The Adults' People and Nature Survey: Small Area Estimation

June 2024

Natural England Research Report NECR551

www.gov.uk/natural-england

About Natural England

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

Further Information

This report can be downloaded from the Natural England Access to Evidence Catalogue. For information on Natural England publications or if you require an alternative format, please contact the Natural England Enquiry Service on 0300 060 3900 or email enquiries@naturalengland.org.uk.

Copyright

Report details

Author(s)

Dr. Alex Gibson, RAE Consulting

Natural England Project Manager

Tom Marshall (Principal Adviser & SRO, PaNS) & Caitlin Clark (Principal Adviser, Statistics).

Email: People_and_Nature@naturalengland.org.uk

All enquiries about the research should, in the first instance, be directed to the Natural England Project Managers.

Keywords

Engagement, green and natural spaces, people and nature.

Citation

Gibson, A. (2024). The Adults' People and Nature Survey: Small Area Estimation. *Natural England Commissioned Report* NECR523. Natural England.

Foreword

Natural England commissioned this work to assess the suitability of Small Area Estimation techniques to develop a model using The Adults' People and Nature Survey (A-PaNS) data. The contractors that deliver the People and Nature Surveys for England (PaNS), Verian, commissioned RAE Consulting to deliver this assessment.

The overall purpose of this exercise was to understand whether Small Area Estimation techniques will be useful in providing more accurate estimates of attitudes and behaviours in small geographical areas such as Local Authorities and Middle Layer Super Output Areas (MSOA) using data collected from A-PaNS.

The findings of this report will inform the future development of the People and Nature Surveys and improve the ways that Natural England monitors and researches how people engage with the natural environment, with particular regard to how it measures engagement with and public perceptions of nature at a Local Authority level.

Disclaimer: Natural England commissions a range of reports from external contractors to provide evidence and advice to assist us in delivering our duties. The views in this report are those of the authors and do not necessarily represent those of Natural England.

Executive summary

This Small Area Estimation (SAE) project explores whether a model-based approach can be used to produce accurate, precise and useful estimates of how people in different Local Authorities (LAs) engage with nature. Using data from *The Adults' People & Nature Survey (A-PaNS)* dataset, it focuses on estimating the proportion of adults who:

(a) have visited 'green & natural spaces' over the previous 14 days (*AnyVisit14*), and
(b) consider that green and natural spaces 'close to where they live' have improved over the previous five years (*ImprovedSpaces*)

This report outlines the principles and implementation of SAE, details the *AnyVisits14* and *ImprovedSpaces* models, and reports on the final set of LA-level estimates, but the key observations for Natural England and LA stakeholders are as follows:

- SAE produces significantly more precise LA-level estimates (i.e. estimates have narrower 95% Confidence Intervals) than can be achieved using traditional survey-based estimation. The estimates are directly comparable across all LAs even though the impact of the COVID-19 pandemic varied between LAs.

- Responses to the *AnyVisit14* and *ImprovedSpaces* questions were strongly influenced by individual-level factors such as age, general health, and educational status.

- Improved estimate precision was achieved largely because SAE is particularly effective at combining census-based evidence on the composition of LA populations with model-based evidence (drawn from the whole A-PaNS dataset) on the role that individual-level factors play in determining how people use and view green and natural spaces.

- The SAE models developed in this study improved LA-level estimate precision, but significant uncertainty remains. Additional data from future A-PaNS and/or the use of better locality data will undoubtedly allow for further improvement.

- It is questionable, however, whether this approach will ever be able to produce estimates that are precise enough to monitor how use of, and attitudes to, nature in specific LAs change over time. This means that the method is unlikely to be useful for measuring changes in response to policy initiatives and interventions, or drawing robust conclusions about differences between LAs, for example to highlight positive or negative impacts of divergent policy environments.

- If the goal is to inform the development of policies that encourage the use of green and natural spaces, striving to improve estimate precision may not represent the best use of A-PaNS. Effort would be better focused on using SAE as a framework for developing a better understanding of how policy-modifiable factors influence how different population cohorts in different localities make use of green and natural spaces.

- A set of five ‘Further Work’ recommendations are made on how this might be achieved.

Contents

List of Tables

List of Figures

Background and Method

Since April 2020, Natural England's People and Nature Surveys for Englandhave gathered evidence about how people in England engage with nature. It is formed of two surveys, the Adults' PaNS (A-PaNS) and the Children's PaNS (C-PaNS).

The Adults' People and Nature (A-PaNS) Survey for England is one of the main sources of data and statistics on how people experience and think about the environment in England. It began collecting data in April 2020 and has been collecting data since. This report includes data that were collected between April 2020 and March 2023. Data are collected via an online panel of adults aged 16 years and older, in line with Government Statistical Service guidance on data collection during the COVID-19 pandemic.

The 'Children's People and Nature Survey' (C-PANS) provides information on how children and young people in England aged between 8 and 15 years old experience and think about the natural environment. It is run twice each year, once in term time and once in holiday time, among c. 4,000 children each year. Further detail about the People and Nature Surveys is included within the technical report[1].

Data from the People and Nature Surveys for England enable users to:

- Understand how people use, enjoy and are motivated to protect the natural environment.
- Monitor changes in use of the natural environment over time, at a range of different spatial scales and for key groups within the population.
- Understand how being in the natural environment can have an effect on wellbeing.
- Understand environmental attitudes and the actions people take at home, in the garden and in the wider community to protect the environment.

These data contribute to Natural England's delivery of statutory duties, inform Defra policy and natural capital accounting, and contribute to the outcome indicator framework for the 25 Year Environment Plan, now the Environmental Improvement Plan (specifically the G indicators). The People and Nature Surveys were awarded Accredited Official Statistics status in November 2023.[2]

Designed to provide insights at a national level, by the end of March 2023 the A-PaNS dataset included responses from a representative sample of 74,968 adults.

