experiencing nonresponse risk factors (see Table 23 and Table 24 in the appendix), we do not know much about whether these factors have different effects in the different groups. They could have a different effect on people who do not have an immigration background or who belong to the white British ethnic group compared to the effect that they have on other groups. This consideration is condensed in the first research question of this paper:

RQ1. Does the risk of attrition predicted by demographic and socio-economic characteristics vary by either ethnic or immigration background?

Second, we do not know much about how these risk factors interact and whether experiencing specific combinations of them can significantly increase the risk of attrition. In other words, are people with an ethnic minority background or immigrants more likely to leave a study sample only because they experience a greater accumulation of risk factors, or do they also experience combinations of risk factors that create additional vulnerabilities to attrition? For example, if being a young private renter in an urban setting increases the risk of nonresponse, is this profile more common among sample members with a white British background, or among those with an ethnic minority background?

This second consideration is broken down into two further research questions, the first one seeking to establish the existence of multiplicative effects between risk factors to attrition, and the second to assess whether these effects are more or less common among individuals with an ethnic minority or immigration background:

RQ2. Do interactions of different demographic and socioeconomic characteristics increase the risk of attrition?
RQ3. Are sample members with either an ethnic minority or an immigration background more likely to suffer from an increased risk of attrition emerging from the interaction of different demographic and socio-economic characteristics?

The approach adopted to address these research questions is exploratory, with no specific hypotheses being formulated. This is due to the fact that, with some exceptions (see Feskens et al., 2007), the current literature has concentrated on the mean effect of

risk factors towards attrition (Watson and Wooden, 2009; Deng et al., 2013; Olsen, 2005; Behr et al., 2005; Kappelhof, 2017; 2015) providing little prior evidence on the understanding of interaction effects influencing attrition for individuals from ethnic minority or immigration backgrounds. Investigating this area is the first contribution of this paper.

A second contribution is utilising a recently developed analysis method that is not yet established in the field of survey methodology. The Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA), originally proposed by Evans (2015) and Evans et al. (2018), enables the exploration of multiplicative effects as they empirically emerge without the need to pre-define the types of interaction effects anticipated in the data. Although it is not entirely organic – the analyst must determine which variables are likely to have specific effects in combination – this method is particularly beneficial in this context, where any interaction effect relating to a set of risk factors could be expected to influence attrition, and taking a judgment call on what to include or exclude as interaction terms in a model can easily appear arbitrary. The main advantage of this method over a standard two-way interaction term used in linear regressions is that it allows for the estimation of the effect for each combination of all the predictors included in the model (which in this analysis amounts to 166 observed combinations across six variables).

Data

Analysis sample

The analysis is based on sample members of *Understanding Society* (also known as *UK Household Longitudinal Study*, UKHLS), the largest longitudinal survey in the UK, designed to represent the household population of the country. The study design and the sample size make Understanding Society data ideal for research on the participation of individuals with ethnic minority identities in longitudinal studies.

Individuals with an ethnic minority and an immigration background represent a relatively small yet extremely heterogeneous segment of the UK population, making large sample sizes for these subgroups essential to understanding the differences that emerge both between and within the various groups. In addition to a large sample size, Understanding Society also includes an oversampling of individuals with an ethnic minority background, the *UKHLS Ethnic Minority Boost*, carried out during the first wave of the study in 2009-10 (Lynn, 2009), alongside two general population samples (one for Great Britain and one for Northern Ireland), which also contain individuals with an ethnic minority background.

The design of Understanding Society is also highly consistent over time, allowing the capture of the necessary variables for the analysis from different study waves. Moreover, sample members were selected according to random probability sampling principles, both for the general population samples and for the ethnic minority boost, enhancing the generalisability of the findings.

Besides these considerations on data suitability for analysis, Understanding Society – the longitudinal study with the largest ethnic minority sample in the UK – is expected to be one of the beneficiaries of this research project. Therefore, using data from Understanding Society also facilitates the direct application of the research findings.

The cases included in this analysis are original sample members of Understanding Society who were enumerated in the study's first wave. They are all classified as "Original Sample Members: (OSM) and belong to the samples UKHLS Great Britain 2009-10, UKLHS Northern Ireland 2009-10, or UKHLS Ethnic Minority Boost 2009-10.

The analysis sample excluded individuals who did not participate in any wave of Understanding Society (the non-random missingness on key variables was too high), those classified as rising 16 (they had different experiences as sample members), and those who passed away while included in the study sample (attrition due to permanent ineligibility).

Sample members who became ineligible after leaving the UK were retained in the analysis sample because, unlike those who passed away, they could potentially rejoin the study sample if they returned to their original household in the UK. Thus, their ineligibility is not necessarily permanent. However, checks were conducted to identify the number of cases in the analysis sample who left the study and were classified by the interviewers as having left the UK. The checks confirmed that only a handful of individuals would have fallen into this group.

The resulting subset of the data included 51,173 sample members.

Analysis outcome

The outcome of interest was having left the study sample by wave 13 – or, in other words, attrition during the first 13 waves of the study. Sample members who ceased to be enumerated – i.e. listed in the sample issued to fieldwork – before wave 13 (variable lwenum_dv), as well as those who were enumerated in wave 13 but whose outcome was "Ineligible for individual interview as non-respondent in last three waves," were considered to have left the sample. Of the 51,173 sample members included in this analysis, 33,941 had left the study sample by wave 13 (66.3%), while 17,232 were still part of the sample (33.7%).

Ethnic background

The two key variables of interest in the analysis were ethnic background and immigration background. For ethnicity, the harmonised variable prepared by the Understanding Society research team (ethn_dv) was used in the analysis. This variable is derived using the most recent value reported across several sources, including the self-reported ethnicity, ethnicity reported by a household member, and the ethnicity of the parents, reducing the volume of missing values.

Table 1 – Association between ethnic backgrounds and attrition

Ethnic background	n	Left study sample	Chi-square p-value

White British	37,445	63%	<0.001
African	1,613	85%	
Any other	990	82%	
Bangladeshi	1,433	77%	
Caribbean	1,255	77%	
Indian	2,269	72%	
Irish	763	64%	
Mixed	989	72%	
Other Asian	1,110	78%	
Other white	1,396	74%	
Pakistani	1,835	74%	

Immigration background

The immigration background was computed using variables that captured whether the respondent was born in the UK, the age at which they moved to the UK, whether their parents were born in the UK, and whether they had both parents present during their upbringing. The design of the variable tended to focus on the effects of socialisation, either in a family or school setting. In the computed variable, later generations tend to have had socialisation that was carried out primarily in the UK setting, while earlier generations are likely to have had socialisation in non-UK settings.

The focus on socialisation is important because it determines paths of social action, expected behaviours, and attitudes (Cuff et al., 2005); for example, taking part in a survey may be acceptable in some social groups, but less so in others. These play a role in the heuristic decision-making conceptual framework that Groves and Couper (1998) and Groves et al. (2000) developed to understand survey nonresponse.

The categories included in the immigration background variable are:

- **Generation 1**: not born in the UK from both parents not born in the UK and moved to the UK when aged 18 or older (i.e. did not complete formal education in the UK).
- **Generation 1.5**: same as Generation 1 but moved to the UK before age 18 (i.e. completed at least some formal education in the UK).
- **Generation 2**: born in the UK from both parents not born in the UK.
- **Generation 2.5**: born in the UK from either (but not both) parent born in the UK (i.e., differently from Generation 2, had in childhood at least one parent born in the UK).
- **Generation 3/3+**: Born in the UK from both parents born in the UK, or not born in the UK from both parents born in the UK but moved to the UK at or before age 5 (i.e. started formal education in the UK). For simplicity, this is presented and discussed as “Generation 3” in this research paper.

For sample members who grew up without one parent and did not provide information about the country of birth of both parents, the derivation of the variable only considered the information from the parent they were living with at age 14.

Table 2 – Association between immigration backgrounds and attrition

Immigration background	n	Left study sample	Chi-square p-value
Generation 1	6,534	78%	<0.001
Generation 1.5	2,133	75%	
Generation 2	3,045	70%	
Generation 2.5	2,323	63%	
Generation 3+	33,680	61%	

Risk factors for attrition

The socio-economic and demographic characteristics associated with survey nonresponse are defined in this research paper as "risk factors". The risk factors were selected considering the two overarching forms that nonresponse is found to take: non-contact and refusal (Lepkowski and Couper, 2002). Refusal – the decision of a sampled individual to not take part in a study – is generally assessed through the framework proposed by Groves and Couper (1998) and Groves et al. (2000), which outlines different decision mechanisms and cost opportunities that people weigh when they need to decide whether to take part in a study. Non-contact – the failure of the research team to establish and maintain contact with a sample member – is usually understood through structural elements, such as the living conditions and settings. The different sample members' characteristics selected in the analysis tend to be strong predictors of one – if not both – forms of nonresponse (Watson and Wooden, 2009; Lepkowski and Couper, 2002; Jenkins, 1997; Gray et al., 1996, Fitzgerald et al., 1998).

The analysis includes two categories of risk factors for attrition. The first set includes factors that were considered fixed in time, meaning that they were not expected to change over time for the same individual. This category includes sex and year of birth. The latter was used to compute the respondents' age in 2010, corresponding to the mid-fieldwork year for the first wave of the study.

The second set includes factors that could vary over time, meaning that the same individual may have different measures in the different waves of Understanding Society. As these factors are expected to influence nonresponse when present, the decision was to take the closest observation to the nonresponse event, meaning the latest known value. This strategy was pursued for housing tenure, whether living in an urban or rural setting, the highest level of formal education achieved, having dependent children, and employment activity.

Speaking English as a second language represents a fixed-in-time feature, but the data were collected in various waves of Understanding Society, with some individuals giving different answers between the different waves. Whenever discrepancies in the

responses were found, the approach was to classify the sample member as speaking English as a second language.

Table 3 - Association between demographics and socio-economic characteristics and attrition (row percentages)

Characteristic	n	Left study sample	Chi-square p-value
Latest known area			<0.001
Rural area	11,037	60%	
Urban area	40,136	68%	
Latest known tenure			<0.001
Owned or other	34,249	60%	
Private renter furnished or employer	3,112	94%	
Private renter unfurnished	4,466	73%	
Social renter	9,292	76%	
Whether English is first language			<0.001
English first language	39,941	62%	
English not first language	6,900	76%	
Formal education			<0.001
Below degree	32,068	71%	
Degree or higher	18,984	58%	
Age group			<0.001
Under 35	16,395	74%	
Age 30-65	27,472	59%	
Over 65	7,305	75%	
Whether had dependent children			<0.001
Yes	15,497	69%	
No	35,676	65%	
Sex			<0.001
Male	23,634	67%	
Female	27,537	65%	
Latest known economic activity			<0.001
Economically Active	33,148	69%	
Economically Inactive	18,016	61%	

Missing values

The dataset used for the analysis had a relatively low level of missingness. Missing values were less than 1% for all analysis variables, with the exception of immigration background and English as a second language, which had missing values in 7% and 8% of the cases in the analysis sample, respectively. As expected, missingness in these variables was correlated with the outcome of interest: cases who left the study sample were more likely to have missing values. Figure 3 and Figure 4 in the appendix include a check on the distribution of missing values.

The approach for the analysis was to proceed with listwise deletion in the presence of missing values. This means that cases with missing values in the analysis variables were excluded.

Weights

The analysis was conducted on unweighted data because the objective was to draw conclusions specific to the Understanding Society sample. While many of the findings could be applied to other studies with designs comparable to Understanding Society, it is important to remember that the analysis sample included the UKHLS Ethnic Minorities sample boost (achieved through oversampling urban areas with a high proportion of residents with an ethnic minority background; see Lynn, 2009) alongside general population samples. For this reason, some results, such as the relation between ethnic background and area type, may be influenced by the inclusion of the sample boost in the analysis.

Analysis and results

Research question 1

The first research question aimed to determine whether the effect of demographic and socio-economic characteristics on the likelihood of attrition from the study sample (leaving the study) in the first 13 waves of Understanding Society varies by ethnic or immigration background.

There are large differences in sample sizes between individuals who identify as part of the ethnic majority (white British) and those who report having an ethnic minority background, as well as between those without a migration background (Generation 3) and those with a migration background. This suggests that our understanding of these risk factors is predominantly influenced by the larger groups, potentially obscuring the dynamics within the smaller subgroups of the populations relevant to this research.

This question was answered using binary logistic models. A model was built for each risk factor, including the risk factor, either the ethnic background or the migration background variables, and an interaction term between these two variables. The results of the model were explored through first and second differences in their average marginal effects (Mize, 2019; Arel-Bundock et al., 2024). First differences explore the average slope effect within each group, and whether it is significantly different from zero (for example, in Table 4, it refers to the effect of living in an urban area over a rural one, within each ethnic or migration group). Second differences report the variations between the effect size in the group and the effect size for white British or Generation 3 (reference categories), and test whether the differences between the two groups are statistically significant.

All the models returned the interaction term predictor to be a statistically significant improvement of the model. While this was expected, given the large sample used in the analysis, only a few differences between groups seem to reveal a consistent trend.

Furthermore, even if there was significant empirical evidence for some differences, the effect size was often particularly small.

The model outputs are included in the appendix (see Table 25 through Table 40). The next few subsections focus on the model predictions and the average marginal effects for each of area type (urban or rural), level of formal education, type of housing tenure, having dependent children, and type of economic activity with ethnic group or immigration status.

The outputs by sex, age, and speaking English as a second language are not discussed, as the differences were particularly small, and it was difficult to identify a clear pattern. The reason why the effect sizes for speaking English as a second language were particularly small may lie in the longitudinal nature of the sample. The cases considered in the analysis are sample members who have already accepted recruitment to the study, meaning they have completed a wave-one survey interview with sufficient language skills or have found adequate language support. The small effect may also be attributable to the level of missing values and patterns of missingness (likely to be not at random) in the analysis variable. In the Great Britain general population sample 8% of the cases missed language information, 7% spoke English as a second language, and 85% spoke it as their first language.

