

RESEARCH PROJECTS

Infectious Intestinal Disease Survey During COVID-19: Additional Analysis on Association of Behaviours With Illness

Mark Jitlal¹, Vinotharan Annarasa¹, Darren Holland¹

¹ Food Standards Agency

Keywords: Foodborne Disease and AMR, Behaviour and perception, Foodborne disease, Modelling

<https://doi.org/10.46756/001c.122449>

FSA Research and Evidence

During the COVID-19 pandemic businesses and individuals changed hygiene, eating, travelling and social behaviours. These changes have the potential to play an important role in the likelihood of people contracting diseases, including Infectious Intestinal Disease (IID). Six waves of a nationally representative survey were carried out to gather information about IID among the general population during this period. Questions were also asked about the behaviours of those who got IID and of a subset of individuals without IID. Separate surveys were run for adults and children (with parents responding on behalf of their children). Waves 1 to 4 were run for both adults and children, wave 5 was just for adults and wave 6 was just for children. Logistic regression models were used to examine the association between the behaviours and IID, for individual waves and for a combined all-waves analysis, for adults and children separately. Adults eating food from takeaways or street food vendors was the single behaviour positively associated with IID across all five waves, whereas two other behaviours were positively associated with IID in four of the five waves: buying ready to eat food outside work/school; and eating food from work/school canteen. All three of these behaviours were among those selected in the model using data from all five waves. For the children's analysis there were few behaviours selected in any of the individual models and those that were selected only featured in one of the five models. For the children's all-waves model the strongest effects were: use of public transport more than weekly; anybody leaving the house; and children eating ready to eat food outside of schools. In the hypothetical scenario of entirely removing eating food from takeaways or street food vendors for adults, the expected reduction in IID cases is 9-24%.

Executive Summary

The COVID-19 pandemic and the associated contingency measures taken by the government, businesses and individuals changed many of our behaviours including hygiene, eating habits, travelling habits and social

interactions. All these changes have the potential to play an important role in the likelihood of people contracting diseases, including Infectious Intestinal Disease (IID).

To investigate this the FSA commissioned Ipsos MORI to undertake six waves of a nationally representative online panel survey to gather information about IID among the general population during this period. Questions were also asked about the behaviours of those who got IID and of a comparison group of individuals who did not get IID. Separate surveys were run for adults and children (with parents/guardians responding on behalf of children). Waves 1 to 4 were run for both adults and children, wave 5 was just for adults and wave 6 was just for children.

This report covers analysis undertaken internally by the FSA's analytics unit using the data from these surveys. This involved producing logistic regression models to evaluate the behaviours together, rather than analysing them individually against IID which was covered in the main Ipsos report. This additional work was undertaken as it is likely some of the behaviours may be correlated.

Separate models were produced for adults and children for each wave. Two additional models were produced (one for adults and one for children) combining data from all five waves.

From the adult data one behaviour was selected in all five models as being associated with a higher likelihood of getting IID. This was:

- Eating food from takeaways or street food vendors in the previous four weeks.

In addition, there were two other behaviours selected in four out of the five models, as being associated with higher likelihood of IID, these being:

- Have bought ready to eat food outside the workplace, school or university in the prior four weeks.
- Eating food from work, school or university canteen in the prior four weeks.

All three of these behaviours were among those selected in the model using data from all five waves. Note these associations do not necessarily infer causation and as such may just be indicators/proxies of other causes such as likelihood of mixing with people outside the household.

For the children analysis there were few behaviours selected in any of the individual models and those that were only featured in one of the five models. For the children's all-waves model the strongest effects were: use of public transport more than weekly; anybody leaving the house; and children eating ready to eat food outside of schools or childcare.

Further analysis on the adults' datasets found that the entire removal of eating food from takeaways or street food vendors indicates an expected reduction in IID cases by 9-24%. For the behaviour of buying ready to eat food outside the workplace, school or university resulted in a 7-14% reduction in IID; and for eating food from work, school or university canteens this results in a 6-8% reduction.

For adults with IID further analysis was undertaken to see if there were any significant differences in the likelihood of seeking medical care depending on variations in demographics. This found that men were more likely to seek healthcare than women. This was true for four waves, the increased odds ranging from 43% to 124%. The other notable result was that other than white ethnicities had a more than doubled odds of seeking healthcare compared to white ethnicity in waves 1 and 3, but not in the other three waves.

1. Introduction

During the COVID-19 pandemic individuals changed many of their behavioural habits, including hygiene, eating habits, travelling habits and social interactions. All these changes have the potential to play a role in the likelihood of people contracting diseases, including Infectious Intestinal Disease (IID).

The FSA commissioned Ipsos MORI to undertake six waves of nationally representative online panel surveys to gather information about IID among the general population during this period. Separate surveys for adults and children (with parents/guardians responding on behalf of children) were designed and administered. Throughout this report the term parents is used to refer to both parents and guardians. Details of this work and the results can be found in "Survey of infectious intestinal disease in the UK: Effect of COVID-19 response and associated measures on IID in the UK".

As part of these surveys a series of questions on behaviours were asked to assess whether there were any relationships between these and the likelihood of someone having IID. In the aforementioned report such relationships were considered between IID and each of these behaviours individually.

It is likely that some behaviours will be correlated with each other. This report describes more in-depth analysis where all the behaviours and demographics collected during the surveys were considered together and a best fitted model for each wave produced. It should be stated from the outset that these models show associations rather than causation.

In addition, for those who had IID, further analysis was undertaken to see if there were significant differences in medical care seeking behaviour depending on variations in demographics.

2. Materials and Methods

2.1. Data Collection

2.1.1. Survey Methodology

Previous self-reported survey research with the public about IID symptoms has been carried out by telephone using Random Digit Dialling (Viviani et al., 2016). In contrast to the previous studies, an online panel survey has been utilised in this study. The move to an online panel survey approach has the benefits of being faster, more cost effective, more flexible and more suited to capturing self-reported findings on a sensitive topic.

Quota sampling was used to achieve representativeness of the UK population according to key characteristics:

- Age [six adult age bands: 16-24; 25-34; 35-44; 45-54; 55-64; 65+];
- Gender;
- Region [12 regions: East Midlands; East of England; Greater London; North East; North West; Northern Ireland; Scotland; South East; South West; Wales; West Midlands; Yorkshire & Humberside];
- Social grade [4 levels: higher and intermediate managerial, administrative and professional occupations (AB); supervisory, clerical and junior managerial, administrative and professional occupations (C1); skilled manual occupations (C2); and semi-skilled and unskilled manual and lowest grade occupations (DE)].

For the children survey, adult panellists were screened for the presence of dependent children in their household, and the selected adults answered questions on behalf of under 16-year-old children in the surveys (age bands: 0-4; 5-9; 10-15 years).

It was assumed that around 11,000 respondents sampled would be sufficient to answer the research questions, 9,000 in the adults' survey and 2,000 in the children's survey. A previous study that used a telephone

survey with 28-day recall had an IID incidence rate of 0.533 per person-years (Viviani et al., 2016). This equates to an incidence rate of 0.0409 per person-28 days. From this we would anticipate 450 IID cases during a 28 day period from 11,000 respondents.

Quota sampling was used to ensure that the original sample was representative of the UK population according to the aforementioned quota variables. All survey participants were asked if they had vomiting or diarrhoea (V/D) in the previous four weeks. If they reported diarrhoea only, a follow-on question was asked whether they had three or more instances over a 24 hour period. If they met these criteria or if they reported any instances of vomiting the participants went on to answer the full survey.

Quota sampling was then used again to select a control group amongst those who had no such symptoms. A similar number of controls to those who answered the full survey were selected for each wave. Quota sampling led to weights being applied such that respondents were representative of the UK population (with respect to the quota variables). Those in the control group were asked a set of questions relating to behaviour. Participants who did not have V/D in the previous four weeks or were not selected for the control group were asked no further questions.