This is sufficient for producing national estimates of how people access and interact with natural spaces, but many stakeholders have requested estimates that are specific to

[1] https://assets.publishing.service.gov.uk/media/6570814d746930000d488913/PaNS_Technical_Report_2023_update_published_Dec23.pdf

[2] https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-ian-lonsdale-assessment-of-statistics-from-the-people-and-nature-survey-england/

smaller geographies, particularly at Local Authority (LA) level. This poses a challenge to traditional survey-based estimation because of the small number of responses available for individual LAs.

This Small Area Estimation (SAE) Project was set up to explore an alternative, model-based, approach to analysing the A-PaNS dataset. It focussed on estimating the proportion of adults in each LA who:

(a) have visited 'green and natural spaces' over the previous 14 days (henceforth *AnyVisit14*), and
(b) consider that green and natural spaces 'close to where they live' have improved over the previous five years (*ImprovedSpaces*).

These variables were chosen for this initial exercise because visits to nature and perceptions of quality of green space are important across a range of key policy areas, including Defra's Environmental Improvement Plan and the Department for Levelling Up, Housing and Communities' 'Levelling Up' mission.

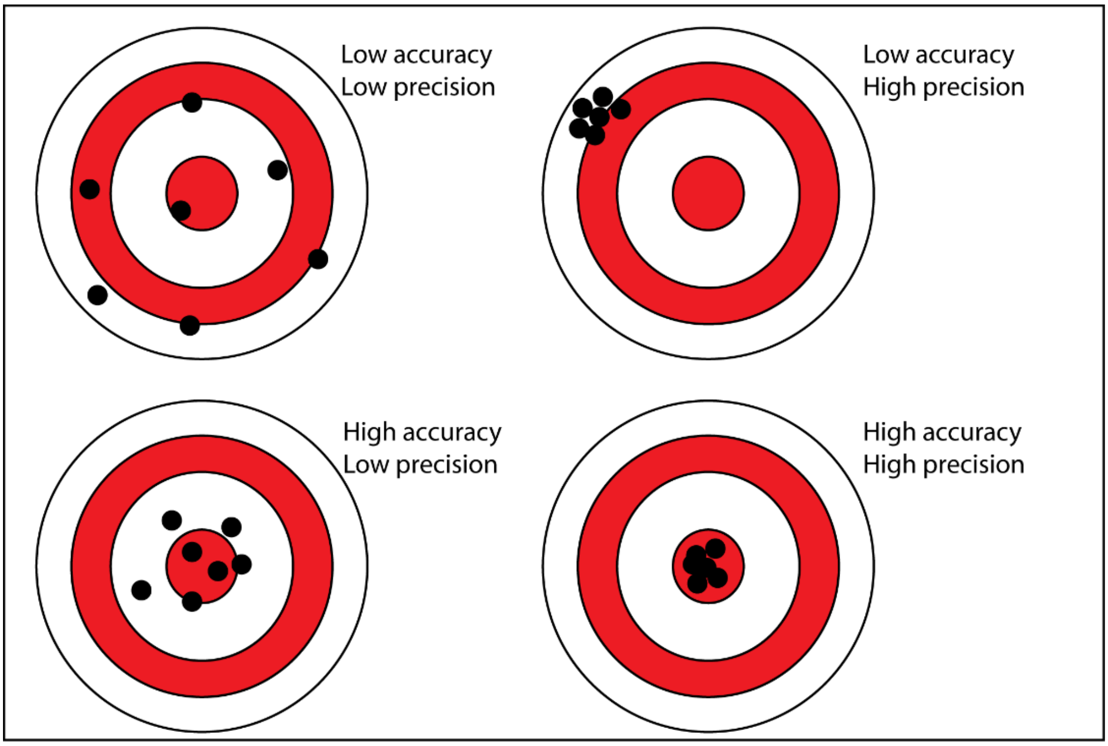
Estimate Accuracy, Precision and Usefulness

It is impossible to ask everybody in England about how they interact with nature, so surveys such as A-PaNS ask questions of a fraction of the whole population and use this sample to estimate likely population or cohort responses. The goal is to provide estimates that are accurate, precise and useful.

Accuracy: This refers to how close a survey-based estimate of, for instance, *AnyVisit14*, is to the 'true' population-level value. This largely depends on a well-constructed survey of carefully worded questions being asked of a genuinely representative sample. If this is repeated many times the survey-based estimates will tend to cluster around the 'true' population or cohort value, as illustrated by the bottom two 'targets' on Figure 1 below.

Precision: These repeated estimates, even if accurate, may not be closely clustered. This, as illustrated by the bottom-left-hand target, undermines the level of confidence in precisely where the true (but unknown) value lies. In practice, estimate precision is usually expressed as the range over which one is 95% sure that the true value lies and, in traditional survey-based estimation, this is almost entirely a function of sample size: the larger the sample the more precise the estimate.

Figure 1: Estimate precision and accuracy



Usefulness: The goal is to achieve high accuracy and high precision, but it is not always obvious when the estimates are sufficiently accurate and precise. This depends on their usefulness to potential users.

For LAs with a relatively small number of respondents in A-PaNS, such as the 100 valid[3] responses available for Uttlesford District Council, the 95% CIs are wide and the estimates have reduced accuracy, precision and therefore usefulness. The *AnyVisit14* estimate for Uttlesford is 52%, but the 95% CI is 40.2% – 65.2%.

There are 60 LAs with 100 or fewer valid responses to the *AnyVisit14* question and in these cases the estimates are not robust enough to draw confident conclusions about local behaviour or perceptions. Even for LAs with the largest sample sizes (Table 1), the 95% CIs are relatively wide.