Urban or rural settings

First differences indicate that, across all the ethnic backgrounds, including white British, living in an urban area was associated with an increase of the likelihood of attrition. However, these differences were not significant for some groups (African, Bangladeshi, Caribbean, Indian), probably due to the low number of sample members with these backgrounds living in rural areas. Second differences indicate that the urban effect on attrition is greater for sample members with an ethnic minority background than it is for white British. However, also in this case, the difference is not always statistically significant.

These findings are mirrored when examining how the effect of living in an urban area on attrition varies by migration background. Across all migration groups, living in an urban area increases the likelihood of attrition. The second differences indicates that this effect is larger for sample members with a Generation 1 and Generation 1.5 background, compared to Generation 3.

Table 4 – Predicted probabilities of attrition and average marginal effects of area type and ethnic or migration background

	Predicted probabilities		First differences	Second differences
Ethnic background	**Rural area**	**Urban area**	**Urban effect[1]**	**Urban effect[2]**
white British	0.60	0.64	0.04*	Reference
African	0.64	0.86	0.21	0.17
Any other	0.65	0.83	0.18*	0.15*
Bangladeshi	0.67	0.77	0.10	0.06
Caribbean	0.54	0.77	0.23	0.19

Indian	0.59	0.73	0.14	0.10
Irish	0.60	0.67	0.08*	0.04
Mixed	0.47	0.73	0.26*	0.22*
Other Asian	0.54	0.79	0.25*	0.21*
Other white	0.64	0.76	0.11*	0.08*
Pakistani	0.20	0.74	0.54*	0.50*
Migration background	**Rural area**	**Urban area**	**Urban effect[1]**	**Urban effect[2]**
Generation 3	0.58	0.62	0.04*	Reference
Generation 2.5	0.57	0.64	0.07*	0.02
Generation 2	0.59	0.70	0.11*	0.06
Generation 1.5	0.55	0.75	0.20*	0.16*
Generation 1	0.62	0.79	0.16*	0.12*

[1] Difference in the predicted probabilities (urban areas minus rural areas) in the group.
[2] Difference between the urban effect in the group and the urban effect in the reference category.
* p-value < 0.05 (null hypothesis: the effect is zero).

Level of formal education

The association between education level and attrition varies significantly across ethnic backgrounds. First differences show that – on average – higher levels of formal education decrease the likelihood of attrition. The only group where formal education is not associated with a decrease in attrition is sample members with a Bangladeshi background; however, the reduction in attrition for higher levels of formal education is also relatively small for Pakistani, Indian, and other Asian groups.

Second differences indicate that the effect of education on lowering attrition was greater for sample members with white British, Irish, other white or mixed backgrounds. The other ethnic backgrounds had a greater likelihood of attrition at higher levels of education, compared to white British.

A similar variation can be observed by migration background. Across all groups, the probability of attrition for those with a level of education equivalent to a university degree or higher was lower than that for those with no formal qualification. However, higher levels of education were more strongly correlated with attrition across Generations 1, 1.5, and 2, compared to Generations 2.5 and 3.

Table 5 – Predicted probabilities of attrition and average marginal effects by level of formal education and ethnic or migration background

	Predicted probabilities			**First differences**		**Second differences**	
Ethnic background	**No qual.**	**Below degree**	**Degree or higher**	**Below degree effect[1]**	**Degree or higher effect[1]**	**Below degree effect[2]**	**Degree or higher effect[2]**
white British	0.79	0.66	0.51	-0.13*	-0.28*	Reference	Reference
African	0.94	0.89	0.82	-0.05*	-0.12*	0.08*	0.16*
Any other	0.92	0.84	0.77	-0.08*	-0.15*	0.05	0.13*
Bangladeshi	0.77	0.77	0.75	0.00	-0.02	0.13*	0.26*
Caribbean	0.89	0.78	0.69	-0.10*	-0.20*	0.03	0.08*
Indian	0.79	0.74	0.70	-0.05	-0.09*	0.08*	0.19*

Irish	0.74	0.69	0.54	-0.05	-0.19*	0.08	0.09
Mixed	0.87	0.79	0.62	-0.08	-0.24*	0.05	0.04
Other Asian	0.84	0.82	0.75	-0.02	-0.09*	0.11*	0.19*
Other white	0.91	0.80	0.68	-0.11*	-0.23*	0.02	0.05
Pakistani	0.78	0.72	0.72	-0.05*	-0.06*	0.07*	0.23*
Migration background	**No qual.**	**Below degree**	**Degree or higher**	**Below degree effect[1]**	**Degree or higher effect[1]**	**Below degree effect[2]**	**Degree or higher effect[2]**
Generation 3	0.78	0.64	0.49	-0.14*	-0.29*	Reference	Reference
Generation 2.5	0.78	0.68	0.53	-0.10*	-0.25*	0.03	0.04
Generation 2	0.80	0.75	0.64	-0.05	-0.16*	0.09*	0.13*
Generation 1.5	0.76	0.78	0.69	0.02	-0.07*	0.16*	0.22*
Generation 1	0.83	0.78	0.76	-0.05*	-0.07*	0.09*	0.22*

[1] Difference in the predicted probabilities (below degree minus no qualification/degree or higher minus no qualification) in the group.
[2] Difference between the formal education category effect in the group and the formal education category effect in the reference category.
* p-value < 0.05 (null hypothesis: the effect is zero).

Housing Tenure

First differences indicate that across all the ethnic backgrounds, renting a furnished accommodation was associated with a greater likelihood of attrition. However, the effect of being a social renter or renting an unfurnished accommodation on attrition was not different from the effect of owning the dwelling across some ethnic groups, such as Bangladeshi or Caribbean. Second differences show that the housing tenure effect on attrition was stronger among white British compared to other groups. The only exception is represented by sample members with an other white background; in this group, being a renter – regardless of the type – was more strongly linked to attrition compared to white British.

The effect of housing tenure also showed some variations by migration background. In particular, second differences indicate that the effect of privately renting a furnished accommodation had a milder effect on attrition for sample members with Generations 1, 1.5, and 2, compared to the effect it had on Generation 3.

Table 6 – Predicted probabilities of attrition and average marginal effects of housing tenure and ethnic or migration background

	Predicted probabilities				**First differences**			**Second differences**		
Ethnic background	**Own or other**	**Furn. renter**	**Unfurn. renter**	**Social renter**	**Furn. renter effect[1]**	**Unfurn. renter effect[1]**	**Social renter effect[1]**	**Furn. renter effect[2]**	**Unfurn. renter effect[2]**	**Social renter effect[2]**
white British	0.58	0.93	0.71	0.74	0.35*	0.13*	0.16*	Ref	Ref	Ref
African	0.76	0.93	0.89	0.87	0.17*	0.13*	0.11*	-0.18*	0.00	-0.05
Any other	0.72	0.96	0.82	0.89	0.24*	0.10*	0.17*	-0.11*	-0.03	0.01
Bangladeshi	0.73	0.85	0.78	0.77	0.12*	0.05	0.04	-0.23*	-0.07	-0.11*
Caribbean	0.74	0.93	0.76	0.77	0.19*	0.02	0.03	-0.17*	-0.11*	-0.13*
Indian	0.66	0.97	0.85	0.81	0.31*	0.18*	0.14*	-0.05*	0.06	-0.02
Irish	0.58	0.97	0.72	0.77	0.39*	0.14*	0.19*	0.04	0.01	0.03

Mixed	0.65	0.98	0.66	0.76	0.33*	0.01	0.12*	-0.02	-0.11*	-0.04
Other Asian	0.68	0.96	0.79	0.81	0.28*	0.11*	0.13*	-0.07*	-0.02	-0.03
Other white	0.57	0.98	0.80	0.82	0.41*	0.23*	0.25*	0.05*	0.10*	0.09*
Pakistani	0.71	0.90	0.79	0.74	0.19*	0.09*	0.03	-0.16*	-0.04	-0.13*
Migration background	**Own or other**	**Furn. renter**	**Unfurn. renter**	**Social renter**	**Furn. renter effect[1]**	**Unfurn. renter effect[1]**	**Social renter effect[1]**	**Furn. renter effect[2]**	**Unfurn. renter effect[2]**	**Social renter effect[2]**
Generation 3	0.56	0.93	0.69	0.73	0.37*	0.13*	0.17*	Ref	Ref	Ref
Generation 2.5	0.56	0.96	0.65	0.76	0.40*	0.09*	0.20*	0.03	-0.04	0.03
Generation 2	0.66	0.87	0.74	0.74	0.21*	0.08*	0.09*	-0.15*	-0.05	-0.08*
Generation 1.5	0.66	0.95	0.84	0.83	0.29*	0.18*	0.17*	-0.08*	0.05	0.00
Generation 1	0.67	0.95	0.84	0.82	0.28*	0.17*	0.15*	-0.09*	0.04*	-0.02

[1] Difference in the predicted probabilities (private renter furnished minus owner or other/private renter unifurnished minus owner or other/social renter minus owner or other) in the group.
[2] Difference between the housing tenure category effect in the group and the housing tenure category effect in the reference category.
* p-value < 0.05 (null hypothesis: the effect is zero).

Dependent children

The effect of having dependent children on attrition also showed some substantial variations between ethnic groups. If for the ethnic majority, white British, having dependent children was associated with a slightly higher probability of attrition, the association was reversed for sample members with Bangladeshi or Indian ethnic backgrounds. For these groups, there was a higher predicted probability of attrition among those without dependent children compared to those with dependent children. Turning to second differences, on average, the effect of having dependent children on attrition was much greater for white British than it was for most of the other ethnic backgrounds, such as Bangladeshi, Pakistani, Indian, Caribbean, other Asian, and other white.

The association between having dependent children and attrition was also reversed among sample members with a Generation 2 background. In this group, those with dependent children had a lower predicted probability of attrition, 0.65, compared to those without dependent children, 0.72.

Table 7 – Predicted probabilities of attrition and average marginal effects of having dependent children and ethnic or migration background

	Predicted probabilities		**First differences**	**Second differences**
Ethnic background	**No**	**Yes**	**Child effect[1]**	**Child effect[2]**
white British	0.62	0.66	0.04*	Reference
African	0.84	0.87	0.02	-0.02
Any other	0.80	0.85	0.04	0.00
Bangladeshi	0.79	0.74	-0.05*	-0.09*
Caribbean	0.77	0.75	-0.02	-0.06*
Indian	0.75	0.67	-0.08*	-0.13*
Irish	0.62	0.69	0.06	0.02
Mixed	0.72	0.71	-0.01	-0.06

Other Asian	0.80	0.76	-0.04	-0.08*
Other white	0.75	0.72	-0.03	-0.07*
Pakistani	0.75	0.72	-0.02	-0.06*
Migration background	**No**	**Yes**	**Child effect[1]**	**Child effect[2]**
Generation 3	0.60	0.65	0.05*	Reference
Generation 2.5	0.62	0.64	0.02	-0.03
Generation 2	0.72	0.65	-0.07*	-0.11*
Generation 1.5	0.74	0.76	0.02	-0.03
Generation 1	0.78	0.78	0.00	-0.05*

[1] Difference in the predicted probabilities (not having children minus having children) in the group.
[2] Difference between the child effect in the group and the child effect in the reference category.
* p-value < 0.05 (null hypothesis: the effect is zero).

Economic activity

This research conceptualises economic activity according to the definition of the Labour Force Survey (ONS, 2024), with unemployed individuals considered economically active.

The association between economic activity and attrition in most ethnic groups was either not significant or in line with survey literature, which considers economic activity to be mildly associated with attrition due to cost-opportunity considerations (Watson and Wooden, 2009). However, the association seems reversed for sample members with a Bangladeshi background, where economic inactivity is more strongly associated with attrition. Overall, second differences indicate that economically inactive sample members with an African, Caribbean, Bangladeshi or Pakistani background were more likely to incur attrition compared to economically inactive white British.

Some differences could also be observed across migration backgrounds. The probability of leaving the study was significantly lower for economically inactive sample members across all the migration backgrounds, with the exception of Generation 1. In this group, the rates of attrition were not different between economically inactive and active sample members.

Table 8 – Predicted probabilities of attrition and average marginal effects of economic activity and ethnic or migration background

	Predicted probabilities		**First differences**	**Second differences**
Ethnic background	**Active**	**Inactive**	**Inactive effect[1]**	**Inactive effect[2]**
white British	0.66	0.58	-0.08*	Reference
African	0.85	0.85	0.00	0.07*
Any other	0.83	0.80	-0.03	0.04
Bangladeshi	0.75	0.81	0.06*	0.13*
Caribbean	0.77	0.77	0.00	0.08*
Indian	0.73	0.69	-0.04	0.03
Irish	0.68	0.59	-0.09*	-0.02
Mixed	0.73	0.66	-0.06	0.01
Other Asian	0.80	0.72	-0.08*	0.00

Other white	0.76	0.69	-0.06*	0.01
Pakistani	0.73	0.75	0.03	0.10*
Migration background	**Active**	**Inactive**	**Inactive effect[1]**	**Inactive effect[2]**
Generation 3	0.64	0.57	-0.06*	Reference
Generation 2.5	0.65	0.56	-0.09*	-0.03
Generation 2	0.70	0.65	-0.06*	0.00
Generation 1.5	0.78	0.66	-0.12*	-0.06*
Generation 1	0.78	0.77	-0.01	0.05*

[1] Difference in the predicted probabilities (being economically inactive minus being economically active) in the group.
[2] Difference between the inactive effect in the group and the inactive effect in the reference category.
* p-value < 0.05 (null hypothesis: the effect is zero).

Research question 2

The second research question was answered using an application of multilevel modelling called Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy, or MAIHDA (Evans, 2015; Evans et al., 2018).