2.1.2. Questionnaire development

Ipsos MORI worked closely with the FSA to develop a suitable questionnaire for the study, which included questions from previous IID research, alongside new questions designed to meet the aims of the current study. The adult questionnaire, once finalised, was then adapted for parents to answer on behalf of their child.

Due to the dynamic nature of the pandemic and due to the urgency to collect data to build an evidence base, the questionnaire evolved between the waves.

2.1.3. Timelines

The fieldwork dates for the six surveys were:

- Wave 1: 27th August – 17th September 2020
- Wave 2: 2nd December – 18th December 2020
- Wave 3: 15th February – 3rd March 2021
- Wave 4: 26th August – 20th September 2021
- Wave 5: 9th December 2021 – 5th January 2022 (survey conducted for adults only)

- Wave 6: 15th February – 10th March 2022 (survey conducted for children only)

The emergence of the Coronavirus Omicron variant in late 2021 brought about the possibility of new government restrictions being introduced. It was decided to bring forward Wave 5 of the survey from February/March 2022 to December 2021/January 2022 to try to run the survey outside any further restrictions. This was a late decision in response to a changing situation and as such it was only possible to run the adults survey at that time. As the Omicron variant was less severe than feared, it was possible to run a separate children's survey as previously planned in February/March 2022. For simplicity, wave 6 of the children only survey will henceforth be referred to as wave 5.

2.2. Methodology

2.2.1. Data cleaning and formatting

Many of the questions in the survey include "Don't know" and "Prefer not to say" options. Whilst it is important to have these options in the questionnaire, they do not necessarily add any value to the statistical analysis. Where participants have selected these responses for any of the questions used in the analysis they have been removed from the datasets. This is because best subset selection models require non-missing data for all variables used in the modelling process.

For some questions two or more options have been grouped together to increase numbers in each group and reduce the number of possible options.

Respondents who gave contradictory responses to behaviours (such as stating that they had not left their home in the previous four weeks, but also stated they had eaten at restaurants or workplace eateries in the same time period) were excluded from the IID models. Additional exclusions for the IID models included those who reported their gender as being other than female or male (due to small numbers), respondents who didn't indicate their ethnicity, and those who didn't indicate how frequently they wash their hands under a number of scenarios.

As the seeking healthcare models were restricted to those respondents with IID, exclusions based on issues with responses related to behaviours were not required. Therefore respondents who gave inconsistent answers or "Don't know" / "Prefer not to say" responses to behaviours were not removed for the seeking healthcare analyses. However, those respondents who gave contradictory responses by not declaring their source of seeking healthcare while separately indicating that they had sought healthcare for IID, were excluded from these latter analyses.

2.2.2. Explanatory variables

The variables used in the statistical analysis of the adults' data is given in [Table 1](#). Other than for demographic details the question number in the survey has been used as the variable name. The variables utilised for the statistical analysis of the adults' and children's datasets are largely similar. In the children's data the variables relate to the child the parent is completing the survey for, rather than the respondent themselves. The exception is social grade which is that of the parent. Any other differences between the adults' and children's variables are highlighted in [Table 2](#). All behaviour variables relate to the previous four weeks.

Details of the different subgroup levels and reference categories for each variable, for use in the modelling, can be found in Appendix A Tables A1-A2. Note that Q5_4 for adults and Q5_5 for children have their baseline category switched from "No" to "Yes", ensuring that the reference category is the theoretically least risky behaviour, as these two questions are effectively asking about negative behaviours.

2.2.3. Domestic IID definition

Domestic IID was defined in three ways as follows:

Definition 1: Domestic IID cases (baseline definition): amongst those with V/D symptoms, not all were considered to be IID cases. Exclusions included those who answered specific questions as follows:

- If the V/D symptoms were thought to be caused by any of the following: pre-existing illness that causes regular sickness and / or diarrhoea; medication; morning sickness (due to pregnancy); hangover - as a result of alcohol consumption; other (for these responses further details were given in a text box and these were accessed by three people. A decision was then made as to the cause was likely to be IID or non-IID); or those who preferred not to say;
- If the respondent had tested positive for COVID-19 around the time that their V/D symptoms had started (or the respondent didn't know or preferred not to say). This is because diarrhoea can also be a symptom of COVID-19;
- If the respondent had travelled outside the UK in the two weeks prior to V/D symptoms starting (or the respondent preferred not to say).

Definition 2: Following waves 1 and 2 it became clear that a proportion of those who had answered "other" as the likely cause of their vomiting and diarrhoea put down some form of food hypersensitivity in their detailed

Table 1. Description of each explanatory variable for adults' data.

Variable Name	Description
Age	Age of the respondent
Gender	Gender of the respondent
Region	Region that respondent resides in
Social_grade	Social grade of the respondent. Parent's social grade is used in the children's dataset.
Ethnicity	Ethnicity of respondent
People_household	Number of people in the respondent's household
Q5_1	Has the respondent left the house to go to work, school, university, or another childcare setting?
Q5_2	Has other adult members in the respondent's house left to go to work, school, university, or another childcare setting?
Q5_3	Has other household members under the age of 16 left the house to go to work, school, university, or another childcare setting?
Q5_4	No one in the household has left the house to go to work, school, university, or another childcare setting.
Q6_1	When going to work, school, or university, I have taken packed lunches from home.
Q6_2	When going to work, school, or university, I have had food provided in the setting, e.g., cafeteria food.
Q6_3	When going to work, school, or university, I have bought 'ready to eat' food, e.g. pre-packed sandwiches from outside the setting.
Q6_4	When going to work, school, or university, I have not eaten meals in these settings.
Q7	Have you left the house? This could be for any reason, not just to attend work, school, or university.
Q8	On average, how long have you spent outside of your home on the days you have gone out?
Q9	How many times, if at all, have you personally used public transport?
Q10_1	How often have you eaten from takeaways or street food vendors?
Q10_2	How often have you eaten from restaurants (eating in)? Note: As restaurants were closed during the 3 rd wave, this variable has been excluded from the analysis for the 3 rd wave.
Q10_3	How often have you eaten from work/school/university canteen?
Q10_4	How often have you eaten food prepared in someone else's house?
Q10_5	How often have you eaten food from organised meal delivery services, e.g., 'meals on wheels'?
Q11_1	How often do you wash your hands when you get home after a trip out of your home?
Q11_2	How often do you wash your hands after blowing your nose, sneezing, or coughing into your hands?
Q11_3	How often do you wash your hands after going to the toilet?
Q11_4	How often do you wash your hands before cooking a meal or preparing food?
Q11_5	How often do you wash your hands before eating?

response. We had previously assumed that this would have been captured by the pre-existing illness option, but clearly in some circumstances this was not the case. Therefore, in subsequent waves respondents were asked a follow-up question specifically relating to whether they thought that their symptoms were caused by a food hypersensitivity. Thus, for waves 3 to 5 we were able to undertake a sensitivity analysis with the following extra exclusion applied:

Table 2. Description of each explanatory variable in the children's data, where this differs from the adults' data.

Variable Name	Description
Q5_1	Has the respondent's parent/guardian left the house to go to work, school, university, or another childcare setting?
Q5_3	Has the child left the house to go to work, school, university, or another childcare setting?
Q5_4	Has other household members under the age of 16 left the house to go to work, school, university, or another childcare setting?
Q5_5	No one in the household has left the house to go to work, school, university, or another childcare setting.

- Respondents who thought that their V/D symptoms were caused by a food allergy, a food intolerance, coeliac disease or preferred not to say were excluded from the analyses.

This is referred to as domestic IID (strict hypersensitivity exclusion) in the rest of the report. Note in waves 3 to 5, we found that between 61% to 67% of adults and 61% to 78% of children, whose illness may have been related to food hypersensitivities, had already been excluded under the previous definition.

Definition 3: Finally, as COVID-19 travel restrictions eased in later waves (4-5) we asked an additional question as to whether respondents had travelled outside the UK in the previous six weeks (asked of both potential cases and controls; those who preferred not to say were excluded). The condition of travelling outside the UK superseded the travel condition imposed in previous definitions of domestic IID. This allowed further sensitivity analysis of the models with this exclusion criteria applied to both the cases and comparison groups. This is referred to as domestic IID (strict hypersensitivity and travel exclusions) in the rest of the report.