Table 1: *AnyVisit14* Rates & 95% CIs for LAs with the largest A-PaNS samples

Local Authority	Valid Responses	14-day visit rate	95% Cis
Sheffield	631	75.5%	[68.2%–76.6%]
Cornwall	663	71.3%	[66.8%–75.7%]

[3] Estimates are based on respondents who provided a postcode and who responded Yes or No to the survey question. It excludes those who responded ‘Don’t know’ or ‘Prefer not to say’.

Leeds	923	73.6%	[70.1%–77.1%]
Birmingham	1357	68.8%	[65.8%–71.7%]

An alternative approach: Small Area Estimation (SAE)

In contrast to traditional direct survey-based estimation – which only uses data relating to the specific LA for which an estimate is being produced – Small Area Estimation (SAE) is a long-established technique[4] which pools evidence from across a survey as a whole to generate estimates about smaller geographies. Applied to A-PaNS, this has the potential to significantly improve LA estimate precision.[5]

Individual-level predictive models are constructed using data on the personal characteristics of all individuals in the survey and, through postcode linkage, the nature of the localities and LAs in which they live.

These models are then used to predict the likelihood that each adult in each LA will, in this study, (a) have visited green and natural spaces over the previous 14 days (*AnyVisit14*), and (b) agree that local green and natural spaces have improved over the last five years (*ImprovedSpaces*). LA-level rates are calculated by aggregating these individual-level likelihoods to LA-level and then dividing by the total adult population.

If a reasonably strong relationship exists between available individual-, locality- and LA-level predictor variables and the *AnyVisit14* and *ImprovedSpaces* response variables then the LA-level estimates should be far more precise than can be achieved using direct survey-based estimation. Importantly, the precision of these modelled estimates can be calculated and compared with the direct survey-based estimates.

Small Area Estimation: Evidential and Analytical Challenges

As the individual-level model which underpins SAE is constructed using survey data, and is then applied to individuals in local areas, it is necessary to describe individuals in the survey and the wider population in precisely the same terms. It is also important to capture as many relevant individual- and place-based characteristics as possible.

Data on the detailed composition of local populations

This analysis uses the 2021 Census to obtain information on local populations. The Census publishes a series of cross-tabulations of the number of people in broadly defined categories. The census does not, of course, contain data on all potentially relevant factors.

4 Rao, J.N.K. (2003) *Small Area Estimation* New York, Wiley

5 CITY SCIENCE. 2019. *Small Area Estimation feasibility: MENE survey*. Natural England Commissioned Reports, Number 268. (https://publications.naturalengland.org.uk/publication/5051507248726016)

Analysis of A-PaNS shows that having a dog in a household is a very strong predictor of *AnyVisit14*, but no British census has ever asked about dog ownership.

Information about the composition of local areas is also limited as the UK Census does not publish data that could be used to identify individuals. Instead, it publishes a series of cross-tabulations of the number of people in broadly defined categories.

It is thus necessary to use a technique called microsimulation to estimate the number of people in each local area Middle Super Output Area (MSOA) with each unique combination of characteristics. A simple non-iterative approach was adopted using the categories listed in Table 2 below. The resulting 'full joint distribution' approximates the detailed socio-economic and demographic composition of each MSOA population.

Table 2: 2021 Census 4-way cross-tabulations used in microsimulation

Content
Ageband / Sex / Access to Car / Whether Children in HH[6]
Ageband / Sex / Access to Car / Economic Activity Status[7]
Ageband / Sex / Access to a Car / General Health Status[8]
Ageband / Sex / Access to a Car / Highest Ed Qual[9]
Ageband / Sex / Access to a Car / Marital Status[10]
Ageband / Sex / General Health Status / Marital Status[11]
Ageband / Sex / General Health Status / Highest Ed Qual[12]
Ethnic Group (20)[13]

[6] https://www.ons.gov.uk/datasets/create/filter-outputs/862f3998-7fac-445e-9a09-30cfb17a0abf#get-data

[7] https://www.ons.gov.uk/datasets/create/filter-outputs/2de95132-baf7-4b4f-a1c8-7bf24e0b6d2d#get-data

[8] https://www.ons.gov.uk/datasets/create/filter-outputs/0e2d8e3e-76ea-49c7-b714-c340c1fe8912#get-data

[9] https://www.ons.gov.uk/datasets/create/filter-outputs/a325bf2c-66af-4863-8770-f456d6089b58#get-data

[10] https://www.ons.gov.uk/datasets/create/filter-outputs/2de95132-baf7-4b4f-a1c8-7bf24e0b6d2d#get-data

[11] https://www.ons.gov.uk/datasets/create/filter-outputs/711631c6-4404-4904-9c94-e5377cbe2e7f#get-data

[12] https://www.ons.gov.uk/datasets/create/filter-outputs/ad72b22f-8edf-4abb-8ed7-71a75330c973#get-data

[13] https://www.ons.gov.uk/datasets/create/filter-outputs/e1ee359b-c447-4c5f-83e6-bca8735c8c78#get-data

Data on the characteristics of where people live

MSOA-level data on the characteristics of the localities in which people live were included in the SAE models. It seemed intuitively likely that use of, and attitudes to, green and natural spaces would be influenced by the local accessibility of such spaces, as well as by the overall socio-economic and demographic characteristics of local populations.

It was not possible to develop bespoke MSOA-level measures. Instead, the analysis was restricted to testing the potential contribution of various existing measures that could be attributed to 2021 MSOAs. These are listed in Table 3 below, along with the geographical units at which the original data were available and whether the attribution to 2021 MSOAs (the geographic units used in this analysis) was based on a population- or area-based weighting.

Table 3: MSOA-level locality data

Dataset	Geography	Attribution basis
Index of Multiple Deprivation 2019	2011 LSOA	Population
ONS 2011 rural-urban classification	2011 MSOA	Population
ONS 2011 residential-based area classifications	2011 LSOA	Population
Access to Health Assets and Hazards (AHAH)	2011 LSOA	Population
Spatial Signatures of Great Britain	2011 LSOA	Area (Hectares)

The inclusion of MSOA-level data meant that only A-PaNS respondents who provided a postcode could be included in the predictive *AnyVisit14* and *ImprovedSpaces* models. While this reduced sample size, it was necessary to ensure that the models and estimates were sensitive to context as well as the local composition of populations.