The underlying theory behind using this model is that each individual occupies a specific social position, defined by their combination of demographics and socio-economic characteristics. Each social position has a corresponding level of power, which places individuals in a position of relative advantage or disadvantage towards the outcome being analysed. At a societal level, individuals who share the same social position have similar power relations with the outcome; thus, they experience comparable advantages or disadvantages. From a statistical perspective, the model considers respondents clustered in social positions.

The multilevel model proposed in this analysis has two levels. The effect of a predictor – or risk factor – on the outcome of interest is estimated at the first level. The predictor's coefficient represents the mean risk: this is the average effect that the specific predictor has on the outcome across the entire analysis sample. However, the mean risk is a central estimate; many cases would have a risk that deviates from the mean risk, resulting in positive or negative residuals.

In MAIHDA, these residuals are expected to correlate with the individual's social position. This implies that, for certain combinations of demographic and socio-economic characteristics, the residuals are likely to be consistently positive, signalling an increased risk for individuals in that social position, or consistently negative, indicating a decreased risk. The residuals for each observed social position are represented in the second level of the model. These residuals illustrate the multiplicative risk, or the level of advantage and disadvantage towards a specific outcome that is proper of a particular social position, above and beyond the mean risk across the analysis sample.

Model design

Six of the eight risk factors used in this research paper were included in the model: age, level of formal education, housing tenure, whether the individual has dependent children, whether the individual lives in an urban setting, and economic activity. These six risk factors were combined, resulting in 166 different strata – or mutually exclusive combinations of risk factors – in the analysis sample.

Sex and speaking English as a second language were excluded from the model due to the need to limit the proportion of strata with a small number of observations. The decision to remove sex came from the fact that its association with the likelihood of attrition was not as strong as that of other risk factors. However, this consideration does not exclude that sex could interact with other variables (this was tested as a robustness check using a separate MAIHDA model, as discussed later in this paper). English second language was removed because it seemed to be a relatively weak predictor for attrition, but also due to the large volume of missing values in the variable.

The number of observations across the 166 strata in the sample ranged from 1 to 3,512, with six strata containing only one observation, 80% of the strata having 10 or more observations, and 70% having 20 or more. MAIHDA is found to tolerate a small proportion of strata with small sample sizes (Evans et al., 2024), as the estimates for these strata are mitigated by residual shrinkage (the estimates are conservatively brought closer to the mean of all the strata, reducing the influence of outliers from strata with a small number of observations). However, exceeding the number of strata with too few observations would have posed challenges for the model's convergence, as the model would not have been able to detect variance in the binary outcome for a large proportion of strata.

Strata inequity and residual variance

The analysis started from a null MAIHDA model, which included only the intercept for both the fixed-effect and random-effect components, with clustering by strata. This provides a summary of the "overall inequity in the sample" (Evans et al., 2024: 6), revealing the extent of variation in the outcome (attrition) that can be attributed to the complex interactions in strata. The Variance Partition Coefficient (VPC) indicated that the variance at level 2 was 0.23, obtained in logistic models by treating the level-1 variance as a fixed constant (3.29), following the method suggested by Snijders and Bosker (2012). This indicated that 23% of the likelihood of attrition can be attributed to the strata formed by the six variables included in the model.

The model was then expanded to include fixed-effect predictors at level 1 (see Table 9 and Table 10). The model returned a strata-level variance of 0.36, which corresponds to a VPC of 10%. Within the MAIHDA interpretation, this suggests that after controlling for the mean effect of the various risk factors in the fixed-effect component of the model, the residual variance between the strata, attributed to interactions among these six variables, was 10%. The strata (level-2) variance between the two models decreased by 64%, as indicated by the Proportional Change in Variance (PCV) coefficient. This shows that the mean effect of the predictors accounted for 64% of the overall inequity between strata towards leaving the study sample.

Table 9 – MAIHDA null and main effect model – intercept random effects

Model	N obs	N groups	Std. Dev.	Variance	VPC	PCV
Null model	50,996	166	1.00	0.99	0.23	-
Main effect model	50,996	166	0.60	0.36	0.10	0.64

Table 10 – MAIHDA Null and Main effect model – fixed effects

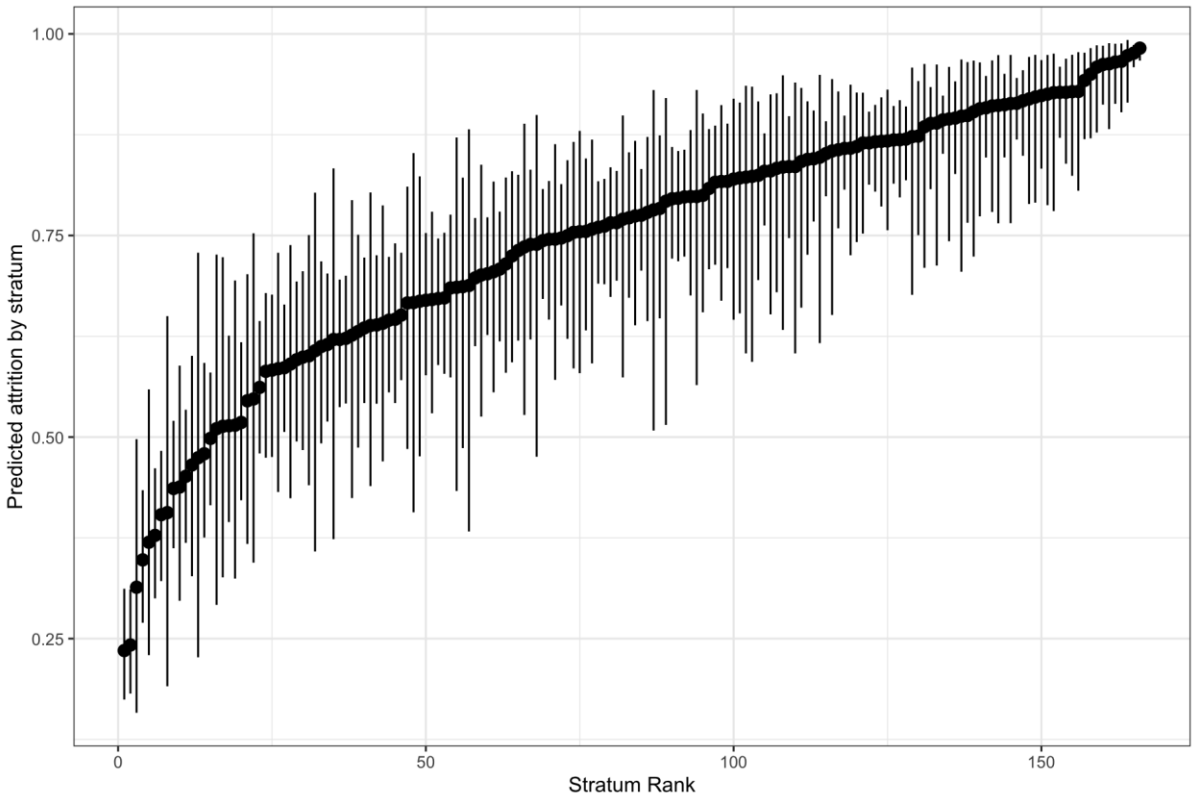
NULL MODEL			
Terms	**Log odds**	**Std. Error**	**p-value**
Intercept	1.19	0.09	<0.001
MAIN EFFECT MODEL			
Terms	**Log odds**	**Std. Error**	**p-value**
Intercept	-0.51	0.17	<0.001
Age group (ref: 35-65)			
Under 35	0.25	0.12	0.04
Over 65	0.96	0.19	<0.001
Formal education (ref: degree of higher)			
Below degree	0.67	0.12	<0.001
Tenure (ref: owner)			
Private renter furnished or from employer	1.80	0.18	<0.001
Private renter unfurnished	0.39	0.15	0.01
Social renter	0.45	0.15	<0.001
Whether living in urban setting (ref: rural area)			
Urban area	0.38	0.12	<0.001
Whether has dependent children (ref: no)			
Yes	0.40	0.12	<0.001
Economic activity status (ref: Inactive)			
Active	0.23	0.12	0.05

Total risk

In the MAIHDA framework, the fixed effects predict the *additive effect* or *mean risk* (i.e., the risk of attrition associated with the additive presence of given risk factors), while the random intercepts of the different strata (where each stratum is an observed combination of risk factors) predict the *interaction effect* or *multiplicative risk* (i.e. the residual risk not covered in the mean risk but explained by the strata). The sum of mean risk and multiplicative risk determines the *total risk*, emerging from the combination of mean and multiplicative risks.

The total risk predicted by the main effect model varied strongly between the different strata (Figure 1), ranging from 0.24 for the strata with the lowest total risk to 0.98 for the strata with the highest total risk.

Figure 1 – Total risk in the main effect model by stratum rank



The 10 strata with the highest total risk included primarily sample members aged over 65 or below 35, with formal educational attainment below a university degree, renting privately furnished accommodation or renting from their employer, and living in an urban area. The 10 strata with the lowest total risk were mostly composed of people aged 35-65, with a level of education equal to or above a degree, owning their accommodation, and living in a rural area. Interestingly, both groups largely included people without dependent children and who were economically inactive.

Table 11 – Overview of strata with the highest and lowest total risk

Top 10 strata with the highest total risk

Age group	Education	Tenure	Area type	Dependent children	Economic activity	Predicted probability	Obs. In Stratum (n)
Under 35	degree or higher	Private renter (furnished/employer)	Urban	Yes	Inactive	0.94	40
Under 35	below degree	Private renter (furnished/employer)	Urban	No	Inactive	0.95	20
Over 65	below degree	Private renter (furnished/employer)	Rural	No	Inactive	0.96	17
Under 35	below degree	Private renter (furnished/employer)	Urban	Yes	Inactive	0.96	73
Over 65	below degree	Private renter (furnished/employer)	Rural	No	Active	0.96	7
35-65	below degree	Private renter (furnished/employer)	urban	Yes	Inactive	0.97	49
Over 65	below degree	Private renter (furnished/employer)	Urban	No	Inactive	0.97	35
Over 65	below degree	Private renter (furnished/employer)	Urban	No	Active	0.97	4
Under 35	below degree	Private renter (furnished/employer)	Urban	No	Active	0.98	681
Under 35	degree or higher	Private renter (furnished/employer)	Urban	No	Active	0.98	851

Top 10 strata with the lowest total risk

Age group	Education	Tenure	Area type	Dependent children	Economic activity	Predicted probability	Obs. In Stratum (n)
35-65	degree or higher	Owned or other	Rural	No	Inactive	0.24	736
35-65	degree or higher	Owned or other	Urban	No	Inactive	0.24	1436
35-65	degree or higher	Social renter	Rural	No	Inactive	0.31	26
Under 35	degree or higher	Owned or other	Rural	Yes	Active	0.35	391
35-65	degree or higher	Private renter unfurnished	Rural	No	Inactive	0.37	27
35-65	below degree	Owned or other	Rural	No	Inactive	0.38	1028
Under 35	degree or higher	Owned or other	Urban	Yes	Active	0.40	1389
Under 35	degree or higher	Owned or other	Rural	No	Inactive	0.40	6
35-65	below degree	Owned or other	Urban	No	Inactive	0.44	2631
35-65	degree or higher	Private renter unfurnished	Urban	No	Inactive	0.44	60

Multiplicative risk

The multiplicative risk, represented by the stratum-level random intercept effect, has negative values in some strata and positive values in others, as shown in Figure 2. This indicates that the multiplicative effects can also be mitigative: strata with negative residuals include sample members who are less likely to leave the study sample, relative to what is predicted by the mean risk emerging from the independent variables added to the model.

Figure 2 also highlights that the multiplicative effects were statistically significant (different from zero) only for strata at the two extremes of the distribution. This is expected in a MAIHDA output, and the common approach is to review the composition of the strata at either end of the multiplicative effects distribution to determine which variables are likely to be responsible for these effects (Evans et al., 2024).

The review of the ten strata with the highest and the ten with the lowest multiplicative effects highlights two key considerations. Firstly, sample members who were both economically active and under the age of 35 were likely to be found in strata with the highest and lowest multiplicative risks. The discriminant element between these two groups appeared to be whether they had dependent children (marked in light green in Table 12):

- High risk: age under 35, economically active, no dependent children.
- Low risk: age under 35, economically active, with dependent children.

Similarly, sample members, both aged 35 to 65 and economically inactive, were likely to be found in strata with either the highest or the lowest multiplicative risk. Also in this instance, the discriminant for the type of multiplicative risk – whether high or low – was having dependent children (marked in light blue in Table 12):

- High risk: age 35-65, economically inactive, with dependent children.
- Low risk: age 35-65, economically inactive, no dependent children.

These findings seem to provide empirical evidence of a multiplicative risk effect between economic activity, having dependent children, and age, in determining attrition.