2.3. Statistical methods

2.3.1. Behaviour models for domestic IID

Models for domestic IID were considered separately for adults and children, for each of the five waves. In total there were 20 IID models across the individual waves (waves 1-2: one model each; wave 3: two models; waves 4-5: three models each). The additional models in waves 3 onwards were for the sensitivity analysis on the different definitions of domestic IID. In this case we assessed how the variables selected under the three IID definitions varied and whether there were any marked differences in effect sizes.

For the adult data, we used a logistic regression modelling approach for the domestic IID response. We modelled various independent variables with IID, using the best fitted model to reduce the impact of overfitting models. The aim of the modelling was to ascertain the key variables associated

with domestic IID across the waves. The variables ([Tables 1-2](#)) included demographic factors (gender, age, region, social grade, ethnicity and number of people in household) as well as behavioural factors (leaving the house for work or education; food eaten out during the work/school day; and the frequency of the following: leaving the house for any reason; duration of leaving the house; use of public transport; eating food prepared outside the house; and handwashing behaviour). As quota variables were used in the selection process, we used unweighted logistic regression models, adjusting for the quota variables (age, gender, region and social grade). Age was modelled as a categorical variable as it was found to have a non-linear relationship with the log odds of domestic IID; therefore, all variables used in the analysis are categorical. Other assumptions we checked for the logistic regression models include: the responses are binary (IID case or not); the observations within each wave are independent of one another (this can't be formally checked, but we assume the members of each sample are independent of one another); the independent variables exhibit no obvious multicollinearity; and there are a sufficient number of cases in order to model the parameters in each model.

We attempted to use a best subset selection approach for all variables, but this proved to be computationally infeasible (there are at least 23 independent variables beyond the quota variables, resulting in over 8 million potential models). Instead, we fitted three (automated) stepwise regression models (forwards, backwards and both approaches) and any variables found in any of these three models were then considered in a best subset selection approach, using AIC (Akaike information criterion) as the criteria of best model fit.

As it is feasible that the above stepwise regression approach may miss important variables, from the pool of variables not in the best subset selection model, we added each in turn to the final selected model to ascertain if the fit was improved, by AIC. We checked the final fitted model for multicollinearity by assessing the Variance Inflation Factor (VIF) for each variable. We dropped any terms that were collinear with other variables, using $VIF < 3$ as a threshold (Zuur et al., 2010), though others have used less stringent thresholds (James et al., 2021 recommend $VIF < 5$). The confidence intervals for the odds ratios were calculated using the profile likelihood.

We also fitted a Lasso (Least Absolute Shrinkage and Selection Operator) model to assess whether similar independent variables were selected, compared to the best subset selection approach. From these models we only give the odds ratio, without its confidence interval, as the standard error is difficult to estimate (Tibshirani, 1996). The Lasso model is a shrinkage method used for variable selection (Tibshirani, 1996). It allows

coefficients to be regularised and shrunk to zero, to avoid overfitting. The Lasso model uses the weighted data, and no variables were coerced into any models. The Lasso model uses k-fold cross-validation, meaning that the dataset is split into k random groups of approximately equal size, and each group in turn is used as the test data once, with the other k-1 groups acting as the training data. This results in k estimates of the mean square error, and the cross-validated error is the mean of these k errors (James et al., 2021). In our case, we used 10-fold cross-validation.

In addition to the data, the key input into the Lasso model is the lambda tuning parameter, which identifies the variance-bias trade-off (or trade-off between over-fitting and under-fitting the model). The higher the value of lambda, the fewer variables will be selected. The best values of lambda can be estimated via cross-validation for predictive purposes. The optimal solution (lambda-min) is sometimes used, though this has been criticised for resulting in overfitted models; instead it has been suggested using the most parsimonious model whose prediction error is within one standard error of the cross-validated error of the best model (lambda-1se) (Breiman et al., 1984; Hastie et al., 2009; Krstajic et al., 2014); this was the approach we used.

For the children's individual wave data, there were insufficient number of IID cases (often < 150, and, as with the adult data, always lower than the number of controls) and so we did not fit any best subset selection models to this data, due to overfitting issues (Harrell, 2001); (Riley et al. (2019) suggest a more nuanced approach). Instead, we ran a Lasso model (using weighted data), using the most parsimonious model whose prediction error is within one standard error of the cross-validated error of the best model (lambda-1se), as with the adult data. Due to low numbers, we retained in the statistical analysis those respondents who answered "Not applicable" to questions relating to their children washing their hands. However, we do not present any results relating to these uninformative parameters.

In addition to the models for the individual waves we produced two all-waves models (one for adults and one for children) using best subset selection (logistic regression). As selected panellists (respondents) were not identified by Ipsos at each wave, strictly speaking we cannot treat such samples as being independent across waves. However, as approximately 9,000 adults were sampled (to assess adults, not children) at each wave from an approximate panel size of 325,000 (the numbers for both changes over time), then the probability of a given individual being randomly selected for 2+ waves of the adult survey is approximately 0.7%. About 2,000 adults were surveyed in each wave for the behaviour of children in their household and assuming that half of the households in the panel have children aged under 16 years, the probability of such an individual

being randomly selected for 2+ waves is approximately 0.2%. For the combined wave analysis of the children's data we excluded the "Not Applicable" responses for the handwashing behaviours. Note that we did not use combined data for weighted analyses as the weights for the individual waves would need to be recalibrated for the combined dataset. As Q10_2 (how often have you eaten food from at a restaurant in the last 4 weeks) was not asked in wave 3, as restaurants were closed during this period, it was excluded from the all-waves analysis.

We were not able to run the sensitivity analysis for the different definitions of IID for the all-waves models, as the necessary information to produce these definitions was not gathered in earlier waves.

2.3.2. Population Proportional Attributable Risk

Population Proportional Attributable Risk (PPAR) was calculated for the adult models for domestic IID (baseline definition). PPAR is used to assess how much the outcome would be reduced by if an exposure were to be entirely removed. PPAR is typically used with Relative Risk metrics, but we have used Odds Ratios via the case-control study design. This has two main assumptions:

- the prevalence of our exposures in the sample (of cases and the selected controls) truly reflect that in the general population; and
- the prevalence of IID for the given time period (prior four weeks) in the population is relatively small (as the PPAR is based upon risk and not odds, but these are similar when the prevalence of disease is small). For example, a risk of 5% results in an odds of 5.3%; a risk of 10% results in an odds of 11.1%; and a risk of 20% results in an odds of 25%.

The below PPAR calculation adjusts for multiple risk factors in an adjusted logistic regression model (Bruzzi et al., 1985; Sjölander & Vansteelandt, 2011):

$$PPAR = 1 - \frac{Pr(Y_0 = 1)}{Pr(Y = 1)}$$

Where $Pr(Y_0 = 1)$ refers to the counterfactual, that is the theoretical probability of the outcome when the behaviour has been eliminated from the population and $Pr(Y = 1)$ is the observed probability of the outcome.

The PPAR estimates are multiplied by 100 to convert the proportion to a percentage. Note that we only present the PPAR estimate for those parameters with a positive odds ratio (>1), as removal of a negative exposure does not make sense for what we are trying to ascertain. Only

the PPAR estimates are presented, as confidence intervals using the Wald Statistic (resulting in symmetric confidence intervals) could conceivably fall outside the (-100, +100)% boundary.

PPAR was also assessed for the all-waves analysis for adults and children, separately, for those variables found in the best fit model as described above.

2.3.3. Seeking healthcare

Amongst those individuals with domestic IID (baseline definition) we used weighted logistic regression to model seeking healthcare and the association that this has with the various demographic characteristics (age, gender, social grade, ethnicity, no. of people in household and country (England vs. other, rather than region, due to the low number of IID responses in some regions)), amongst adult cases for each wave. The sources of seeking healthcare were: via the GP (in-person; online or over the telephone); consulting a pharmacist; visiting A&E at hospital; speaking to somebody after contacting 111; visiting NHS or other medical websites; or via another form of medical advice.