LA as a ‘dummy’ variable in the SAE models

Pre-April 2023 lower-tier local authorities (n=308) were included as a potential fixed effect in both the *AnyVisit14* and *ImprovedSpaces* models.

It was deemed unlikely that responses to the *AnyVisit14* and *ImprovedSpaces* questions will have been strongly influenced by LA-specific policies or other factors encouraging or facilitating the use of green and natural spaces. It is far more likely that any LA-level effect in the *AnyVisit14* and *ImprovedSpaces* models captures residual geographic influences not accounted for by the available MSOA-level variables. Thus, given that multilevel models are far more time-consuming to fit, a single-level generalised linear model (GLM) approach has been adopted.

Month and Covid-19 restrictions as additional variables

Finally, any understanding of how people engage with nature during the survey period has to recognise the impact of seasonal effects and the evolving set of restrictions put in place to supress the spread of COVID-19. The former was accomplished using a simple ‘*month of interview’* variable, but the latter required a day-by-day analysis and categorisation of the restrictions in place in each LA (Table 4)[14].

Table 4: Categorisation of Local COVID-19 Restrictions

Category	Description
0	No restrictions or government advice concerning precautions individuals should take regarding the Covid Pandemic
1	Government advice against travel and/or social mixing and/or minor regulations regarding personal freedoms (broadly equivalent to Covid Phase 2 Tier 1 restrictions)
2	Significant restrictions on personal freedoms (broadly equivalent to Covid Phase 2 Tier 2 & 3 restrictions)
3	Late-Covid ‘Say-at-home’ order, which allowed outdoor exercise
4	Early-Covid ‘Stay-at-Home’ order, which included a prohibition on exercise.

A COVID-19 ‘Category’ was attached to each A-PaNS respondent based on interview date and the LA in which they lived. Following a review of the data, this variable was not included in the final *ImprovedSpaces* model but was included in the *AnyVisit14* model.

[14] Institute for Government, *Timeline of UK government coronavirus lockdowns and restrictions*, December 2022. [Available at https://tinyurl.com/ye4jasmd]; House of Commons Library, *Coronavirus: the lockdown laws*, July 2022. [Available at https://commonslibrary.parliament.uk/research-briefings/cbp-8875/]

LA-Level SAE Estimates: *AnyVisit14* & *ImprovedSpaces*

The SAE models are used to predict individual-level likelihoods. LA-level *AnyVisit14* and *ImprovedSpaces* proportions are obtained by summing these likelihood estimates and dividing by the total LA adult population. However, it is also necessary to establish the precision (95% CIs) of the SAE estimates so they can be compared with the original direct, survey-based estimates. This approach used bootstrapping to estimate precision.

Estimating Estimate Precision: Bootstrapping

Bootstrapping is a long-established method for assessing the uncertainty of sample estimates. The classic illustration of this process concerns estimating the mean height of people worldwide. A tiny but representative sample can be taken, but how reliable (precise) is the sample estimate as a guide to the actual (unknown) mean global height?

The bootstrap approach involves re-sampling (with replacement) the original sample to obtain a large number of 'bootstrap samples', for each of which a mean height is calculated. If there is little variability in the underlying data the bootstrap samples will all return similar mean heights. However, if there is lots of variability, the set of bootstrap sample means will also be more varied. The distribution of bootstrap means captures the precision of the estimate, with the range within which 95% of the bootstrapped means lie describing the range within which we can be 95% confident that the (unknown) actual global mean lies.

In the present analysis the bootstrap samples are taken from the original A-PaNS dataset.[15] The model is then re-fitted using each new bootstrap sample and a new set of likelihood estimates and, ultimately, LA-level estimates are obtained. The range within which 95% of a LA's bootstrap estimates lie represents the range within which we are 95% confident that the true (but unknown) *AnyVisit14* or *ImprovedSpaces* value for that LA truly lies.

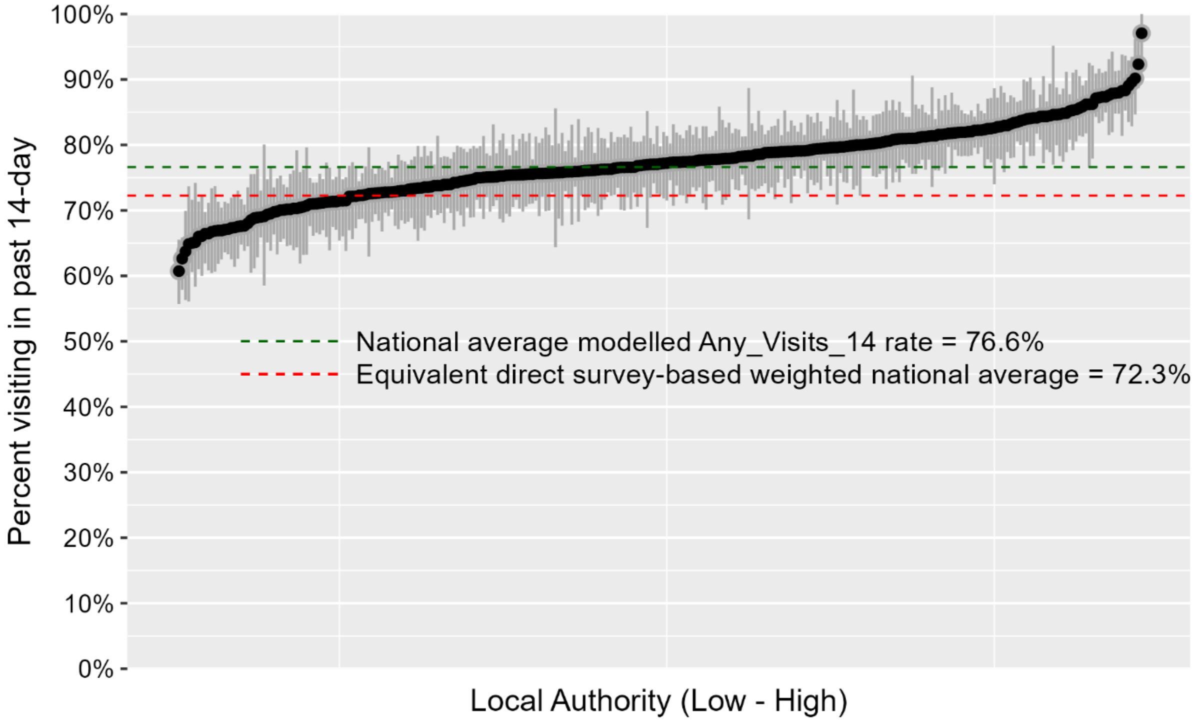
Comparing SAE and Direct LA-level Estimates