Figure 2 – log-odds of random-effect intercept residuals by stratum rank (multiplicative risk)

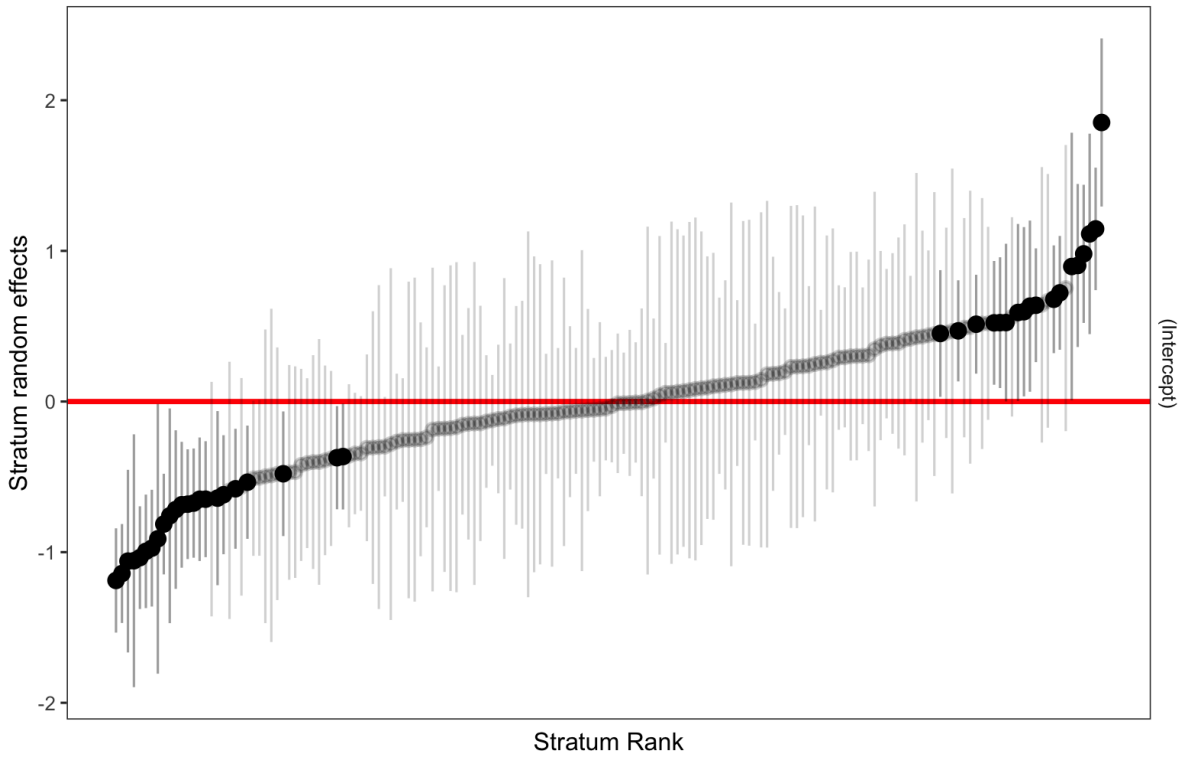


Table 12 – Overview of strata with the highest and lowest multiplicative risk

Top 10 strata with the highest multiplicative risk

Age group	Education	Tenure	Area type	Dependent children	Economic activity	Sig. (p<0.05)[1]	Predicted multiplicative risk	Observations in Stratum (n)
Under 35	degree or higher	Private renter furnished or employer	urban area	No	Economically Active	TRUE	0.87	851
Under 35	below degree	Owned or other	rural area	No	Economically Active	TRUE	0.76	548
age 35-65	degree or higher	Social renter	urban area	Yes	Economically Inactive	TRUE	0.75	58
age 35-65	below degree	Social renter	urban area	Yes	Economically Inactive	TRUE	0.73	347
Under 35	below degree	Private renter furnished or employer	urban area	No	Economically Active	TRUE	0.71	681
Over 65	degree or higher	Owned or other	rural area	No	Economically Active	FALSE	0.70	19
age 35-65	degree or higher	Private renter unfurnished	rural area	Yes	Economically Inactive	FALSE	0.68	8
Under 35	degree or higher	Owned or other	rural area	No	Economically Active	TRUE	0.68	357
Under 35	below degree	Owned or other	urban area	No	Economically Active	TRUE	0.67	2080
age 35-65	below degree	Private renter unfurnished	rural area	Yes	Economically Inactive	FALSE	0.66	18

Top 10 strata with lowest multiplicative risk

Age group	Education	Tenure	Area type	Dependent children	Economic activity	Sig. (p<0.05)[1]	Predicted multiplicative risk	Observations in Stratum (n)
Under 35	below degree	Owned or other	urban area	Yes	Economically Active	TRUE	0.24	1058
Under 35	degree or higher	Owned or other	urban area	Yes	Economically Active	TRUE	0.24	1389
Under 35	below degree	Private renter furnished or employer	urban area	Yes	Economically Active	TRUE	0.26	133
Under 35	degree or higher	Private renter furnished or employer	rural area	Yes	Economically Active	TRUE	0.26	17
age 35-65	degree or higher	Owned or other	urban area	No	Economically Inactive	TRUE	0.27	1436
Under 35	degree or higher	Owned or other	rural area	Yes	Economically Active	TRUE	0.27	391
Under 35	below degree	Owned or other	rural area	Yes	Economically Active	TRUE	0.27	283
age 35-65	below degree	Private renter furnished or employer	rural area	Yes	Economically Active	FALSE	0.30	16
age 35-65	below degree	Owned or other	urban area	No	Economically Inactive	TRUE	0.31	2631
age 35-65	degree or higher	Social renter	rural area	No	Economically Inactive	FALSE	0.33	26

[1]Null hypothesis for significance testing: the strata residual is not different from zero.

Testing for the multiplicative effect with an interaction term

The multiplicative effect of age, dependent children and economic activity was tested by fitting a three-way interaction term to a binary logistic model. The model was first tested with a Type 3 ANOVA to assess whether the predictor, as a whole, significantly improves the model (Table 13). The test found strong evidence in support of a model with this interaction term.

Table 13 – Analysis of Deviance table (type III test) for a model with a three-way interaction term

Term	LR Chisq	DF	p-value
Age group	1699.21	2	< 0.001
Level of formal education	944.15	1	< 0.001
Tenure	1598.51	3	< 0.001
Area type	51.11	1	< 0.001
Dependent children	676.09	1	< 0.001
Economic activity	650.14	1	< 0.001
Age group : Economic activity	1.15	2	0.564
Dependent children : Economic activity	188.58	1	< 0.001
Age group : Dependent children	101.61	2	< 0.001
Age group : Dependent children : Economic activity	9.33	2	< 0.01

The model output (Table 14) revealed a particularly strong effect of age, dependent children, and economic activity on attrition. This would have been missed in a model without interaction terms.

Table 14 – Outputs of model 1 (mean effect) and model 2 (three-way interaction terms)

Terms	Model 1 (Mean effect model)			Model 2 (Model with interaction terms)		
	Log odds	Std. Error	p-value	Log odds	Std. Error	p-value
(Intercept)	-0.87	0.03	<0.001	-1.11	0.03	<0.001
Age group (ref: 35-65)						
Over 65	1.21	0.03	<0.001	1.46	0.04	<0.001
Under 35	0.35	0.02	<0.001	1.04	0.12	<0.001
Level of formal education (ref: degree and above)						
Below degree	0.64	0.02	<0.001	0.65	0.02	<0.001
Tenure (ref: Onwer or other)						
Private renter furnished or employer	2.27	0.08	<0.001	2.10	0.08	<0.001
Private renter unfurnished	0.51	0.04	<0.001	0.48	0.04	<0.001
Social renter	0.65	0.03	<0.001	0.64	0.03	<0.001
Area type (ref: rural)						
Urban	0.21	0.02	<0.001	0.17	0.02	<0.001

Dependent children (ref: No)						
Yes	0.21	0.02	<0.001	1.80	0.08	<0.001
Economic activity (ref: Inactive)						
Active	0.67	0.02	<0.001	0.79	0.03	<0.001
Interaction terms						
Age over 65 : Economically Active				0.20	0.19	0.30
Age under 35 : Economically Active				0.02	0.12	0.84
Dep children Yes : Economically Active				-1.10	0.08	<0.001
Age over 65 : Dep child Yes				9.42	77.09	0.90
Age under 35 : Dep child Yes				-1.50	0.15	<0.001
Age over 65 : Dep child Yes : Economically Active				-11.46	77.10	0.88
Age under 35 : Dep child Yes : Economically Active				-0.40	0.16	0.01
N observations	50996					

Table 15 summarises changes in the predicted probabilities. In the cases presented, the predicted probabilities from the model with interaction terms (model 2) show strong differences from those from the model based on mean effects only (model 1).

Table 15 – Predicted probability for model 1 and model 2 on key groups

Age group	Education	Tenure	Area type	Dependent children	Economic activity	Prediction (model 1)	Prediction (model 2)
Under 35	degree or higher	Owned or other	Urban	No	Active	0.59	0.71
				Yes		0.64	0.43
35-65				No	Inactive	0.34	0.28
				Yes		0.39	0.70

Robustness check for the multiplicative effect

The analysis of the combination of risk factors revealed that multiplicative effects arise from the interaction of age group, having dependent children, and economic activity. However, these effects could be influenced by sex, a variable that may capture the variance explained by the combinations of these risk factors, but which was omitted from the model due to the need to avoid too many strata with small sample sizes.

A second MAIHDA model was fit to the data to understand if including sex in the model could influence the multiplicative risk and its determinants. In order to achieve a sufficient strata sample size, the housing tenure variable was reduced to a binary measure (owner or other vs any renters) to accommodate new strata combinations by sex. The resulting model had 167 strata, with 27 strata presenting 10 or fewer observations.

The resulting model identified that 10% of the model variance could be explained by multiplicative effects (Table 16), confirming the same estimates of the previous model without accounting for sex (Table 9). Differences between the two models were found in the sample inequity (the VPC in the null model); the sample inequity in the model with sex (18%) was smaller than the inequity explained by the model without sex (23%). This is a likely consequence of the recoding of the housing tenure variable into a binary measure. Housing tenure – in particular renting privately furnished or from an employer – is one of the most powerful predictors of attrition; reducing this term to a binary measure led to a loss of information that reduced the explanatory power of the combinations of risk factors.

Table 16 – MAIHDA (incl. sex) null and main effect model – intercept random effects

Model	N obs	N groups	Std. Dev.	Variance	VPC	PCV
Null model	50,994	167	0.85	0.85	0.18	-
Main effect model	50,994	167	0.59	0.35	0.10	0.52

The composition of the strata associated with high or low multiplicative risk also appeared unchanged in the new model, confirming the robustness of the findings in the model without the term sex. However, the inclusion of the sex variable has offered further understanding of the multiplicative risk:

- Across the strata with high multiplicative risk, the combination of age 35-65, with dependent children, and being economically inactive was consistently encountered among women.
- In the low multiplicative risk strata, the combination of age 35-65, no dependent children, and being economically inactive was also associated with being men.

The other combinations of risk factors (age below 35, economically active, either with or without dependent children) were encountered across both men and women.

Research question 3

This last research question seeks to extend the learning from MAIHDA by exploring the association between ethnic or migration background and the multiplicative risk predicted by different strata. The analysis began by examining the proportion of sample members who fell into strata associated with different levels of multiplicative risk, based on their ethnic or migration backgrounds. The descriptive findings were then tested in an expanded MAIHDA model.

The distribution of ethnic and migration backgrounds across the different risk strata

The strata were divided into quintiles based on their estimation of attrition in log odds, as shown in Table 17 and Table 18. The first quintile included strata with the strongest form of mitigative effects – or negative multiplicative risk – which effectively reduces the mean risk of leaving the study sample. The last quintile included strata with the highest

multiplicative risk, which increases the risk of leaving the study sample above and beyond the mean risk.

The association between ethnic background and multiplicative risk was statistically significant ($p < 0.001$). White British, Irish, and Caribbean sample members were more likely to fall in strata with mitigative risk effects. This means that they are more likely to be associated with multiplicative effects that reduce their risk of leaving the study sample. The likelihood of experiencing the highest multiplicative risks was seen among other Asian – with 37% of sample members falling in the top multiplicative risk quintile – followed by mixed and Bangladeshi.

When focusing on the migration background, we can see that second-generation individuals are significantly more likely to be associated with high multiplicative risks. Over half (55%) of Generation 2 sample members fall in the top two quintiles of multiplicative risk: 37% in the top quintile and 28% in the second. The likelihood of being in the bottom two quintiles – those exhibiting strong mitigative multiplicative effects – is higher for later-generation individuals, such as Generation 2.5 and 3.

The interaction between age, economic activity, and having dependent children, which was found to have strong multiplicative effects, was also examined to understand how the likelihood of experiencing it varies across ethnic and migration backgrounds (Table 19 and Table 20). Sample members with a mixed or other Asian background and Generation 2 were particularly likely to be under 35, economically active, and without dependent children. The likelihood of being aged 35-65, economically inactive, and having dependent children was higher among sample members with a Bangladeshi, Pakistani, or Generation 1 background.

Checking for compositional effects

The bivariate associations between multiplicative effects and ethnic or migration backgrounds could be influenced by compositional effects of the strata. This means that the level of risk associated with the different strata could be largely explained by the greater or lower presence of individuals who have ethnic minority or immigration backgrounds, and not by the combination of risk factors that compose them.

If this were the case, multiplicative effects could not explain the greater risk of attrition for sample members with ethnic minority or immigration background; instead, they would only reflect it. Is the variation in the mean risk in the strata reduced when we account for the ethnic and immigration background effects within each stratum?

To determine if this is the case, the MAIHDA model was expanded to include ethnic and migration background information. The first model, Model A, included both ethnic and migration backgrounds as fixed-effect variables. In the second model, Model B, these variables were also added as random slopes within the different strata.