This analysis was only undertaken for adults as there were insufficient number of IID cases for children.

2.3.4. Statistical software

We used R (v4.2.2) to clean and model the data. We used the step command (stats package) to automate the stepwise regression procedure and then double-checked using the dredge command (MuMIn package) to select the best fitted model (based on AIC). We checked for multicollinearity using the car package. For Lasso we used the glmnet package and for PPAR we used the AF package.

3. Results

All references to domestic IID in the results section are to the baseline definition, unless specified otherwise. In all tables when a variable is not fitted for the relevant model, it is simply excluded from that table. Similarly, if a variable is not fitted across a series of waves indicated in a table that variable will also be excluded from the given table.

3.1. Behavioural models domestic IID

The percentage of those adults in the study with IID is split by demographics for each wave in [Table 3](#). Though region is used in the modelling, for reasons of space, we present the breakdown by country instead.

The three stepwise regression models contained exactly the same variables in 9 out of 10 of the adult wave scenarios (this includes the sensitivity analysis using slightly different definitions of domestic IID) and the best subset selection model was always the same as a model found in one of the stepwise regression models for each wave scenario.

[Table 4](#) gives the odds ratios for the best fitted model effects (excluding the quota variables) on domestic IID for the five waves of adult data. We can see that Q10_1 (eating from takeaways or street food vendors in the previous four weeks) is the sole factor which is included in all five models, with three factors included in four models, Q6_3 (have bought ready to eat food outside the workplace or school in the prior four weeks), Q7 (frequency of leaving the house for any reason in the past four weeks) and Q10_3 (eating at work, school or university canteen in the prior four weeks). Q10_1 has a consistently positive relationship with domestic IID, as does Q6_3 and Q10_3. However, Q7 has a negative association with domestic IID for waves 2, 3 and 5, but is very strongly positive in wave 4. The latter result is primarily due to the reference category for Q7 only having one respondent from 18 (6%) with domestic IID, as opposed to the other two levels which have 85 (43%) and 511 (40%) respondents with domestic IID, resulting in very wide confidence intervals.

Appendix B Table B1 gives the odds ratios for the best fitting models for the sensitivity analysis, namely domestic IID (strict hypersensitivities exclusions) (waves 3-5) and domestic IID (strict hypersensitivities and travel exclusions) (waves 4-5), for just the non-quota variables. Parameter estimates are broadly similar for Q10_1, Q6_3, Q7 and Q10_3 under the three definitions of domestic IID. This shows that the slightly different definitions of domestic IID have no material impact on the overall results or conclusions.

There is little evidence of multicollinearity amongst the variables in the above logistic regression models, assessed by VIF. However, for wave 3, Q5_1 has a VIF ~ 2.5 for both models, indicating higher multicollinear relationship with the other fitted variables. However, as the criteria of VIF < 3 for all covariates in the models is met, Q5_1 remains in the respective models.

3.1.1. Lasso (Least Absolute Shrinkage and Selection Operator)

For the adults' models the variables selected from the Lasso models are given in the following tables: for domestic IID (Appendix B Table B2), domestic IID (strict hypersensitivity exclusions) (Appendix B Table B3) and domestic IID (strict hypersensitivity and travel exclusions) (Appendix B Table B3). From Table B2 we observe the variables selected for the most parsimonious models (λ_{1se}), with Q10_1 and Q10_3 are selected

Table 3. Percentage breakdown (no. of cases) for each demographic variable with domestic IID (baseline definition) for those adults in the studies for waves 1-5.

Demographics: levels	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Overall (N)	436	454	390	597	489
Age: 16-24	34.6% (47)	42.8% (80)	31.6% (37)	45.8% (54)	41.3% (62)
Age: 25-34	46.7% (105)	40.1% (105)	33.8% (78)	45.4% (129)	41.7% (110)
Age: 35-44	41.7% (91)	35.7% (79)	35.9% (84)	44.7% (115)	41.3% (104)
Age: 45-54	35.6% (79)	28.5% (65)	31.9% (79)	39.4% (109)	32.5% (79)
Age: 55-64	32.8% (64)	33.8% (75)	32.4% (69)	39.6% (103)	28.1% (59)
Age: 65+	20.3% (50)	18.9% (50)	16.3% (43)	28.2% (87)	25.3% (75)
Gender: Female	38.4% (241)	33.5% (233)	30.3% (224)	40.3% (323)	34.8% (233)
Gender: Male	31.8% (195)	32.1% (221)	29.3% (166)	39.0% (274)	34.3% (256)
Social Grade: AB	41.0% (158)	37.8% (155)	29.3% (112)	40.7% (185)	32.0% (127)
Social Grade: C1	35.7% (126)	32.1% (126)	29.3% (115)	39.7% (174)	32.0% (125)
Social Grade: C2	27.9% (61)	28.6% (80)	28.3% (70)	41.5% (112)	38.2% (104)
Social Grade: DE	31.9% (91)	30.7% (93)	32.6% (93)	36.7% (126)	37.5% (133)
Ethnicity: White	34.0% (386)	32.4% (409)	30.1% (359)	39.4% (538)	34.3% (433)
Ethnicity: Other	46.7% (50)	36.6% (45)	27.2% (31)	42.1% (59)	37.1% (56)
No. people in household: 1	26.9% (70)	26.3% (67)	26.0% (66)	38.6% (114)	27.5% (73)
No. people in household: 2+	37.3% (366)	34.2% (387)	30.8% (324)	39.9% (483)	36.2% (416)
Country: England	34.9% (364)	31.9% (368)	29.2% (314)	39.6% (496)	34.2% (412)
Country: Scotland, Wales & NI	36.4% (72)	37.2% (86)	32.8% (76)	40.2% (101)	36.3% (77)

for all waves and that Q6_3 and Q10_5 are selected in four waves (and largely the same variables are selected when the additional conditions to domestic IID are applied – Table B3). If we focus on the domestic IID models (Tables 4 and B2) those variables selected in the best subset selection models and those in the parsimonious lambda Lasso models, then we find that Q10_1 is selected in both model types for all five waves, whereas Q10_3 and Q6_3 are selected for at least four waves for both model types and Q10_5 for at least three waves for both model types.

For the children models no parameter appears in more than a single wave amongst the selected variables in the domestic IID Lasso model (Appendix B Table B4). Only three variables appear in any of the sensitivity Lasso models (Appendix B Table B5).

3.2. Population Proportional Attributable Risk

The variables that appear in the best subset selection model for adults waves 1-5, domestic IID (Table 4), are assessed for the Population Proportional Attributable Risk (Appendix B Table B6). For each wave the prevalence of domestic IID is approximately 5-10%. The entire removal of the Q10_1 behaviour indicates an expected reduction in IID cases by 9-24% (eating from takeaways or street food vendors in the previous four weeks),

Table 4. Odds ratios (95% CI) using best subset selection models for domestic IID (baseline definition) for adults across waves 1-5 for non-quota variables. Note that for independent factors that are not selected in the modelling for a specific wave this is given as “Excluded”.