SAE LA-level estimates and 95% CIs for *AnyVisit14* or *ImprovedSpaces* are given in Appendices 5 and 6 respectively. These are accompanied by the original direct survey-based estimates and 95% CIs. Three observations need to be made.

What is Accuracy? The implications of modelling COVID-19 Restrictions

It is immediately obvious in Appendices 5 and 6 that the SAE estimates tend to be slightly higher than those obtained through direct, survey-based estimation. Indeed, as illustrated in Figure 2, whereas the overall national *AnyVisit14* weighted proportion using direct estimation is 72.3%, the SAE estimate is 76.6%. The two estimates are comparable in that both are based on the same subset of the A-PaNS dataset which excludes ‘Prefer not to say’ and ‘Don’t know’ responses as well as those for individuals without postcodes. However, they differ because the modelled estimate assumes that individuals respond in June (when responses rates tend to be a bit higher than average) and when there are no COVID-19 restrictions or recommendations against travel and social mixing (when rates are very much higher than when advice/restrictions are in place).

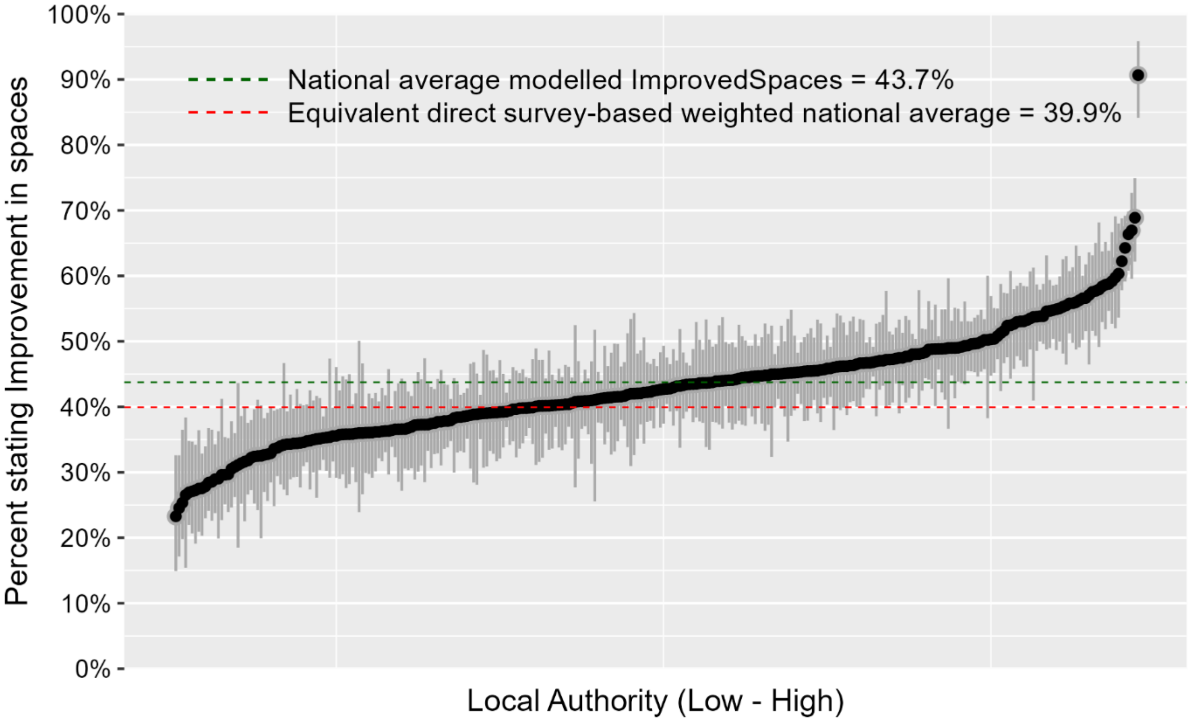
Figure 2: LA-level modelled estimates of *AnyVisit14*



The equivalent comparison for the overall national *ImprovedSpaces* proportion is 39.9% when using direct survey-based weighted estimation and 43.7% when based on SAE (Figure 3 below). This is larger than one might expect as COVID-19 restrictions do not materially affect responses to the *ImprovedSpaces* questions, though the direct estimates will still be supressed relative to SAE estimates because interviews take place across the year rather than just in June, the nominal month to which SAE estimates refer.

These systematic differences, particularly with respect to *AnyVisit14*, raise the question: what is meant by accuracy? Direct estimates must reflect the answers supplied notwithstanding any seasonal or COVID-19 influences on behaviour or attitudes, whereas SAE must control for any significant seasonal and COVID-19 effects. SAE estimates are therefore inaccurate in one sense, but they are comparable between LAs and, should the exercise be repeated, would be comparable with future estimates.