Table 17 – Multiplicative risk strata residual by ethnic background

Risk quintile	White British	Other white	Pakistani	African	Any other	Bangladeshi	Caribbean	Indian	Irish	Mixed	Other Asian	p-value[1]
[-1.17,-0.644]	22%	15%	21%	8%	12%	16%	8%	20%	25%	13%	11%	<0.001
(-0.644,-0.274]	26%	19%	14%	16%	19%	13%	34%	18%	17%	19%	14%	
(-0.274,-0.0258]	21%	15%	20%	14%	17%	17%	19%	17%	22%	14%	15%	
(-0.0258,0.457]	15%	25%	19%	33%	24%	20%	19%	16%	20%	20%	23%	
(0.457,1.88]	16%	26%	26%	30%	29%	34%	20%	28%	16%	34%	37%	
Observations	37,349	1,391	1,824	1,605	986	1,427	1,247	2,253	761	984	1,100	

[1] Pearson's Chi-squared test

Table 18 – Multiplicative risk strata residuals by immigrant generation

Risk quintile	Gen 1	Gen 1.5	Gen 2	Gen 2.5	Gen 3+	p-value[1]
[-1.17,-0.644]	14%	18%	15%	20%	23%	<0.001
(-0.644,-0.274]	22%	20%	14%	21%	26%	
(-0.274,-0.0258]	19%	18%	15%	18%	21%	
(-0.0258,0.457]	26%	16%	18%	17%	15%	
(0.457,1.88]	19%	28%	37%	24%	15%	
Observations	6,512	2,126	3,039	2,321	33,641	

[1] Pearson's Chi-squared test

Table 19 – Selected multiplicative risk factor combinations by ethnic background

Selected combinations	White British	Other white	Pakistani	African	Any other	Bangladeshi	Caribbean	Indian	Irish	Mixed	Other Asian	p-value[1]
Risk-increasing combinations												<0.001
35-65, Inactive, Child Yes	2%	3%	8%	4%	6%	8%	3%	3%	3%	2%	5%	
Under 35, Active, Child No	15%	28%	24%	27%	25%	29%	19%	27%	14%	35%	34%	
Risk-decreasing combinations												
35-65, Inactive, Child No	18%	9%	10%	5%	9%	7%	9%	11%	19%	7%	6%	
Under 35, Active, Child Yes	10%	16%	16%	14%	13%	17%	8%	13%	10%	14%	10%	
Observations	37,349	1,391	1,824	1,605	986	1,427	1,247	2,253	761	984	1,100	

[1] Pearson's Chi-squared test

Table 20 – Selected multiplicative risk factor combinations by immigrant generation

Multiplicative risk	Gen 1	Gen 1.5	Gen 2	Gen 2.5	Gen 3+	p-value[1]
Risk-increasing combinations						<0.001
35-65, Inactive, Child Yes	6%	4%	3%	2%	2%	
Under 35, Active, Child No	16%	27%	36%	24%	14%	
Risk-decreasing combinations						
35-65, Inactive, Child No	10%	13%	6%	16%	18%	
Under 35, Active, Child Yes	13%	13%	15%	10%	10%	
Observations	6,512	2,126	3,039	2,321	33,641	

[1] Pearson's Chi-squared test

Due to model computational requirements, the ethnic or migration background variables were reduced to binaries. For ethnic background, white British were compared against the other ethnic groups, while for migration generation, the comparison was between Generation 2.5 or 3 and Generation 1, 1.5 and 2. Several tests were carried out using variables with a greater number of categories, but all the models failed to converge or encountered singularity.

The output of the models (Table 21 and Additionally, the model indicates a mild negative correlation between the random slope for immigrant background and the stratum intercept. This means that the greater the attrition risk associated with the strata, the smaller the difference between Generation 2.5-3 and Generation 1-2. In other terms, there are larger differences by migrant generations for strata associated with mitigative risk (a reduction in the risk of attrition caused by multiplicative effects) than among strata with an aggravation of the attrition risk. There was no correlation between the strata intercept and the ethnic background slope.

Overall, the effect of ethnic backgrounds on the model appears to be making a minimal contribution to the explanation of attrition risk. Ethnic background appears to be a not statistically significant fixed-effect predictor and a weak random slope when controlling for the other variables. Migrant generation was a stronger fixed-effect predictor, with Generation 1-2 being 1.4 times more likely to leave the study sample compared to Generation 2.5-3, all other things being equal. It was also associated with a random slope, which had greater explanatory power compared to ethnic background.

These findings suggest that the attrition risk, as explained by the complex interactions of risk factors, was not largely driven by a compositional effect. The variance in the outcome explained by the combinations of risk factors did not strongly vary after controlling for ethnic or migration backgrounds across the different strata. This provides stronger evidence to infer that the attrition of individuals holding ethnic minority or immigration backgrounds is partly shaped by their greater likelihood of experiencing multiplicative risk effects.

Table 22) indicates that the risk emerging from the multiplicative effects is only slightly reduced when we account for the effects of ethnic and migration backgrounds across the different strata. After including ethnic or migration backgrounds as fixed effects, we observe a reduction in the random intercept variance of less than 1% of the VPC (to 9% of the total risk of attrition). The addition of the random slopes for ethnic or migration backgrounds, nested in the different strata, does not influence the strata random intercept, which remains at 9%. The risk variance explained by the random slopes (3% for migration background and 1% for ethnic background) does not reduce the multiplicative risk emerging from the strata, but the residual variance attributable to differences between individuals.

Table 21 – Model A and Model B: random effects

	Model A Fixed effects	**Model B** Random slopes
N obs	47638	47638

	N groups	165	165
Strata intercept	Std. Dev.	0.57	0.59
	Variance	0.32	0.35
	VPC	0.09	0.09
Migrant generation random slope	Std. Dev.	-	0.36
	Variance	-	0.13
	VPC	-	0.03
	Corr	-	-0.36
Ethnic background random slope	Std. Dev.	-	0.15
	Variance	-	0.02
	VPC	-	0.01
	Corr	-	-0.03

Additionally, the model indicates a mild negative correlation between the random slope for immigrant background and the stratum intercept. This means that the greater the attrition risk associated with the strata, the smaller the difference between Generation 2.5-3 and Generation 1-2. In other terms, there are larger differences by migrant generations for strata associated with mitigative risk (a reduction in the risk of attrition caused by multiplicative effects) than among strata with an aggravation of the attrition risk. There was no correlation between the strata intercept and the ethnic background slope.

Overall, the effect of ethnic backgrounds on the model appears to be making a minimal contribution to the explanation of attrition risk. Ethnic background appears to be a not statistically significant fixed-effect predictor and a weak random slope when controlling for the other variables. Migrant generation was a stronger fixed-effect predictor, with Generation 1-2 being 1.4 times more likely to leave the study sample compared to Generation 2.5-3, all other things being equal. It was also associated with a random slope, which had greater explanatory power compared to ethnic background.

These findings suggest that the attrition risk, as explained by the complex interactions of risk factors, was not largely driven by a compositional effect. The variance in the outcome explained by the combinations of risk factors did not strongly vary after controlling for ethnic or migration backgrounds across the different strata. This provides stronger evidence to infer that the attrition of individuals holding ethnic minority or immigration backgrounds is partly shaped by their greater likelihood of experiencing multiplicative risk effects.

Table 22 – Model A and Model B: fixed effects

Terms	**Model A** Fixed effects			**Model B** Random slopes		
	Log odds	**Std. Error**	**p-value**	**Log odds**	**Std. Error**	**p-value**
(Intercept)	-0.59	0.16	<0.001	-0.58	0.16	<0.001
Age group (ref: 35-65)						

Over 65	0.99	0.18	<0.001	0.95	0.18	<0.001
Under 35	0.23	0.12	0.053	0.23	0.12	0.051
Level of formal education (ref: degree and above)						
Below degree	0.70	0.11	<0.001	0.65	0.12	<0.001
Tenure (ref: Onwer or other)						
Private renter furnished or employer	1.74	0.17	<0.001	1.77	0.17	<0.001
Private renter unfurnished	0.44	0.15	<0.01	0.50	0.15	<0.01
Social renter	0.46	0.14	<0.001	0.48	0.14	<0.001
Area type (ref: rural)						
Urban	0.23	0.11	<0.05	0.29	0.12	<0.05
Dependent children (ref: No)						
Yes	0.39	0.12	<0.001	0.36	0.12	<0.01
Economic activity (ref: Inactive)						
Active	0.21	0.11	0.064	0.18	0.11	0.110
Migration BG (ref: Gen 2.5-3)						
Generation 1-2	0.40	0.05	<0.001	0.34	0.08	<0.001
Ethnic BG (ref: white British)						
Other ethnic background	0.04	0.05	0.454	0.03	0.06	0.612
N observations	47638					

Discussion

This research paper examined attrition among sample members from different ethnic or migration backgrounds in *Understanding Society*, focusing on interactions and multiplicative effects between risk factors.

The first research question sought to understand whether differences between groups could be explained by a different impact of the predictors on attrition in the different ethnic or migration groups. Overall, there was evidence of variations in the sample, but, in many instances, they had a very small effect size.

The risk factor that varied the most was the area type. After accounting for variations of the risk factor for each group, it was possible to see that the divide between urban and rural sample members strongly widens for those with an ethnic minority or an immigration background. Specifically, the effect of urbanicity on nonresponse was stronger among sample members with an ethnic minority background than it was for white British or Generation 3. While this was also found in earlier studies in the Netherlands (Feskens et al., 2007), the variation by area type may also be caused by sampling strategies in Understanding Society. The 2009-10 Ethnic Minority Boost was recruited by oversampling urban areas known to have a greater proportion of residents with an ethnic minority background (Lynn, 2009). These areas also have specific socio-economic profiles, such as high levels of material deprivation, meaning that the

comparison in the model may not be between sample members with ethnic minority or immigrant backgrounds living in rural versus urban areas, but between those living in rural versus deprived urban areas. While weights in Understanding Society correct for this, the analysis presented in this paper was carried out on unweighted data, as it aimed to understand the behaviour of sample members. This suggests that this geographical variation must be validated on a general population sample.

Further interesting differences were found across housing tenure and level of formal education. Both predictors could explain more variation in attrition among white British or third-generation sample members than among those with an ethnic minority or immigration background. In other terms, these predictors had the same association across the different groups, but a different strength: a sample member who has completed higher education is more likely to remain in the study than a sample member with lower educational attainments, but this effect is stronger among white British than it is in other ethnic groups.

The second and third research questions shifted the attention to the multiplicative effects that emerge in the data, and to the association between these effects and ethnic or immigration background. This was done using an innovative method for the study of multiplicative risk effects, MAIHDA. The six variables included in the model (age, economic activity, having dependent children, level of formal education, housing tenure, and area type) revealed a very high level of total sample inequity, at 23%. This means that 23% of the variance of attrition could be attributed to the 166 combinations observed in the data for these six variables. The analysis revealed that 36% of this total sample inequity was multiplicative, or due to interactions of risk factors. Overall, 10% of the variance of attrition in the sample was multiplicative.

This level of multiplicative effects is particularly large compared to what would be typically expected in a MAIHDA model (Evans et al., 2024). This is probably attributable to the fact that survey nonresponse incorporates both structural considerations – such as housing tenure – but also attitudinal aspects – such as perceived salience, for which formal education, age, and other structural elements may serve as a proxy – leading to significant variations between the different strata.

The multiplicative risk was also associated with ethnic and immigration backgrounds. While sample members with white British, Irish, Caribbean, or Generation 3 backgrounds were more likely to experience mitigative multiplicative effects, those with Bangladeshi, other Asian, mixed identities, or Generation 2s, were more likely to experience an intensification of risk.

Some strata contributed strongly towards increasing the risk of attrition in some groups. For example, sample members aged 35-65, economically inactive, with dependent children were particularly likely to be from Pakistani or Bangladeshi backgrounds. This strata may identify a combined increase in attrition risk linked to cost opportunities, emerging from caring responsibilities, material deprivation, or a combination of both, unfolding in the context of multigenerational households, which are common in these ethnic groups (ONS, 2023).

Being economically active, below the age of 35, with or without dependent children, was a multiplicative stratum likely to be found in second generations. These are all characteristics generally associated with no-contact outcomes. Young people who are economically active are difficult to trace due to their high mobility and frequent address changes. The presence of dependent children may make this population more stationary, regardless of housing tenure type, providing mitigation for no-contact outcomes.

Implications for survey methodology and future research

The first contribution of this paper for survey methodology is to identify where additional fieldwork and design solutions for the mitigation of attrition among sample members with ethnic minority backgrounds are likely to be more effective. The findings encourage the development of additional solutions that can target characteristics associated with attrition, defined in this paper as risk factors, responsible for both additive and multiplicative attrition risk, rather than solutions designed to retain participation in a specific ethnic group. After controlling for these risk factors for attrition, the effect of having an ethnic minority background on attrition is not significant. While these considerations also hold true for sample members with an immigration background, being a first or second-generation immigrant maintains a significant association with attrition after controlling for the risk factors. However, the effect size appears to be particularly small.

The second contribution of this paper is to introduce a new method for the study of survey nonresponse, MAIHDA. This method enables precise estimation of multiplicative effects in longitudinal attrition and offers a powerful exploratory tool for statisticians seeking to capture them in longitudinal weighting. Survey weighting strategies generally account for interactions or multiplicative effects that may emerge among the weighting variables. For instance, the Understanding Society team iteratively tests a large set of interaction terms, retaining those with greater predictive power in the final models. MAIHDA can support statisticians in early-stage decisions about which combinations of risk factors for attrition they want to test in their models, as the findings can reveal unexpected multiplicative effect structures. Indeed, important three-way interactions, such as the one identified in this paper between age, economic activity, and dependent children, may be missed in an iterative weighting process, but would appear in a MAIHDA exploration.

This research paper offered an initial examination of multiplicative effects but did not address some aspects that could enhance our understanding of this dimension of nonresponse. For example, it did not specifically account for household dynamics, which are particularly important in a study like Understanding Society, where individual forms of nonresponse can be mitigated or exacerbated by what happens in a household. Future research could also focus on the longitudinal component of attrition to understand how multiplicative effects influence the time of attrition in a study. MAIHDA has already been applied to longitudinal analysis (Bell et al., 2024), meaning that the method used in this paper could also be employed to understand trajectories of nonresponse over time.

Reference list

Arel-Bundock, V., Greifer, N., and Heiss, A. (2024). How to Interpret Statistical Models Using marginaleffects for R and Python. Journal of Statistical Software, Vol 111, Issue 9.

Atrostic, B., Baytes, N. and Silberstein, A. (2001). *Nonresponse in U.S. Government household surveys: consistent measures, recent trends, and new insights*. Journal of Official Statistics, 17, 209–226.

Behr, A., Bellgardt, E. and Rendtel, U. (2005). *Extent and determinants of panel attrition in the European Community Household Panel*. European Sociological Review, 23, 81–97.

Bell, A., Evans, C., Holman, D., Leckie, G. (2024). *Extending intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) to study individual longitudinal trajectories, with application to mental health in the UK*. Social Science & Medicine, 351, 116955.