Independent factor: levels	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Ethnicity: White	Excluded	Excluded	1	Excluded	Excluded
Ethnicity: Other	Excluded	Excluded	0.62 (0.37, 1.00)	Excluded	Excluded
No. people in household: 1	1	Excluded	Excluded	1	Excluded
No. people in household: 2+	1.30 (0.94, 1.82)	Excluded	Excluded	0.79 (0.59, 1.05)	Excluded
Q5_1 (adult left house for work/education): No	Excluded	Excluded	1	Excluded	Excluded
Q5_1 (adult left house for work/education): Yes	Excluded	Excluded	0.60 (0.39, 0.91)	Excluded	Excluded
Q5_2 (other adult left house): No	Excluded	Excluded	Excluded	Excluded	1
Q5_2 (other adult left house): Yes	Excluded	Excluded	Excluded	Excluded	1.25 (0.96, 1.61)
Q5_3 (child left house): No	1	Excluded	Excluded	1	1
Q5_3 (child left house): Yes	1.89 (1.04, 3.48)	Excluded	Excluded	2.15 (1.48, 3.14)	1.41 (1.02, 1.93)
Q5_4 (nobody left house): Yes	Excluded	Excluded	1	Excluded	Excluded
Q5_4 (nobody left house): No	Excluded	Excluded	1.40 (0.99, 1.98)	Excluded	Excluded
Q6_1 (home pack lunches): No	Excluded	1	Excluded	Excluded	Excluded
Q6_1 (home pack lunches): Yes	Excluded	1.48 (1.12, 1.94)	Excluded	Excluded	Excluded
Q6_2 (ate food at workplace): No	Excluded	Excluded	1	Excluded	Excluded
Q6_2 (ate food at workplace): Yes	Excluded	Excluded	2.24 (1.30, 3.86)	Excluded	Excluded
Q6_3 (ate food outside workplace): No	1	Excluded	1	1	1
Q6_3 (ate food outside workplace): Yes	1.45 (0.98, 2.14)	Excluded	3.13 (1.97, 5.02)	1.56 (1.13, 2.14)	1.69 (1.25, 2.27)
Q7 (left house for any reason): No	Excluded	1	1	1	1
Q7 (left house for any reason): Up to weekly	Excluded	0.70 (0.35,	0.65 (0.37,	9.58 (1.87,	0.91 (0.34,

Independent factor: levels	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
		1.44)	1.13)	175.69)	2.63)
Q7 (left house for any reason): More than weekly	Excluded	0.50 (0.26, 1.00)	0.47 (0.28, 0.83)	7.50 (1.50, 136.37)	0.52 (0.20, 1.45)
Q9 (public transport usage): None	Excluded	1	1	Excluded	Excluded
Q9 (public transport usage): Up to weekly	Excluded	1.48 (1.03, 2.12)	1.72 (1.13, 2.60)	Excluded	Excluded
Q9 (public transport usage): More than weekly	Excluded	1.80 (1.16, 2.78)	1.41 (0.80, 2.44)	Excluded	Excluded
Q10_1 (eaten takeaways or street food): Less than weekly	1	1	1	1	1
Q10_1 (eaten takeaways or street food): Weekly or more	1.75 (1.32, 2.32)	1.95 (1.50, 2.54)	1.68 (1.26, 2.25)	1.26 (0.98, 1.62)	1.56 (1.20, 2.03)
Q10_2 (eaten at restaurants): Less than weekly	Excluded	Excluded	Excluded	1	Excluded
Q10_2 (eaten at restaurants): Weekly or more	Excluded	Excluded	Excluded	1.68 (1.27, 2.22)	Excluded
Q10_3 (eaten at work/education canteen): Less than weekly	1	1	Excluded	1	1
Q10_3 (eaten at work/education canteen): Weekly or more	1.64 (1.02, 2.64)	1.58 (1.07, 2.33)	Excluded	1.46 (1.02, 2.08)	1.49 (1.05, 2.11)
Q10_4 (eaten food from other house): Less than weekly	1	Excluded	Excluded	1	1
Q10_4 (eaten food from other house): Weekly or more	1.43 (0.98, 2.09)	Excluded	Excluded	1.42 (1.02, 1.97)	1.34 (0.95, 1.89)
Q10_5 (eaten organised meal delivery food): Less than weekly	1	1	Excluded	Excluded	1
Q10_5 (eaten organised meal delivery food): Weekly or more	2.34 (1.19, 4.78)	1.88 (1.07, 3.37)	Excluded	Excluded	2.20 (1.24, 4.00)
Q11_1 (always wash hands on arrival): Yes	Excluded	Excluded	1	Excluded	Excluded
Q11_1 (always wash hands on arrival): No	Excluded	Excluded	0.79 (0.58, 1.08)	Excluded	Excluded
Q11_2 (always wash hands after blowing nose, sneezing or coughing into hands): Yes	Excluded	1	Excluded	1	Excluded
Q11_2 (always wash hands after blowing nose, sneezing or coughing into hands): No	Excluded	0.80 (0.62, 1.02)	Excluded	0.80 (0.63, 1.02)	Excluded
Q11_3 (always wash hands after toilet): Yes	1	Excluded	1	Excluded	Excluded

Independent factor: levels	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Q11_3 (always wash hands after toilet): No	1.49 (0.98, 2.24)	Excluded	1.91 (1.27, 2.85)	Excluded	Excluded
Q11_4 (always wash hands before cooking): Yes	Excluded	Excluded	Excluded	1	Excluded
Q11_4 (always wash hands before cooking): No	Excluded	Excluded	Excluded	1.33 (1.02, 1.74)	Excluded
Q11_5 (always wash hands before eating): Yes	1	Excluded	Excluded	Excluded	1
Q11_5 (always wash hands before eating): No	0.83 (0.63, 1.07)	Excluded	Excluded	Excluded	0.77 (0.61, 0.98)

whereas for Q6_3 (have bought ready to eat food outside the workplace or school in the prior four weeks) this results in a 7-14% reduction in IID and for Q10_3 (eating at work, school or university canteen in the prior four weeks) 6-8% reduction. Other parameters which appear in three of the models are reasonably consistent: Q5_3 (adults living with children in the same household attending school in the previous four weeks) 3-9% reduction; Q10_4 (consumption of food prepared in somebody else's house in the prior four weeks) 5-7% reduction; and Q10_5 (eating from meal delivery services in the prior four weeks) 4-6% reduction.

3.3. All-waves analysis

For adults the combined five wave analysis for domestic IID results in 12 variables selected (excluding the wave variable), more than for the individual waves (7-10 variables, after excluding Q10_2). However, for the variables in common we find that the results are broadly consistent with the findings in [Table 4](#), but with tighter confidence intervals. Namely, the strongest associations with domestic IID are for ([Table 5](#)):

- Q10_5 (eating from meal delivery services in the prior four weeks), where such consumption of at least weekly has an increased odds of domestic IID by 66% (95% CI: 28-116%) compared to those eating such food less than once per week;
- Q10_1 (eating from takeaways or street food vendors in the previous four weeks), where such consumption once per week or more frequently leads to a greater odds of domestic IID by 59% (95% CI: 41-79%) compared to consumption of less than once per week;
- Q6_3 (have bought ready to eat food outside the workplace or school in the prior four weeks), has an increased odds of domestic IID by 55% (95% CI: 33-80%) compared to those not eating such food;
- Q10_3 (eating at work, school or university canteen in the prior four weeks), where such consumption once per week or more frequently has a greater odds of domestic IID by 51% (95% CI: 26-80%) compared to those eating such food less frequently;
- Q5_3 (adults living with children in the same household attending school in the previous four weeks), has an increased odds of domestic IID by 36% (95% CI: 14-61%) compared to those adults with children in the same household who have not been attending school, or have no children in the household;

- Q9 (use of public transport in the previous four weeks), whereby those using it up to once per week have an increased odds of domestic IID of 28% (95% CI: 11-47%) and those using public transport more than weekly have an increased odds of domestic IID by 36% (95% CI: 14-63%), compared to those not using public transport in the previous four weeks;
- Q10_4 (consumption of food prepared in somebody else's house in the prior four weeks), where such consumption of at least once per week has an increase in odds by 32% (95% CI: 12-56%) compared to those eating such food less than once per week;
- Conversely, Q7 (those leaving the house for any reason), where such behaviour more than weekly in the previous four weeks have a reduced odds of domestic IID by 37% (95% CI: 13-54%), compared to those not leaving the house.

The elimination of the behaviour of eating from takeaways and fast food vendors indicates that domestic IID would be reduced by 16.8% ([Table 5](#)) amongst adults. These PPAR results indicate no other behaviour whose elimination may result in a reduction of domestic IID greater than 10%.