Figure 3: LA-level modelled estimates of *ImprovedSpaces*



Improved Estimate Precision using SAE

Appendices 5 and 6 also show the extent to which SAE improves estimate precision for individual LAs. The size of the effect for different LAs does vary, largely depending on the size of the LA-specific sample (larger sample sizes improve direct estimates and thus tend to reduce the 'space' for SAE improvement) and the detailed composition of local populations.

In general terms, however, the relative size of the 95% CIs around LA-level SAE estimates is substantially smaller than around LA-level direct estimates. The Coefficient of Variation (CV) is a widely used relative measure of dispersion based on the ratio of the standard deviation to the mean.

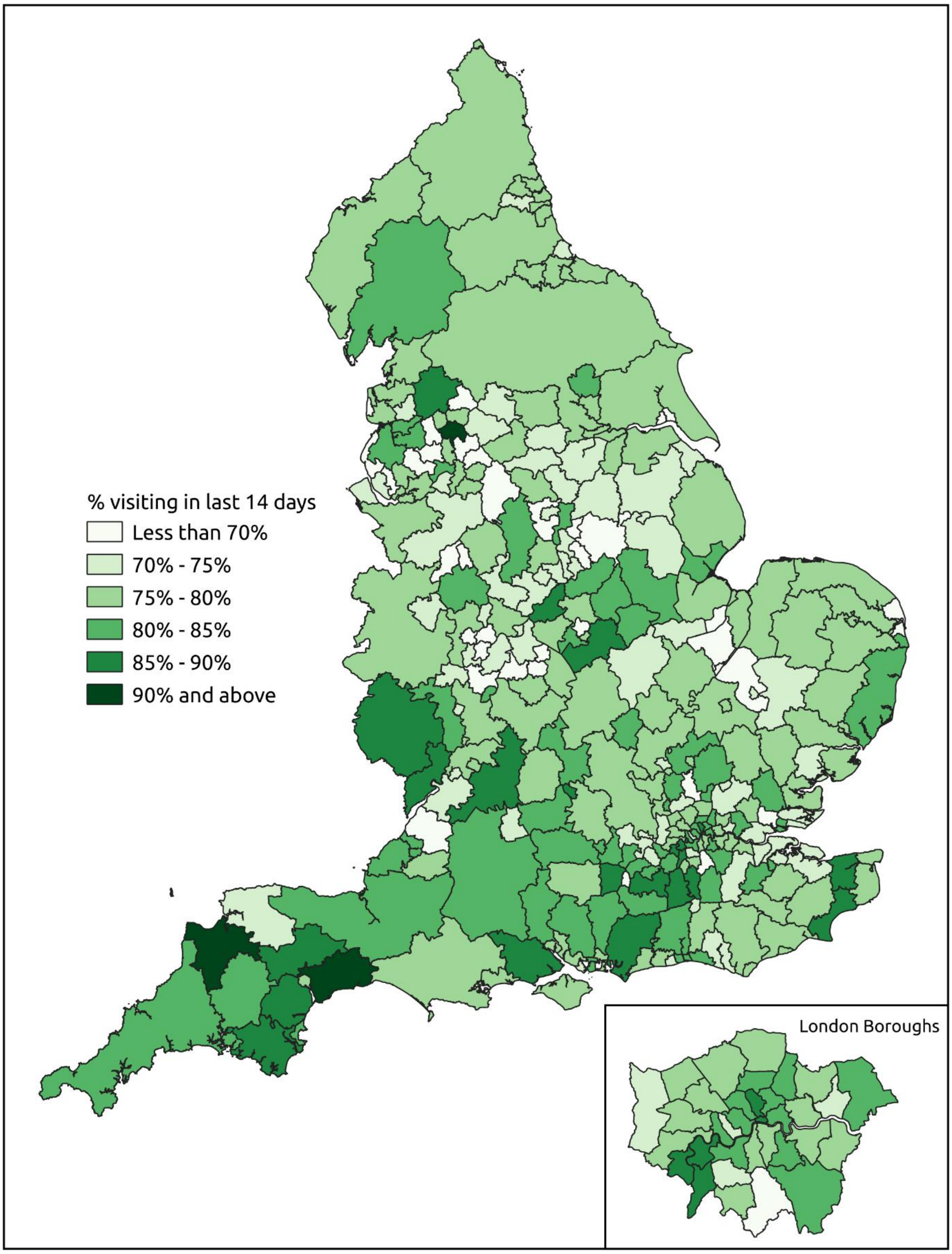
- For *AnyVisit14* LA-level direct estimates the average CV across all LAs is 6.65%, whereas for SAE it is smaller at 4.6%.
- For *ImprovedSpaces* estimates the average for direct estimates is 12.5%, whereas for SAE it is smaller at 9.8%.

These represent substantial improvements in precision, but does it now mean the estimates are useful for local authorities?

The usefulness of SAE Estimates

This, of course, depends on purpose. There is no doubt that improved precision around genuinely comparable LA estimates improves confidence in our understanding of how the use of green and natural spaces varies across the country, for instance. The pattern revealed by Figure 4 is likely to be as genuine as it is unsurprising: namely that, other than with respect to parts of London, rates tend to be higher in rural and coastal districts, and lower in urban and metropolitan areas.

Figure 4: Modelled SAE Estimates of *AnyVisit14*



At the level of individual LAs the issue is less clear cut. The SAE models developed in this analysis improve estimate precision relative to direct survey-based estimates, but significant uncertainty remains. Moreover, even though more data from future A-PaNS and/or the use of more relevant data about places will undoubtedly allow for increasingly precise estimates, it is questionable whether SAE estimates will, in themselves, be able to provide the information LA stakeholders really need.