Cabrera-Alvarez, P., and Lynn, P. (2023). *Trends in Panel Attrition in Understanding Society: Waves 1 to 13*. Understanding Society Working Paper Series, No. 2023-16. Available at https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2023-16.pdf (accessed 25 Jan 2025).

Cuff, E., Sharrock, W., and Francis, D. (2005). *Perspectives in Sociology*. 5th Edition, London: Routledge.

De Leeuw, E., and De Heer, W. (2002). *Trends in household survey nonresponse: a longitudinal and international comparison*. In R. Groves, D. Dillman, J. Eltinge and R. Little (Eds), Survey Nonresponse (pp. 41–54). New York: John Wiley & Sons, Inc.

Deng, T., Hillygus, S., Reiter, J., Si, Y., and Zheng, S. (2013), *Handling Attrition in Longitudinal Studies: The Case for Refreshment Samples*, Statist. Sci. 28(2), 238-256.

Evans, C. (2015). *Innovative approaches to investigating social determinants of health-social networks, environmental effects and intersectionality*. Boston, MA: Harvard T.H. Chan School of Public Health. ScD Dissertation.

Evans, C., Leckie, G., Subramanian, S., Bell, A., and Merlo, J. (2024). *A tutorial for conducting intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA)*, SSM - Population Health, Volume 26, 101664.

Evans, C., Williams, D., Onnela, J., and Subramanian, S. (2018). *A multilevel approach to modeling health inequalities at the intersection of multiple social identities*. Social Science & Medicine, 203, 64–73.

Feskens, R., Hox, J., Lensvelt-Mulders, G., and Schmeets, H. (2007). *Nonresponse Among Ethnic Minorities: A Multivariate Analysis*. Journal of Official Statistics, 23 (3), 387-408.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. A. (1998). *An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics*. National Bureau of Economic Research. Technical Working Paper, 0220. Available at: https://www.nber.org/papers/t0220 (accessed 25 Jan 2025).

Font, J., and Mendez, M. (2013). *Surveying Ethnic Minorities and Immigrant Populations Methodological Challenges and Research Strategies*. Amsterdam: Amsterdam University Press.

Gray, R., Campanelli, P., Deepchand, K., & Prescott-Clarke, P. (1996). *Exploring Survey Non-Response: The Effect of Attrition on a Follow-Up of the 1984-85 Health and Life Style Survey*. Journal of the Royal Statistical Society. Series D (The Statistician), 45(2), 163–183

Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.

Groves, R.M., Singer, E., and Corning, A. (2000). *Leverage-saliency theory of survey participation: description and an illustration*. Public Opinion Quarterly, 64 (3): 299–308.

Jenkins, S. P. (1997). *Attrition in longitudinal surveys: Evidence from the British Household Panel Survey*. Journal of the Royal Statistical Society: Series A (Statistics in Society), 160(3), 331–354.

Kappelhof, J. (2017). *Survey Research and the quality of survey data among ethnic minorities*. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, N. Tucker, B. West (eds.), Total Survey Error in practice, pp. 235-252. Wiley Series in Survey Methodology. Hoboken, NJ: Wiley and Sons.

Kappelhof, J. (2015). *Face-to-Face or Sequential Mixed-Mode Surveys Among Non-Western Minorities in the Netherlands: The Effect of Different Survey Designs on the Possibility of Nonresponse Bias*. Journal of Official Statistics, 31(1), 1-30.

Lepkowski, J. M. and Couper, M. P. (2002). *Nonresponse in the second wave of longitudinal household surveys*. In R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little (Eds), Survey Nonresponse (pp. 259–272). New York: John Wiley & Sons, Inc.

Loosveldt, G., Pickery, J., and Billiet J. (2002), *Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey*. Journal of Official Statistics, 18(4), 545-557.

Luiten, A., Hox, J., and de Leeuw, E. (2020). *Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys*. Journal of Official Statistics, *36*(3), 469-487.

Lynn, P. (2009). *Sample Design for Understanding Society*. Understanding Society Working Paper Series, No. 2009-01. Available at https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2009-01.pdf (accessed 25 Jan 2025).

Mize, T. (2019). *Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects*. Sociological Science, 6: 81-117.

Office for National Statistics, ONS (2024). *Labour Force Survey (LFS) QMI*, 15 October 2024. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveylfsqmi (accessed 25 Jan 2025).

Office for National Statistics, ONS (2023). *Families in England and Wales: Census 2021*, 10 May 2023. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/articles/familiesinenglandandwales/census2021 (accessed 25 Jan 2025).

Olsen, R. (2005). *The problem of respondent attrition: Survey methodology is key*. Monthly Labor Review 128 63–71.

Pyper, D. (2021). *Race and ethnic disparities*. House of Commons Library, Briefing Paper CBP 8960, 24 May 2021. Available at https://researchbriefings.files.parliament.uk/documents/CBP-8960/CBP-8960.pdf (accessed 25 Jan 2025).

Snijders, T., and Bosker, R. (2012). *Multilevel analysis: an introduction to basic and advanced multilevel modelling*. Los Angeles: Sage Publications.

Watson N., Wooden M. (2009). *Identifying Factors Affecting Longitudinal Survey Response*. In P. Lynn (Ed.), Methodology of Longitudinal Surveys, (pp. 157-181). Wiley Series in Survey Methodology. Hoboken, NJ: Wiley & Sons.

Data

University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Waves 1-13, 2009-2022 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 18th Edition. UK Data Service. SN: 6614, DOI: http://doi.org/10.5255/UKDA-SN-6614-19

Appendix

Table 23 – Prevalence of demographics and socio-economic characteristics associated with attrition in the different ethnic groups

Characteristic	white British N = 37,445	Other white N = 1,396	Pakistani N = 1,835	African N = 1,613	Any other N = 990	Bangladeshi N = 1,433	Caribbean N = 1,255	Indian N = 2,269	Irish N = 763	Mixed N = 989	Other Asian N = 1,110	Chi-sq p-value
Immigration background												<0.001
Generation 1	0%	80%	43%	67%	58%	45%	29%	52%	18%	22%	69%	
Generation 1.5	0%	7%	18%	19%	16%	26%	20%	16%	8%	7%	18%	
Generation 2	1%	4%	34%	12%	14%	27%	38%	28%	9%	14%	12%	
Generation 2.5	5%	4%	4%	1%	5%	1%	6%	3%	9%	34%	1%	
Generation 3+	93%	4%	0%	1%	7%	1%	6%	1%	55%	23%	1%	
Whether English is first language												<0.001
Yes	99%	29%	32%	38%	35%	23%	98%	38%	99%	84%	24%	
No	1%	71%	68%	62%	65%	77%	2%	62%	1%	16%	76%	
Latest known area												<0.001
Rural area	27%	13%	0%	1%	6%	1%	1%	2%	41%	5%	3%	
Urban area	73%	87%	100%	99%	95%	99%	99%	98%	59%	95%	97%	
Latest known tenure												<0.001
Owned or other	72%	45%	72%	23%	40%	40%	48%	76%	71%	47%	51%	
Private renter furnished or employer	3%	23%	10%	15%	16%	13%	6%	14%	5%	10%	27%	
Private renter unfurnished	8%	21%	8%	11%	12%	6%	8%	5%	8%	13%	10%	
Social renter	16%	12%	11%	50%	32%	41%	38%	6%	16%	29%	12%	
Formal education												<0.001
Below degree	66%	44%	65%	46%	53%	71%	65%	47%	61%	54%	40%	
Degree or higher	34%	56%	35%	54%	47%	29%	35%	53%	39%	46%	60%	
Age group												<0.001
Under 35	27%	47%	51%	46%	44%	57%	29%	42%	28%	54%	49%	
35-65	56%	47%	44%	51%	49%	40%	56%	50%	58%	41%	48%	
Under 35	27%	47%	51%	46%	44%	57%	29%	42%	28%	54%	49%	
Whether has dependent children												<0.001
Yes	27%	37%	50%	48%	43%	53%	29%	36%	30%	33%	37%	
No	73%	63%	50%	52%	57%	47%	71%	64%	70%	67%	63%	
Sex												<0.001
Male	46%	42%	49%	44%	48%	50%	42%	52%	45%	42%	48%	
Female	54%	58%	51%	56%	52%	50%	58%	48%	55%	58%	52%	

Latest known economic activity												<0.001
Economically Active	61%	80%	66%	83%	73%	71%	71%	76%	61%	81%	82%	
Economically Inactive	39%	20%	34%	17%	27%	29%	29%	24%	39%	19%	18%	

Table 24 – Prevalence of demographics and socio-economic characteristics associated with attrition in the different immigration backgrounds

	Gen 1	Gen 1.5	Gen 2	Gen 2.5	Gen 3+	Chi-square p-value
	N =	N =	N =	N =	N =	
Characteristic	6,534	2,133	3,045	2,323	33,680	
Ethnic background						<0.001
white British	2%	4%	14%	69%	97%	
African	15%	14%	6%	1%	0%	
Any other	8%	7%	4%	2%	0%	
Bangladeshi	9%	16%	11%	1%	0%	
Caribbean	5%	11%	14%	3%	0%	
Indian	17%	16%	19%	2%	0%	
Irish	2%	3%	2%	3%	1%	
Mixed	3%	3%	4%	13%	1%	
Other Asian	11%	8%	4%	1%	0%	
Other white	16%	5%	2%	2%	0%	
Pakistani	11%	15%	19%	3%	0%	
Whether English is first language						<0.001
Yes	23%	41%	81%	98%	99%	
No	77%	59%	19%	2%	1%	
Latest known area						<0.001
Rural area	4%	3%	4%	21%	28%	
Urban area	96%	97%	96%	79%	72%	
Latest known tenure						<0.001
Owned or other	46%	55%	65%	68%	72%	
Private renter furnished or employer	20%	10%	6%	5%	3%	
Private renter unfurnished	12%	7%	7%	9%	9%	
Social renter	22%	28%	22%	18%	17%	
Formal education						<0.001
below degree	50%	65%	52%	57%	66%	
degree or higher	50%	35%	48%	43%	34%	
Age group						<0.001
Under 35	35%	46%	56%	38%	27%	
35-65	54%	51%	43%	53%	55%	
Over 65	11%	3%	1%	9%	18%	
Whether has dependent children						<0.001
Yes	46%	37%	37%	28%	27%	
No	54%	63%	63%	72%	73%	
Sex						<0.001
Male	46%	49%	44%	44%	45%	
Female	54%	51%	56%	56%	55%	
Latest known economic activity						<0.001
Economically Active	68%	73%	85%	70%	60%	
Economically Inactive	32%	27%	15%	30%	40%	

Figure 3 – Distribution and prevalence of missing values in the analysis sample

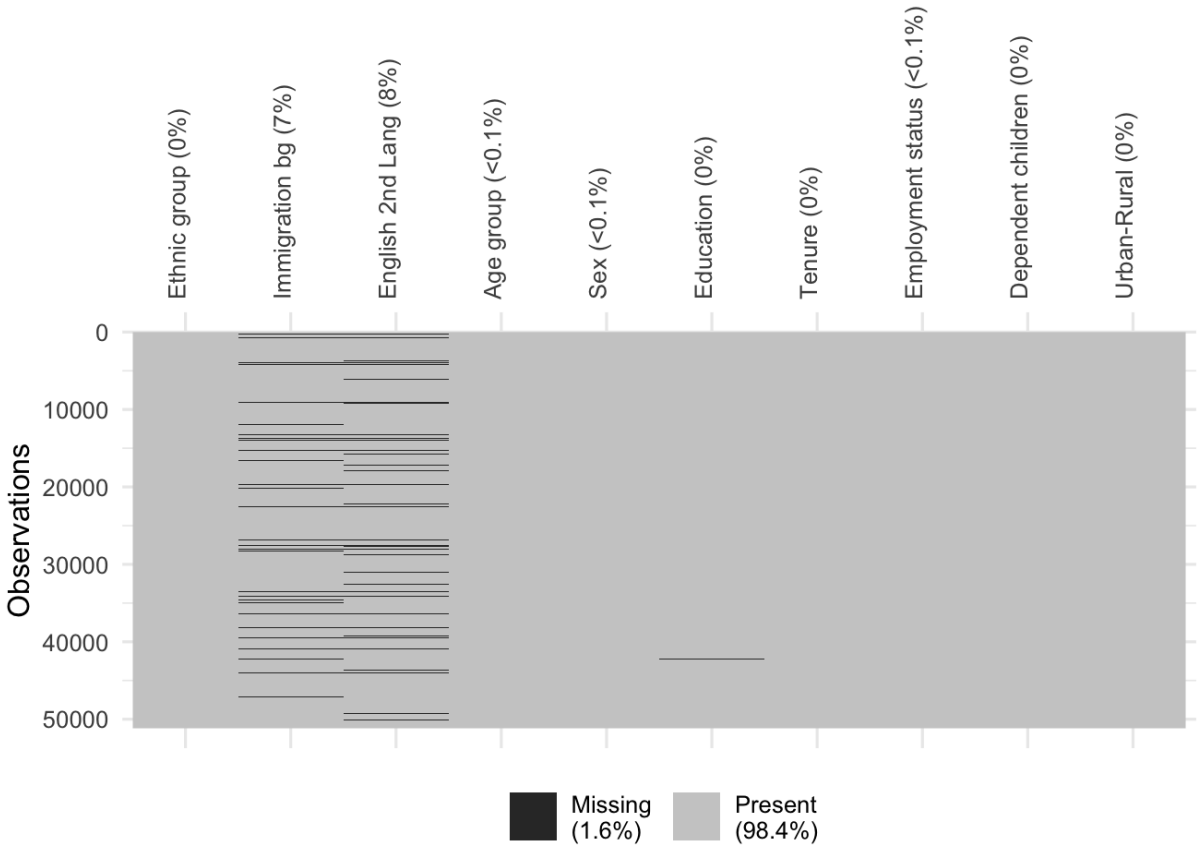


Figure 4 – Distribution and prevalence of missing values in the analysis sample, by attrition