When the five waves were combined for children there were 544 domestic IID cases. This is a sufficient number of cases to run best subset selection for potentially all variables, under the assumption that we require at least ten events per parameter. The best fitted model is given in [Table 6](#). From this the strongest results are:

- Q9 indicating that those children that use public transport more than weekly have an increased odds of domestic IID by 87% (95% CI: 29-169%) compared to those that have not used public transport in the previous four weeks;
- Q5_5 indicating that somebody leaving the adult respondent's household in order to go to school or work in the previous four weeks results in an increase in the odds of domestic IID by 84% (95% CI: 32-158%);
- Q6_3 indicating that children eating ready to eat food outside of schools or childcare setting have an increased odds of domestic IID of 77% (95% CI: 26-148%);
- Q10_4, children eating food from another house weekly or more frequently increases the odds of domestic IID by 55% (95% CI: 20-100%);

Table 5. Odds ratios using best subset selection for domestic IID (baseline definition) for adults across all five waves combined for the non-quota variables; and Population Proportional Attributable Risk (PPAR) for the behaviours with positive effects. Note that for independent factors that are not behaviours the corresponding PPAR is “Not applicable”; for baseline reference categories this is given as “Baseline”; and for behaviours where the odds ratio is <1 this is given as “Negative”.

Independent factor: levels	Odds ratio (95% CI)	PPAR (%)
Ethnicity: White	1	Not applicable
Ethnicity: Other	0.86 (0.71, 1.04)	Not applicable
No. people in household: 1	1	Not applicable
No. people in household: 2+	1.12 (0.97, 1.29)	Not applicable
Q5_3 (child left house): No	1	Baseline
Q5_3 (child left house): Yes	1.36 (1.14, 1.61)	3.9
Q6_3 (ate food outside workplace): No	1	Baseline
Q6_3 (ate food outside workplace): Yes	1.55 (1.33, 1.80)	8.9
Q7 (left house for any reason): No	1	Baseline
Q7 (left house for any reason): Up to weekly	0.84 (0.60, 1.17)	Negative
Q7 (left house for any reason): More than weekly	0.63 (0.46, 0.87)	Negative
Q9 (public transport usage): None	1	Baseline
Q9 (public transport usage): Up to weekly	1.28 (1.11, 1.47)	4.8
Q9 (public transport usage): More than weekly	1.36 (1.14, 1.63)	4.4
Q10_1 (eaten takeaways or street food): Less than weekly	1	Baseline
Q10_1 (eaten takeaways or street food): Weekly or more	1.59 (1.41, 1.79)	16.8
Q10_3 (eaten at work/education canteen): Less than weekly	1	Baseline
Q10_3 (eaten at work/education canteen): Weekly or more	1.51 (1.26, 1.80)	6.7
Q10_4 (eaten food from other house): Less than weekly	1	Baseline
Q10_4 (eaten food from other house): Weekly or more	1.32 (1.12, 1.56)	4.4
Q10_5 (eaten organised meal delivery food): Less than weekly	1	Baseline
Q10_5 (eaten organised meal delivery food): Weekly or more	1.66 (1.28, 2.16)	3.8
Q11_2 (always wash hands after blowing nose, sneezing or coughing into hands): Yes	1	Baseline
Q11_2 (always wash hands after blowing nose, sneezing or coughing into hands): No	0.86 (0.77, 0.96)	Negative
Q11_3 (always wash hands after toilet): Yes	1	Baseline
Q11_3 (always wash hands after toilet): No	1.21 (1.03, 1.42)	2.8
Wave: 1	1	Not applicable
Wave: 2	0.86 (0.72, 1.02)	Not applicable
Wave: 3	0.79 (0.66, 0.94)	Not applicable
Wave: 4	1.11 (0.94, 1.31)	Not applicable

Independent factor: levels	Odds ratio (95% CI)	PPAR (%)
Wave: 5	0.86 (0.72, 1.02)	Not applicable

- Q10_1, children eating food from takeaways or street food vendors weekly or more frequently increases the odds of domestic IID by 51% (95% CI: 21-88%);
- Q11_3, children not always washing their hands after using the toilet increases the odds of domestic IID by 45% (95% CI: 16-83%);
- Conversely, Q7 indicates that those children leaving the house for any reason, more than weekly compared to never leaving, have a 49% (95% CI: 12-70%) reduced odds of domestic IID;
- Q5_2, other adults, in addition to those responding to the survey, leaving the house for work or education, reduces the odds of domestic IID by 36% (95% CI: 19-49%).

These effects were generally consistent with the results of the adult behaviours effects upon domestic IID. For example, eating from takeaways or street food vendors increases the odds of domestic IID by 59% in adults and 51% in children.

The removal of children eating from takeaways and fast food vendors indicates that domestic IID would be reduced by 16.7% amongst children ([Table 6](#)), remarkably similar to the finding for adults. The removal of children eating food from other homes may reduce domestic IID by 12.2% and if children always washed their hands after going to the toilet then domestic IID may be reduced by 11.7%. If nobody left the house for school or educational reasons then domestic IID may be reduced by 39.5%.

3.4. Seeking healthcare

The most consistent result across all five waves for adults with IID is that men have greater odds of seeking healthcare than women, ranging from an estimated relative increase of 43% to 124% ([Table 7](#)). The other notable result is that other ethnicities have more than a doubled odds of seeking healthcare compared to white ethnicity, but only in waves 1 and 3. The effect of age is not consistent over time, in that there's no particular pattern amongst age groups seeking healthcare within waves and across waves. For example, those aged 25-34 are more likely than those aged 16-24 to seek healthcare in wave 1 and to a lesser extent wave 3, but less likely in waves 4-5. No strong patterns in seeking healthcare emerge across waves for social grade, number of people in household and country.

Table 6. Odds ratios using best subset selection for domestic IID (baseline definition) for children across all five waves combined for the non-quota variables; and Population Proportional Attributable Risk (PPAR) for the behaviours with positive effects. Note that for independent factors that are not behaviours the corresponding PPAR is “Not applicable”; for baseline reference categories this is given as “Baseline”; and for behaviours where the odds ratio is <1 this is given as “Negative”.

Independent factor: levels	Odds ratio (95% CI)	PPAR (%)
Q5_2 (other adult left house): No	1	Baseline
Q5_2 (other adult left house): Yes	0.64 (0.51, 0.81)	Negative
Q5_4 (Other children have left house): No	1	Baseline
Q5_4 (Other children have left house): Yes	1.38 (1.04, 1.84)	6.3
Q5_5 (Nobody has left house): Yes	1	Baseline
Q5_5 (Nobody has left house): No	1.84 (1.32, 2.58)	39.5
Q6_3 (ate food outside workplace): No	1	Baseline
Q6_3 (ate food outside workplace): Yes	1.77 (1.26, 2.48)	7.3
Q7 (left house for any reason): No	1	Baseline
Q7 (left house for any reason): Up to weekly	0.70 (0.40, 1.23)	Negative
Q7 (left house for any reason): More than weekly	0.51 (0.30, 0.88)	Negative
Q9 (public transport usage): None	1	Baseline
Q9 (public transport usage): Up to weekly	1.08 (0.80, 1.46)	1.3
Q9 (public transport usage): More than weekly	1.87 (1.29, 2.69)	7.7
Q10_1 (eaten takeaways or street food): Less than weekly	1	Baseline
Q10_1 (eaten takeaways or street food): Weekly or more	1.51 (1.21, 1.88)	16.7
Q10_4 (eaten food from other house): Less than weekly	1	Baseline
Q10_4 (eaten food from other house): Weekly or more	1.55 (1.20, 2.00)	12.2
Q11_3 (always wash hands after toilet): Yes	1	Baseline
Q11_3 (always wash hands after toilet): No	1.45 (1.16, 1.83)	11.7
Wave: 1	1	Not applicable
Wave: 2	1.20 (0.83, 1.75)	Not applicable
Wave: 3	1.40 (0.97, 2.03)	Not applicable
Wave: 4	0.91 (0.62, 1.33)	Not applicable
Wave: 5	1.70 (1.19, 2.45)	Not applicable

4. Discussion

4.1. Discussion of results

In evaluating the association of various behaviours with contracting IID, we found a single adult behaviour which was associated with IID in all five waves. Namely, adults who ate food from takeaways or street food vendors (Q10_1) at least once per week were consistently at increased odds of disease, compared to those who ate such food less often or not at all (increased odds of 26-95% for the individual five waves and 59% for combined waves). Two other behaviours had an association with increased odds of IID in four out of the five waves: bought and eaten food outside the workplace or educational institute (Q6_3), 45-213% increase in odds

Table 7. Odds ratios (95% CI) for those seeking healthcare across all seven avenues for those with domestic IID (baseline definition) for adults across waves 1-5 for all demographic variables. The seven routes of seeking healthcare are: via the GP (in-person; online or over the telephone); consulting a pharmacist; visiting A&E at hospital; speaking to somebody after contacting 111; visiting NHS or other medical websites; or via another form of medical advice.