This presumably focuses on two key questions:

1) Is it possible to monitor how use of, and attitudes to, nature change over time in a specific LA, particularly in response to policy initiatives and interventions, and;
2) Is it possible to draw robust conclusions about differences between LAs, again with a view to highlighting the positive or negative impacts of divergent policy environments?

The difficulty lies with two issues clearly illustrated in this study.

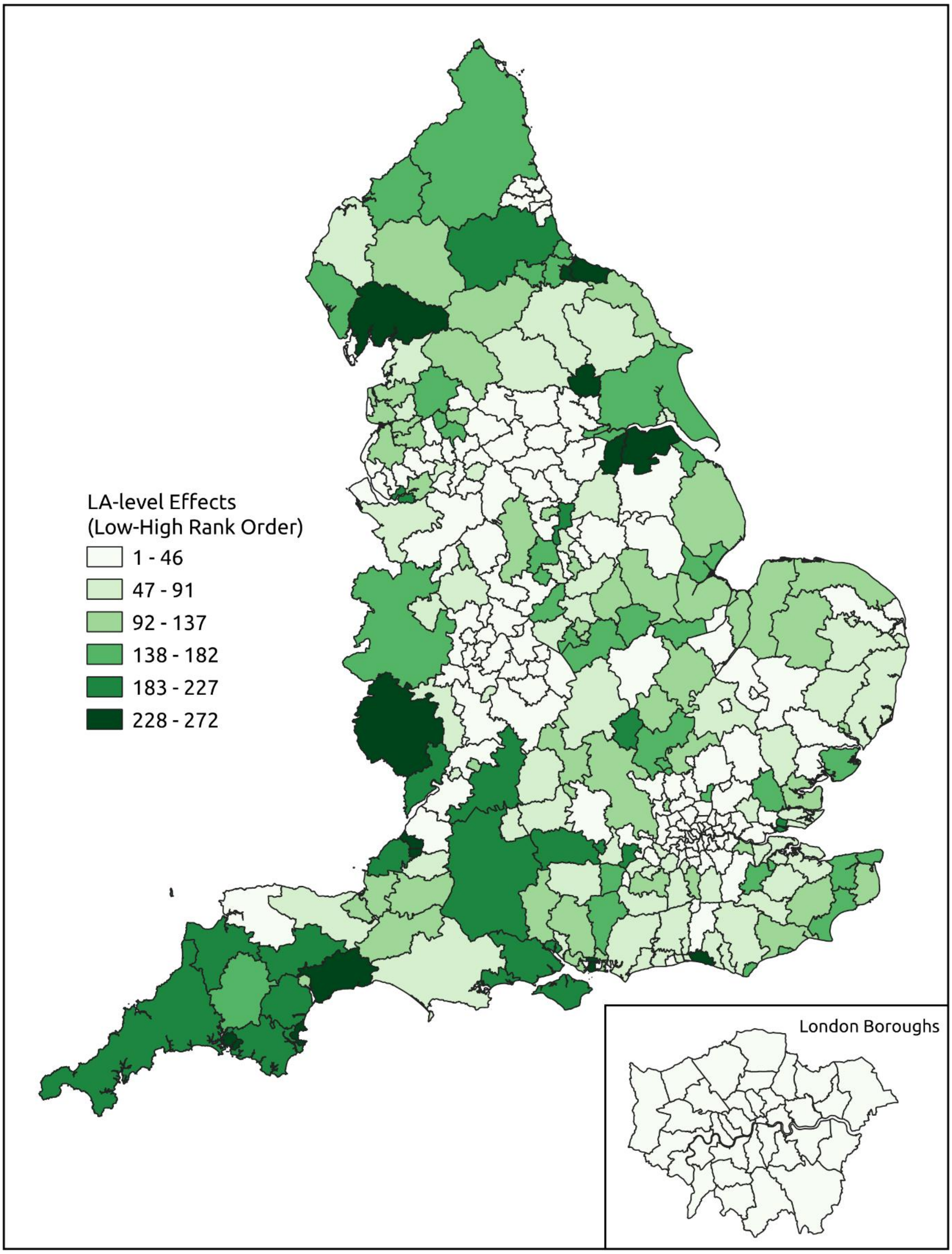
First, that the use of green and natural spaces, and even beliefs as to how those places have changed over time, strongly reflect the personal characteristics of individuals and the places in which they live. As a result, population-level *AnyVisit14* and *ImprovedSpaces* rates primarily respond to the composition of local populations and the nature of the places in which they live.

Second, that the accuracy and precision of SAE-based estimates is largely dependent on the explanatory power of models that use data for the country as a whole, and thus improved estimate accuracy and precision will reflect a better understanding of the nature of the overall way in which individual and locality factors influence the response variables (e.g. *AnyVisit14* and *ImprovedSpaces*) in which we are interested.

The inclusion of an explicit LA dummy variable to capture how responses in each LA are higher or lower than might be expected may provide insights into policy effects (Figure 5) but, unless it is possible to include very informative MSOA- or, preferably, LSOA-level locality data, these LA-level effects are likely to remain both imprecise and confounded by a range of unidentified geographic factors.

This is not to argue that SAE should not be pursued, but rather that the emphasis would be better placed on embedding SAE within a broader model-based analysis of A-PaNS data. Possible options are considered in the next section.

Figure 5: LA-level Effects in the *AnyVisits14* model



Conclusions and Recommendations for Further Work

Small Area Estimation (SAE) is a long-established technique which pools evidence from across a survey as a whole to generate estimates about smaller geographies. This report explored whether a model-based approach can be used to produce accurate, precise and useful estimates of how people in different Local Authorities (LAs) engage with nature.

Using data from A-PaNS dataset, it has focused on estimating the proportion of adults who:

(a) have visited 'green and natural spaces' over the previous 14 days (henceforth *AnyVisit14*), and
(b) consider that green and natural spaces 'close to where they live' have improved over the previous five years (*ImprovedSpaces*).