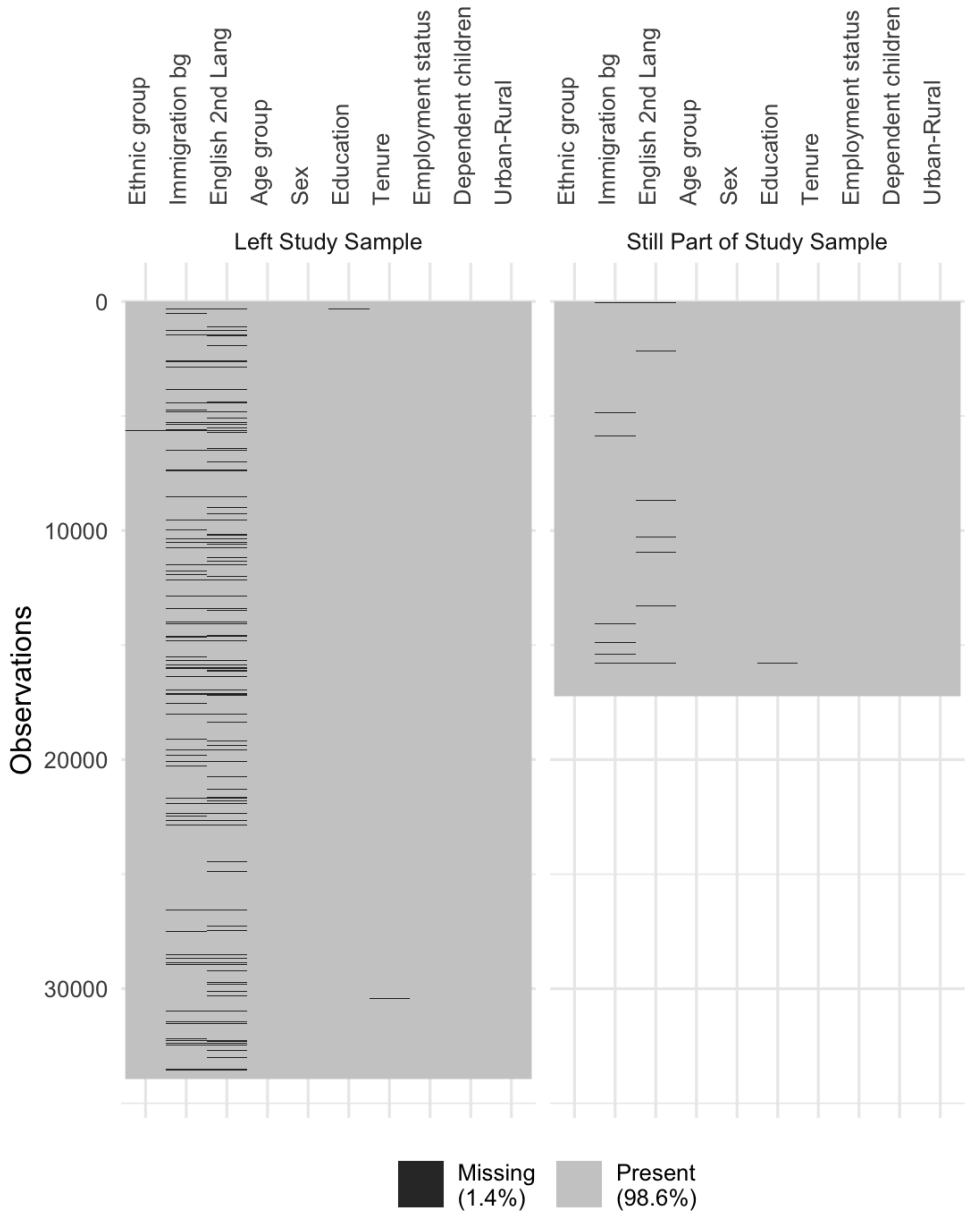


Table 25 – Interaction terms between ethnic background and sex

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.57	0.01	<0.001	0.55	0.02	<0.001
Ethnicity (ref: white British)						
African	1.25	0.07	<0.001	1.35	0.11	<0.001
Any other	1.00	0.08	<0.001	1.14	0.13	<0.001
Bangladeshi	0.65	0.06	<0.001	0.73	0.09	<0.001
Caribbean	0.67	0.07	<0.001	0.94	0.11	<0.001
Indian	0.43	0.05	<0.001	0.48	0.07	<0.001
Irish	0.06	0.08	0.43	0.05	0.11	0.64
Mixed	0.41	0.07	<0.001	0.51	0.11	<0.001
Other Asian	0.75	0.07	<0.001	0.84	0.11	<0.001
Other white	0.54	0.06	<0.001	0.74	0.10	<0.001
Pakistani	0.49	0.05	<0.001	0.51	0.08	<0.001
Sex (ref: Male)						
Female	-0.09	0.02	<0.001	-0.05	0.02	0.02
Interaction terms						
African:Female				-0.17	0.15	0.24
Any other:Female				-0.24	0.17	0.16
Bangladeshi:Female				-0.14	0.13	0.27
Caribbean:Female				-0.45	0.14	<0.001
Indian:Female				-0.10	0.10	0.29
Irish:Female				0.01	0.15	0.94
Mixed:Female				-0.17	0.15	0.24
Other Asian:Female				-0.16	0.15	0.27
Other white:Female				-0.33	0.13	0.01
Pakistani:Female				-0.04	0.11	0.71

N observations: 51,069

Table 26 – Interaction terms between ethnic background and age group

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.23	0.01	<0.001	0.22	0.01	<0.001
Ethnicity (ref: white British)						
African	1.27	0.07	<0.001	1.37	0.09	<0.001
Any other	1.00	0.08	<0.001	1.07	0.11	<0.001
Bangladeshi	0.60	0.06	<0.001	0.77	0.10	<0.001
Caribbean	0.69	0.07	<0.001	0.75	0.09	<0.001
Indian	0.43	0.05	<0.001	0.50	0.06	<0.001
Irish	0.08	0.08	0.32	-0.09	0.10	0.37
Mixed	0.35	0.07	<0.001	0.31	0.10	<0.001
Other Asian	0.74	0.07	<0.001	0.76	0.10	<0.001
Other white	0.51	0.06	<0.001	0.41	0.08	<0.001
Pakistani	0.46	0.05	<0.001	0.67	0.08	<0.001
Age group (ref: 35-65)						
Over 65	0.84	0.03	<0.001	0.86	0.03	<0.001
Under 35	0.63	0.02	<0.001	0.65	0.03	<0.001
Interaction terms						
African:Over 65				-1.12	0.39	<0.001
Any other:Over 65				-0.67	0.34	0.05
Bangladeshi:Over 65				-0.55	0.37	0.14
Caribbean:Over 65				-0.17	0.22	0.44
Indian:Over 65				-0.11	0.21	0.60
Irish:Over 65				0.25	0.25	0.31
Mixed:Over 65				-0.14	0.37	0.71
Other Asian:Over 65				-0.59	0.41	0.15
Other white:Over 65				-0.10	0.28	0.71
Pakistani:Over 65				0.04	0.32	0.90
African:Under 35				-0.20	0.15	0.19
Any other:Under 35				-0.08	0.18	0.65
Bangladeshi:Under 35				-0.32	0.13	0.01
Caribbean:Under 35				-0.16	0.16	0.33
Indian:Under 35				-0.17	0.10	0.10
Irish:Under 35				0.58	0.20	<0.001
Mixed:Under 35				0.08	0.15	0.61
Other Asian:Under 35				-0.01	0.15	0.93
Other white:Under 35				0.28	0.13	0.04
Pakistani:Under 35				-0.46	0.11	<0.001

N observations: 51,097

Table 27 – Interaction terms between ethnic background and English second language

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.43	0.01	<0.001	0.42	0.01	<0.001
Ethnicity (ref: white British)						
African	1.10	0.08	<0.001	1.13	0.11	<0.001
Any other	0.82	0.09	<0.001	0.73	0.13	<0.001
Bangladeshi	0.37	0.07	<0.001	0.66	0.14	<0.001
Caribbean	0.65	0.07	<0.001	0.64	0.07	<0.001
Indian	0.27	0.06	<0.001	0.38	0.08	<0.001
Irish	0.12	0.08	0.13	0.11	0.08	0.16
Mixed	0.35	0.07	<0.001	0.30	0.08	<0.001
Other Asian	0.52	0.08	<0.001	0.20	0.14	0.14
Other white	0.38	0.07	<0.001	0.15	0.11	0.15
Pakistani	0.26	0.06	<0.001	0.41	0.10	<0.001
English first language (ref: Yes)						
No	0.27	0.05	<0.001	0.30	0.12	0.01
Interaction terms						
African:No				-0.08	0.19	0.66
Any other:No				0.12	0.21	0.57
Bangladeshi:No				-0.41	0.20	0.04
Caribbean:No				0.71	0.76	0.35
Indian:No				-0.20	0.16	0.21
Irish:No				0.96	1.09	0.38
Mixed:No				0.33	0.25	0.19
Other Asian:No				0.42	0.21	0.04
Other white:No				0.31	0.18	0.08
Pakistani:No				-0.25	0.17	0.14
N observations: 46,839						

Table 28 – Interaction terms between ethnic background and area type

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.38	0.02	<0.001	0.41	0.02	<0.001
Ethnicity (ref: white British)						
African	1.19	0.07	<0.001	0.18	0.56	0.74
Any other	0.96	0.08	<0.001	0.21	0.29	0.47
Bangladeshi	0.60	0.06	<0.001	0.29	0.61	0.64
Caribbean	0.61	0.07	<0.001	-0.25	0.56	0.65
Indian	0.38	0.05	<0.001	-0.06	0.32	0.85
Irish	0.09	0.08	0.25	-0.01	0.12	0.91
Mixed	0.36	0.07	<0.001	-0.52	0.28	0.06
Other Asian	0.70	0.07	<0.001	-0.23	0.34	0.49
Other white	0.51	0.06	<0.001	0.19	0.15	0.23
Pakistani	0.44	0.05	<0.001	-1.79	1.12	0.11
Area type (ref: rural)						
urban area	0.20	0.02	<0.001	0.17	0.02	<0.001
Interaction terms						
African:urban area				1.03	0.56	0.07
Any other:urban area				0.83	0.30	0.01
Bangladeshi:urban area				0.33	0.62	0.59
Caribbean:urban area				0.89	0.56	0.11
Indian:urban area				0.46	0.32	0.15
Irish:urban area				0.17	0.15	0.28
Mixed:urban area				0.95	0.29	<0.001
Other Asian:urban area				0.99	0.35	<0.001
Other white:urban area				0.39	0.17	0.02
Pakistani:urban area				2.25	1.12	0.04
N observations: 51,098						

Table 29 – Interactio terms between ethnic background and level of formal education

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	1.23	0.03	<0.001	1.32	0.03	<0.001
Ethnicity (ref: white British)						
African	1.41	0.07	<0.001	1.39	0.33	<0.001
Any other	1.11	0.08	<0.001	1.17	0.32	<0.001
Bangladeshi	0.60	0.06	<0.001	-0.12	0.13	0.38
Caribbean	0.68	0.07	<0.001	0.72	0.23	<0.001
Indian	0.57	0.05	<0.001	-0.01	0.15	0.92
Irish	0.07	0.08	0.39	-0.29	0.19	0.12
Mixed	0.53	0.07	<0.001	0.55	0.38	0.15
Other Asian	0.94	0.07	<0.001	0.34	0.27	0.22
Other white	0.71	0.06	<0.001	0.95	0.30	<0.001
Pakistani	0.48	0.05	<0.001	-0.08	0.12	0.54
Level of formal education (ref: no qualification)						
below degree	-0.57	0.03	<0.001	-0.65	0.04	<0.001
degree or higher	-1.15	0.03	<0.001	-1.28	0.04	<0.001
Interaction terms						
African:below degree				<0.001	0.35	0.99
Any other:below degree				-0.15	0.35	0.66
Bangladeshi:below degree				0.67	0.16	<0.001
Caribbean:below degree				-0.11	0.25	0.67
Indian:below degree				0.37	0.17	0.03
Irish:below degree				0.43	0.23	0.06
Mixed:below degree				0.10	0.40	0.80
Other Asian:below degree				0.53	0.31	0.08
Other white:below degree				-0.24	0.33	0.46
Pakistani:below degree				0.37	0.15	0.01
African:degree or higher				0.07	0.34	0.83
Any other:degree or higher				0.01	0.33	0.97
Bangladeshi:degree or higher				1.17	0.17	<0.001
Caribbean:degree or higher				0.04	0.25	0.88
Indian:degree or higher				0.82	0.16	<0.001
Irish:degree or higher				0.43	0.22	0.05
Mixed:degree or higher				-0.09	0.39	0.82
Other Asian:degree or higher				0.72	0.29	0.01
Other white:degree or higher				-0.24	0.31	0.45
Pakistani:degree or higher				0.99	0.15	<0.001