Independent factor: levels	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Age: 16-24	1	1	1	1	1
Age: 25-34	2.17 (1.00, 5.02)	1.04 (0.57, 1.92)	1.70 (0.78, 3.89)	0.55 (0.32, 0.96)	0.59 (0.32, 1.06)
Age: 35-44	1.90 (0.84, 4.56)	0.83 (0.43, 1.61)	1.30 (0.59, 2.97)	0.63 (0.35, 1.12)	0.51 (0.27, 0.97)
Age: 45-54	1.30 (0.53, 3.33)	0.34 (0.14, 0.74)	1.41 (0.62, 3.33)	0.34 (0.18, 0.65)	0.16 (0.06, 0.36)
Age: 55-64	1.12 (0.38, 3.18)	0.50 (0.24, 1.02)	0.58 (0.21, 1.56)	0.36 (0.18, 0.69)	0.18 (0.07, 0.42)
Age: 65+	1.15 (0.41, 3.21)	0.34 (0.13, 0.83)	1.09 (0.40, 2.94)	0.44 (0.22, 0.86)	0.31 (0.14, 0.66)
Gender: Female	1	1	1	1	1
Gender: Male	1.96 (1.23, 3.16)	2.24 (1.45, 3.51)	1.48 (0.94, 2.32)	1.43 (1.00, 2.05)	1.77 (1.19, 2.64)
Social Grade: AB	1	1	1	1	1
Social Grade: C1	0.90 (0.49, 1.63)	0.75 (0.43, 1.30)	0.68 (0.38, 1.22)	0.62 (0.38, 1.01)	0.49 (0.26, 0.92)
Social Grade: C2	1.06 (0.52, 2.12)	0.57 (0.31, 1.03)	0.79 (0.41, 1.50)	0.75 (0.45, 1.23)	1.11 (0.64, 1.92)
Social Grade: DE	1.03 (0.55, 1.93)	0.95 (0.52, 1.75)	0.61 (0.31, 1.16)	0.82 (0.50, 1.32)	0.71 (0.40, 1.26)
Ethnicity: White	1	1	1	1	1
Ethnicity: Other	2.04 (1.08, 3.81)	1.08 (0.56, 2.06)	2.54 (1.29, 5.00)	1.11 (0.65, 1.88)	1.20 (0.68, 2.11)
No. people in household: 1	1	1	1	1	1
No. people in household: 2+	0.80 (0.43, 1.51)	1.47 (0.74, 3.12)	1.45 (0.75, 3.00)	1.40 (0.85, 2.37)	1.52 (0.84, 2.85)
Country: England	1	1	1	1	1
Country: Scotland, Wales & NI	0.67 (0.33, 1.27)	1.07 (0.63, 1.80)	0.91 (0.49, 1.62)	0.73 (0.45, 1.17)	0.87 (0.51, 1.46)

against those not exhibiting this behaviour in four waves and 55% for combined data; and eaten at work or educational canteen (Q10_3) at least weekly (46-64% increase in odds against those exhibiting this behaviour less than weekly in four waves and 51% for combined data). Although only occurring in three waves' final models, those adults who ate from organised meal delivery services (Q10_5) at least weekly often exhibited the strongest association with IID: increased odds of IID by 88-134% for those three waves and from the combined waves data this resulted in a 66% increase in odds of disease.

Adults leaving the house for any reason (Q7) had an inconsistent relationship with IID, in that leaving the house, after adjusting for other factors in the models, is surprisingly protective against IID in three waves and strongly associated in another wave. In the combined waves analysis, this behaviour has a negligible unadjusted effect on domestic IID (1% and 9% reductions for up to weekly and more than weekly compared to never leaving the house, respectively, when adjusting for the four quota variables), but when further adjusting for public transport usage (Q9) this increases to 11% and 24% reductions and separately adjusting for ate ready to eat food outside the workplace or educational institute (Q6_3) increases to 5% and 21% reductions, whilst further adjusting for both factors increases the effects to 13% and 31% reductions, respectively for up to weekly and more than weekly frequency of leaving the house (close to the final effects of 16% and 37% reductions). As the negative effects of leaving the house on domestic IID have increased after adjusting for public transport usage (Q9) and eating food outside the workplace or educational institution (Q6_3), this indicates that these two behaviours negatively confound the effects of leaving the house for any reason on contracting domestic IID. This may be because there are some behaviours which are mutually exclusive, such that those not leaving the house (Q7) could not have used public transport (Q9), nor bought food outside of the workplace (Q6_3). The only other behaviour in the final model which is not possible if somebody does not leave the house is eating food at a work or educational canteen (Q10_3). The effect of leaving the house on IID after adjusting for this behaviour results in the next biggest reductions (6% and 9%, respectively) after Q9 and Q6_3.

We also saw a similar negative effect in the children's all-waves model relating to people leaving the house for any reason and domestic IID. Whereas the adjustment of the Q7 effect on domestic IID for adults was largely explained by Q9 (public transport use) and Q6_3 (eating food outside the workplace or educational institution), this is less so for the children's all-waves model. Here an additional confounder appears to be Q5_5 (nobody leaving the house for work or educational purposes).

In addition, from the children's best subset selection model for all five waves combined, there were some behaviours which exhibited a mild association with domestic IID. This included: Q6_3 (children eating ready to eat food from outside the school or childcare setting) having a 77% increase in the odds of domestic IID; Q10_4: children eating food from another home at least weekly had a 55% increase in the odds of domestic IID compared to those eating such food less frequently; Q10_1: eating food from a takeaway or street food vendor at least weekly increased the odds of domestic IID by 51% compared to those eating such food less frequently; Q11_3: not always washing hands after going to the toilet resulted in an increase in the odds of domestic IID by 45%; and Q9: more than weekly usage of public transport increased the odds of domestic IID by 87% compared to those children who had not used public transport in the previous four weeks. The results for these behaviours were largely consistent with the all-waves combined results for adults (55% outside the workplace, 32%, 59%, 21% and 36%, respectively).

Other results found in either the adults' or children's all-waves model, but not both, include for adults: Q10_5: 66% increase in the odds of domestic IID if adults have eaten from an organised meal delivery service at least weekly; Q10_3: 51% increase in the odds of domestic IID if adults have eaten food from a work or educational canteen at least weekly; and Q5_3: 36% increase in the odds of domestic IID if children leave the house). For children this includes: Q5_5: somebody left the house results in a 84% increase in the odds of domestic IID; Q5_4: other children leaving the house results in an increase in the odds of domestic IID by 38% (similar for Q5_3 in the adults model); and Q5_2: other adults (beyond the respondent on behalf of the child) leaving the house results in a reduction of domestic IID by 36%.

For the children analysis there were few behaviours selected in any of the individual wave Lasso models and those that were only featured in one of the five models.

No adult handwashing behaviour appeared in the best fit model for more than two of the five waves. There was consistency in effect sizes between waves for individual handwashing behaviours. However, there was inconsistency between such behaviours in terms of being associated with an increase or a decrease in IID. This inconsistent result might not be surprising given that such behaviours have the potential to suffer most from social desirability bias (when respondents give answers to questions that they believe will make them look good to others, concealing their true experiences) and so claimed behaviour may be less likely to reflect actual behaviours.