N observations: 50,983

Table 30 – Interaction terms between ethnic background and housing tenure

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.33	0.01	<0.001	0.32	0.01	<0.001
Ethnicity (ref: white British)						
African	0.83	0.07	<0.001	0.82	0.12	<0.001
Any other	0.72	0.09	<0.001	0.61	0.11	<0.001
Bangladeshi	0.37	0.07	<0.001	0.66	0.09	<0.001
Caribbean	0.50	0.07	<0.001	0.74	0.09	<0.001
Indian	0.39	0.05	<0.001	0.36	0.05	<0.001
Irish	0.04	0.08	0.61	0.01	0.09	0.95
Mixed	0.19	0.07	0.01	0.28	0.10	<0.001
Other Asian	0.44	0.08	<0.001	0.42	0.09	<0.001
Other white	0.21	0.07	<0.001	-0.03	0.08	0.73
Pakistani	0.46	0.06	<0.001	0.56	0.06	<0.001
Tenure (ref: Onwer or other)						
Private renter furnished or employer	2.19	0.08	<0.001	2.31	0.12	<0.001
Private renter unfurnished	0.57	0.04	<0.001	0.56	0.04	<0.001
Social renter	0.69	0.03	<0.001	0.72	0.03	<0.001
Interaction terms						
African:Private renter furnished or employer				-0.91	0.30	<0.001
Any other:Private renter furnished or employer				-0.18	0.42	0.67
Bangladeshi:Private renter furnished or employer				-1.57	0.25	<0.001
Caribbean:Private renter furnished or employer				-0.80	0.49	0.10
Indian:Private renter furnished or employer				0.51	0.36	0.16
Irish:Private renter furnished or employer				1.02	1.02	0.32
Mixed:Private renter furnished or employer				0.99	0.73	0.18
Other Asian:Private renter furnished or employer				0.21	0.34	0.55
Other white:Private renter furnished or employer				1.33	0.44	<0.001
Pakistani:Private renter furnished or employer				-1.03	0.28	<0.001
African:Private renter unfurnished				0.35	0.26	0.19
Any other:Private renter unfurnished				<0.001	0.27	0.99
Bangladeshi:Private renter unfurnished				-0.27	0.28	0.33

Caribbean:Private renter unfurnished	-0.45	0.25	0.08
Indian:Private renter unfurnished	0.47	0.27	0.08
Irish:Private renter unfurnished	0.05	0.31	0.87
Mixed:Private renter unfurnished	-0.50	0.21	0.02
Other Asian:Private renter unfurnished	0.01	0.25	0.97
Other white:Private renter unfurnished	0.56	0.17	<0.001
Pakistani:Private renter unfurnished	-0.09	0.22	0.68
African:Social renter	0.03	0.16	0.87
Any other:Social renter	0.42	0.21	0.05
Bangladeshi:Social renter	-0.48	0.14	<0.001
Caribbean:Social renter	-0.56	0.15	<0.001
Indian:Social renter	0.03	0.24	0.90
Irish:Social renter	0.15	0.23	0.51
Mixed:Social renter	-0.15	0.17	0.37
Other Asian:Social renter	-0.02	0.24	0.93
Other white:Social renter	0.51	0.22	0.02
Pakistani:Social renter	-0.56	0.17	<0.001

N observations: 51,044

Table 31 – Interaction terms between ethnic background and having dependent children

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.50	0.01	<0.001	0.48	0.01	<0.001
Ethnicity (ref: white British)						
African	1.22	0.07	<0.001	1.21	0.10	<0.001
Any other	0.99	0.08	<0.001	0.93	0.11	<0.001
Bangladeshi	0.63	0.06	<0.001	0.85	0.10	<0.001
Caribbean	0.67	0.07	<0.001	0.75	0.08	<0.001
Indian	0.42	0.05	<0.001	0.63	0.06	<0.001
Irish	0.06	0.08	0.47	0.03	0.09	0.77
Mixed	0.40	0.07	<0.001	0.47	0.09	<0.001
Other Asian	0.74	0.07	<0.001	0.88	0.09	<0.001
Other white	0.52	0.06	<0.001	0.64	0.08	<0.001
Pakistani	0.47	0.05	<0.001	0.60	0.08	<0.001
Dependent children (ref: No)						
Yes	0.11	0.02	<0.001	0.18	0.02	<0.001
Interaction terms						
African:Yes				-0.01	0.14	0.95
Any other:Yes				0.13	0.17	0.46
Bangladeshi:Yes				-0.45	0.13	<0.001
Caribbean:Yes				-0.28	0.15	0.06
Indian:Yes				-0.59	0.10	<0.001
Irish:Yes				0.09	0.17	0.59
Mixed:Yes				-0.25	0.15	0.10
Other Asian:Yes				-0.39	0.15	0.01
Other white:Yes				-0.34	0.13	0.01
Pakistani:Yes				-0.29	0.11	0.01

N observations: 51,098

Table 32 – Interaction terms between ethnic background and economic activity

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.64	0.01	<0.001	0.65	0.01	<0.001
Ethnicity (ref: white British)						
African	1.18	0.07	<0.001	1.12	0.08	<0.001
Any other	0.97	0.08	<0.001	0.94	0.10	<0.001
Bangladeshi	0.63	0.06	<0.001	0.44	0.07	<0.001
Caribbean	0.64	0.07	<0.001	0.54	0.08	<0.001
Indian	0.39	0.05	<0.001	0.36	0.06	<0.001
Irish	0.06	0.08	0.45	0.09	0.10	0.37
Mixed	0.35	0.07	<0.001	0.34	0.08	<0.001
Other Asian	0.70	0.07	<0.001	0.71	0.08	<0.001
Other white	0.48	0.06	<0.001	0.48	0.07	<0.001
Pakistani	0.48	0.05	<0.001	0.32	0.07	<0.001
Economic activity (ref: Active)						
Inactive	-0.28	0.02	<0.001	-0.32	0.02	<0.001
Interaction terms						
African:Inactive				0.29	0.19	0.12
Any other:Inactive				0.11	0.18	0.57
Bangladeshi:Inactive				0.65	0.14	<0.001
Caribbean:Inactive				0.32	0.15	0.03
Indian:Inactive				0.11	0.11	0.33
Irish:Inactive				-0.08	0.16	0.63
Mixed:Inactive				0.02	0.18	0.93
Other Asian:Inactive				-0.09	0.18	0.60
Other white:Inactive				<0.001	0.15	0.98
Pakistani:Inactive				0.47	0.12	<0.001

N observations: 51,089

Table 33 – Interaction terms between immigration background and sex

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.46	0.02	<0.001	0.44	0.02	<0.001
Migration background (ref. gen 3)						
gen 2.5	0.06	0.04	0.16	0.14	0.07	0.04
gen 2	0.37	0.04	<0.001	0.47	0.06	<0.001
gen 1.5	0.63	0.05	<0.001	0.66	0.07	<0.001
gen 1	0.80	0.03	<0.001	0.95	0.05	<0.001
Sex (ref: Male)						
Female	-0.02	0.02	0.29	0.03	0.02	0.24
Interaction terms						
gen 2.5:Female				-0.13	0.09	0.14
gen 2:Female				-0.17	0.08	0.04
gen 1.5:Female				-0.07	0.10	0.52
gen 1:Female				-0.26	0.06	<0.001
N observations: 47,713						

Table 34 – Interaction terms between immigration background and age group

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.14	0.01	<0.001	0.13	0.01	<0.001
Migration background (ref. gen 3)						
gen 2.5	0.07	0.05	0.14	0.07	0.06	0.23
gen 2	0.33	0.04	<0.001	0.39	0.06	<0.001
gen 1.5	0.64	0.05	<0.001	0.61	0.07	<0.001
gen 1	0.83	0.03	<0.001	0.89	0.04	<0.001
Age group (ref: 35-65)						
Over 65	0.86	0.03	<0.001	0.90	0.03	<0.001
Under 35	0.64	0.02	<0.001	0.65	0.03	<0.001
Interaction terms						
gen 2.5:Over 65				0.02	0.17	0.93
gen 2:Over 65				-0.20	0.37	0.58
gen 1.5:Over 65				-1.02	0.26	<0.001
gen 1:Over 65				-0.43	0.11	<0.001
gen 2.5:Under 35				<0.001	0.10	0.96
gen 2:Under 35				-0.09	0.08	0.26
gen 1.5:Under 35				0.18	0.11	0.11

gen 1:Under 35	-0.08	0.07	0.25

N observations: 47,714

Table 35 – Interaction terms between immigration background and English second language

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.42	0.01	<0.001	0.42	0.01	<0.001
Migration background (ref. gen 3)						
gen 2.5	0.06	0.04	0.22	0.04	0.05	0.36
gen 2	0.34	0.04	<0.001	0.36	0.05	<0.001
gen 1.5	0.58	0.06	<0.001	0.49	0.08	<0.001
gen 1	0.74	0.05	<0.001	0.84	0.07	<0.001
English first language (ref: Yes)						
No	0.07	0.05	0.13	0.34	0.15	0.02
Interaction terms						
gen 2.5:No	NA	NA	NA	0.63	0.40	0.11
gen 2:No	NA	NA	NA	-0.41	0.18	0.02
gen 1.5:No	NA	NA	NA	-0.10	0.18	0.56
gen 1:No	NA	NA	NA	-0.39	0.16	0.02

N observations: 46,162

Table 36 – Interaction terms between immigration background and area type

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.30	0.02	<0.001	0.33	0.02	<0.001
Migration background (ref. gen 3)						
gen 2.5	0.05	0.04	0.29	-0.03	0.09	0.74
gen 2	0.32	0.04	<0.001	0.05	0.18	0.77
gen 1.5	0.58	0.05	<0.001	-0.13	0.24	0.59
gen 1	0.75	0.03	<0.001	0.17	0.12	0.16
Area type (ref: rural)						
urban area	0.21	0.02	<0.001	0.18	0.02	<0.001
Interaction terms						
gen 2.5:urban area				0.10	0.11	0.33
gen 2:urban area				0.29	0.19	0.12
gen 1.5:urban area				0.74	0.25	<0.001
gen 1:urban area				0.62	0.13	<0.001

N observations: 47,715

Table 37 – Interaction terms between immigration background and level of formal education

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	1.14	0.03	<0.001	1.26	0.03	<0.001
Migration background (ref. gen 3)						
gen 2.5	0.15	0.04	<0.001	0.02	0.18	0.92
gen 2	0.52	0.04	<0.001	0.09	0.22	0.67
gen 1.5	0.65	0.05	<0.001	-0.12	0.13	0.37
gen 1	0.90	0.03	<0.001	0.31	0.08	<0.001
Level of formal education (ref: no qualification)						
below degree	-0.56	0.03	<0.001	-0.67	0.04	<0.001
degree or higher	-1.12	0.03	<0.001	-1.30	0.04	<0.001
Interaction terms						
gen 2.5:below degree				0.15	0.19	0.42
gen 2:below degree				0.39	0.23	0.09
gen 1.5:below degree				0.81	0.15	<0.001
gen 1:below degree				0.36	0.10	<0.001
gen 2.5:degree or higher				0.15	0.19	0.43
gen 2:degree or higher				0.51	0.23	0.03
gen 1.5:degree or higher				0.97	0.16	<0.001
gen 1:degree or higher				0.88	0.10	<0.001

N observations: 47,688

Table 38 – Interaction terms between immigration background and housing tenure

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.24	0.01	<0.001	0.24	0.01	<0.001
Migration background (ref. gen 3)						
gen 2.5	0.02	0.05	0.66	<0.001	0.05	0.94
gen 2	0.31	0.04	<0.001	0.41	0.05	<0.001
gen 1.5	0.48	0.05	<0.001	0.41	0.06	<0.001
gen 1	0.51	0.03	<0.001	0.46	0.04	<0.001
Tenure (ref: Onwer or other)						
Private renter furnished or employer	2.19	0.08	<0.001	2.31	0.12	<0.001
Private renter unfurnished	0.61	0.04	<0.001	0.57	0.04	<0.001
Social renter	0.75	0.03	<0.001	0.75	0.03	<0.001
Interaction terms						
gen 2.5:Private renter furnished or employer				0.54	0.47	0.26
gen 2:Private renter furnished or employer				-1.04	0.26	<0.001
gen 1.5:Private renter furnished or employer				-0.06	0.34	0.87
gen 1:Private renter furnished or employer				-0.10	0.18	0.59
gen 2.5:Private renter unfurnished				-0.20	0.16	0.20
gen 2:Private renter unfurnished				-0.17	0.17	0.31
gen 1.5:Private renter unfurnished				0.40	0.23	0.08
gen 1:Private renter unfurnished				0.37	0.11	<0.001
gen 2.5:Social renter				0.18	0.13	0.17
gen 2:Social renter				-0.34	0.11	<0.001
gen 1.5:Social renter				0.19	0.13	0.14
gen 1:Social renter				0.07	0.09	0.44
N observations: 47,668						

Table 39 – Interaction terms between immigration background and having dependent children

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.42	0.01	<0.001	0.40	0.01	<0.001
Migration background (ref. gen 3)						
gen 2.5	0.06	0.04	0.17	0.09	0.05	0.07
gen 2	0.36	0.04	<0.001	0.55	0.05	<0.001
gen 1.5	0.62	0.05	<0.001	0.64	0.06	<0.001
gen 1	0.78	0.03	<0.001	0.86	0.04	<0.001
Dependent children (ref: No)						
Yes	0.12	0.02	<0.001	0.20	0.03	<0.001
Interaction terms						
gen 2.5:Yes				-0.13	0.10	0.19
gen 2:Yes				-0.51	0.08	<0.001
gen 1.5:Yes				-0.09	0.11	0.40
gen 1:Yes				-0.20	0.07	<0.001

N observations: 47,715

Table 40 – Interaction terms between immigration background and economic activity

	Mean-effect model			Model with interaction terms		
	Log odds	**SE**	**p-value**	**Log odds**	**SE**	**p-value**
(Intercept)	0.56	0.01	<0.001	0.56	0.01	<0.001
Migration background (ref. gen 3)						
gen 2.5	0.04	0.04	0.42	0.07	0.05	0.20
gen 2	0.31	0.04	<0.001	0.31	0.05	<0.001
gen 1.5	0.60	0.05	<0.001	0.71	0.06	<0.001
gen 1	0.78	0.03	<0.001	0.72	0.04	<0.001
Economic activity (ref: Active)						
Inactive	-0.26	0.02	<0.001	-0.26	0.02	<0.001
Interaction terms						
gen 2.5:Inactive				-0.11	0.10	0.25
gen 2:Inactive				<0.001	0.11	0.97
gen 1.5:Inactive				-0.36	0.11	<0.001
gen 1:Inactive				0.21	0.07	<0.001

N observations: 47,712