The sensitivity analyses between the three definitions of domestic IID largely produced consistent results from the best subset selection, in that the selected variables and their effect sizes were largely consistent. The selected variables in the Lasso models, though restricted in number, were also largely consistent with the best subset selection models and across waves. However, the odds ratios from the best subset selection were typically larger than those from the corresponding Lasso model.

For adults the entire removal of the Q10_1 behaviour indicates an expected reduction in IID cases by 9-24% (eating food prepared at takeaways or street food vendors in the previous four weeks), whereas for Q6_3 (have bought ready to eat food outside the workplace or school in the prior four weeks) this results in a 7-14% reduction in IID and for Q10_3 (eating prepared at work, school or university canteen in the prior four weeks) 6-8% reduction. Other research has estimated that the proportion of IID attributed to food is around 13% (Holland & Mahmoudzadeh, 2020) and 44-85% of the foodborne element is from eating food prepared outside the home (Redmond et al., 2018). It is therefore likely that IID associated with such behaviours is not all directly related to eating the food itself, or the previous estimates of the proportion of IID attributed to food are too low. Assuming the former, then it is likely that the associations may also be due to other factors associated with eating food produced at these places, such as an increased likelihood of mixing in confined spaces which might increase the likelihood of person-to-person transmission. Separating the effect of these two routes, as well as other possible factors, is not possible from this study.

For adults with IID further analysis was undertaken to see if there were any significant differences in the likelihood of seeking medical care depending on variations in demographics. This found that men were more likely to seek healthcare than women. This was true for four waves, the increased odds ranging from 43% to 124%. The other notable result was that other ethnicities had a more than doubled odds of seeking healthcare compared to white ethnicity in waves 1 and 3, but not in the other three waves.

The finding of men being more likely to seek medical care is in stark contrast to other studies where women were more likely to seek healthcare than men (Smith et al., 2006; Thompson et al., 2016). In addition the Food Standards Agency's Food and You 2 survey found that for those who had self-reported experiencing food poisoning there was no statistically significant difference between men and women in terms of whether they saw a doctor or went to hospital (Food Standards Agency, 2021).

Possible reasons for the result in this study being different to other studies include 1) something has changed or is otherwise different compared to the other research. For example, the context of IID during the pandemic

may have prompted different behaviours because how people accessed care was different, so it may have reduced barriers for men or people from other ethnicities, or conversely increased barriers for women or people of white ethnicity. 2) It could be linked to differences between those on the panel and the general population. In particular younger men are difficult to recruit to such panels. It may be that the types of men and/or people from other ethnicities who join panels may be more likely to seek medical care than those with the same demographics in the wider population. The focus of academic research in this area tends to be around things like political engagement and it has been found, for example, that people on panels tend to be slightly more opinionated than the population on average. The first of these possible reasons could be tested by a similar survey conducted in a post pandemic period to see if this is a sustained difference or occurred just during the pandemic.

4.2. Study limitations

All of these associations do not necessarily infer causation and as such may just be indicators/proxies of other causes such as likelihood and intensity of mixing with people outside the household. Indeed, causation as defined under the Bradford-Hill criteria requires several criteria, of which only one behaviour in our study (Q10_1) indicates a consistently positive relationship with IID. Other criteria include strength of association; specificity; temporality; biological gradient; biological plausibility; and coherence with other research (Hill, 1965).

There is some criticism that Lasso may not offer as much improvement over stepwise regression in variable selection. This was recently highlighted by the authors who originally developed the Lasso method (Hastie et al., 2020). They suggested that best subset selection (we have used AIC in our models; stepwise regression behaves in a similar manner) outperforms Lasso in high signal to noise scenarios, and that Lasso does better in low signal to noise scenarios, indicating that Lasso may be the preferable approach for our data. Possible alternative approaches to those used in this report include machine learning techniques, such as neural networks and random forest (Jomthanachai et al., 2022).

Finally, there were variations in lockdown conditions imposed on the UK during the different periods under examination, which may explain why we don't see strong consistent effects over time. Lockdown dates and how they tally with waves may be useful to explain this as may consideration of other factors such as sampling variability.

.....

Acknowledgements

Ipsos (formerly Ipsos MORI) designed the surveys and collected the data on which this report is based. Ipsos also wrote the main report covering the survey results. Sian Puttock (FSA) performed QA duties on this secondary analysis.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. Taylor & Francis.
- Bruzzi, P., Green, S. B., Byar, D. P., Brinton, L. A., & Schairer, C. (1985). Estimating the population attributable risk for multiple risk factors using case-control data. *Am. J. Epidemiol.*, 122, 904–914. <https://doi.org/10.1093/oxfordjournals.aje.a114174>
- Food Standards Agency. (2021). *Food and You 2: Wave 1 Data Tables – England, Wales and Northern Ireland Combined* (Food & You 2 Survey: Wave 1).
- Harrell, F. E. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer. <https://doi.org/10.1007/978-1-4757-3462-1>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Science & Business Media.
- Hastie, T., Tibshirani, R., & Tibshirani, R. (2020). Best subset, forward stepwise or lasso? Analysis and recommendations based on extensive comparisons. *Stat. Sci.*, 35, 579–592. <https://doi.org/10.1214/19-STS733>
- Hill, A. B. (1965). The Environment and Disease: Association or Causation? *Proc. R. Soc. Med.*, 58, 295–300. <https://doi.org/10.1177/003591576505800503>
- Holland, D., & Mahmoudzadeh, N. (2020). *Foodborne Disease Estimates for the United Kingdom in 2018*. Food Standards Agency. <https://doi.org/10.46756/sci.fsa.squ824>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R, Springer Texts in Statistics*. Springer US. <https://doi.org/10.1007/978-1-0716-1418-1>
- Jomthanachai, S., Wong, W. P., & Khaw, K. W. (2022). An application of machine learning regression to feature selection: a study of logistics performance and economic attribute. *Neural Comput. Appl.*, 34, 15781–15805. <https://doi.org/10.1007/s00521-022-07266-6>
- Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminformatics*, 6, 10. <https://doi.org/10.1186/1758-2946-6-10>
- Redmond, E., Curnin, A., Day, C., Morris, J., Eves, A., & Raats, M. (2018). *Systematic review of the relative proportion of foodborne disease associated with food preparation or handling practices in the home*. <https://www.food.gov.uk/sites/default/files/media/document/fs101098fbdinthehometechnicalrep.pdf>
- Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: PART II-binary and time-to-event outcomes. *Stat. Med.*, 38, 1276–1296. <https://doi.org/10.1002/sim.7992>
- Sjölander, A., & Vansteelandt, S. (2011). Doubly robust estimation of attributable fractions. *Biostatistics*, 12, 112–121. <https://doi.org/10.1093/biostatistics/kxq049>

- Smith, J. A., Braunack-Mayer, A., & Wittert, G. (2006). What do we know about men's help-seeking and health service use? *Med. J. Aust.*, 184, 81–83. <https://doi.org/10.5694/j.1326-5377.2006.tb00124.x>
- Thompson, A. E., Anisimowicz, Y., Miedema, B., Hogg, W., Wodchis, W. P., & Aubrey-Bassler, K. (2016). The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. *BMC Fam. Pract.*, 17, 38. <https://doi.org/10.1186/s12875-016-0440-0>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.*, 58, 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Viviani, L., van der Es, M., Irvine, L., Tam, C. C., Rodrigues, L. C., Jackson, K. A., O'Brien, S. J., Hunter, P. R., & Committee, I. S. E. (2016). Estimating the Incidence of Acute Infectious Intestinal Disease in the Community in the UK: A Retrospective Telephone Survey. *PLOS ONE*, 11, e0146171. <https://doi.org/10.1371/journal.pone.0146171>
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.*, 1, 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>

Supplementary Materials

Appendix A-B

Download: <https://science.food.gov.uk/article/122449-infectious-intestinal-disease-survey-during-covid-19-additional-analysis-on-association-of-behaviours-with-illness/attachment/241247.docx>