The key conclusions from this analysis are:

- SAE produces significantly more precise LA-level estimates (i.e. estimates have narrower 95% Confidence Intervals) than can be achieved using traditional survey-based estimation. The estimates are directly comparable across all LAs even though the impact of the COVID-19 pandemic varied between LAs.
- Responses to the *AnyVisit14* and *ImprovedSpaces* questions were strongly influenced by individual-level factors such as age, general health, and educational status.
- Improved estimate precision was achieved largely because SAE is particularly effective at combining census-based evidence on the composition of LA populations with model-based evidence (drawn from the <u>whole</u> A-PaNS dataset) on the role that individual-level factors play in determining how people use and view green and natural spaces.
- The SAE models developed in this study improved LA-level estimate precision, but significant uncertainty remains. Additional data from future A-PaNS and/or the use of better locality data will undoubtedly allow for further improvement.
- It is questionable, however, whether this approach will ever be possible to produce estimates that are precise enough to monitor how use of, and attitudes to, nature in specific LAs change over time. This means that the method is unlikely to be useful for measuring changes in response to policy initiatives and interventions, or drawing robust conclusions about differences between LAs, for example to highlight positive or negative impacts of divergent policy environments.

Although SAE approaches do provide more accurate small area estimates, it is arguable that pursuing ever more accurate, precise, and directly comparable LA-level SAE estimates is of limited value. Instead, effort would be better focused on developing a better understanding on how policy-modifiable factors influence how different population cohorts in different localities make use of such spaces.

SAE, as a model-based approach to identifying the impact of individual-, locality- and LA-level factors on individuals' use of, and attitudes towards, nature may be the most appropriate mechanism for identifying the potential role of such policy-modifiable factors. We make five recommendations concerning possible further work:

Stakeholder engagement: if the goal is, for health, social cohesion or any other reason, to develop policies to encourage use of green and natural spaces, then it would be invaluable to consult local practitioners as to the range of realistic policy options. This can be used to inform, as detailed below, the future development of A-PaNS and its analysis.

Improved locality data: While we arguably have good, census-based insights into the composition of local populations, evidence on the characteristics of localities that might affect use of green and natural spaces is extremely limited. We would recommend that such data is collected at as granular a level as possible, preferably LSOA level. This would allow for the development of improved SAE models, but many locality characteristics, including factors such as the availability, accessibility, and attractiveness of green and natural spaces, will also be amenable to policy initiatives.

Focus on obtaining postcode data from A-PaNS respondents: The importance of the context on individuals' use of green and natural spaces means identifying precisely where A-PaNS respondents live should be of high priority. Sample size in the current dataset was significantly degraded because of missing postcodes in the dataset.[16]

Incorporate questions in A-PaNS to directly address policy-modifiable factors: Input from stakeholders should identify a range of factors which they believe may impact how people interact with nature and which are amenable to policy. The A-PaNS questionnaire should address these issues directly so their actual importance can be assessed. This could include questions about the availability of transport options, the local provision of formal or informal outdoor activities and/or the availability of suitable maintained spaces.

Focus on understanding the impact of modifiable factors in the SAE model, not on producing LA-level estimates: This represents an important shift in emphasis. The goal is to develop a better understanding of what matters, to whom, and where, rather than the far more limited one of simply improving LA-level estimates of individuals' use of, and attitudes towards, nature.

[16] A-PaNS respondents have the option to opt out of providing a postcode

Appendix 1: Interpretative Issues

SAE is able to provide more precise LA-level estimates than direct survey-based estimates. This is because it pools evidence from <u>across the entire A-PaNS dataset</u> on how a variety of individual- and locality-factors influence how people use and think about green and natural spaces. But this methodological strength brings with it limitations.

In the first place, an implicit assumption behind SAE is that the national model is 'spatially invariant'; in other words, that individual- and MSOA-level factors can be applied to individuals living anywhere in the country. Yet some areas may be so unusual that estimates derived from an analysis of the national dataset may be compromised. The City of London, with a small resident population living in the financial and business heart of a truly global city, is an obvious example.

The inclusion of a LA dummy variable mitigates against the generalising tendency of using nationally based models. This ensures that individual likelihood estimates, and thus overall rates, are adjusted up or down according to how individuals in each LA respond. But, if the LA sample is small, the LA-specific parameter value may be relatively uninformative.

As a socio-economic and geographic outlier with relatively few A-PaNS respondents (95 and 118 for *AnyVist14* and *ImprovedSpaces* respectively), SAE estimates for the City of London should obviously be treated with caution, but the more general observation is that any LA-level estimate represents *"**what would be expected if modelled national relationships between predictor variables and the response variable hold true in that particular LA**"*.

The second, closely related issue, is that estimate *accuracy* and *usefulness* is closely bound up with the overall explanatory power of the predictive model <u>and</u> the relative contribution of its three distinct components: namely individual-level characteristics, place-based factors and the LA-level dummy variable.

As detailed in the appendix 3 and 4, individuals' use of, and attitudes towards, green and natural spaces is very strongly driven by their personal characteristics, most notably by age and health status. This, in turn, means that LA-level rates of *AnyVisit14* and, to a lesser but still surprising extent, *ImprovedSpaces*, are ***very strongly driven by the socio-demographic composition of LA populations***.

For local policy makers the improved precision of SAE estimates, and their ability to discriminate between rates in different LAs, may therefore be of limited *practical value* as it largely reflects the fact that these SAE models are particularly effective at combining evidence on the composition of local populations with evidence on the role that individual-level factors play in determining how people use and view green and natural spaces.

Thus, whilst it is clearly useful to improve the accuracy and precision of LA-level estimates, if local stakeholders' interests lie in gaining a better understanding of the impact of *policy-modifiable factors* then it is necessary, at the very least, to include a wider variety of locality data than has been possible in this study, preferably using more granular LSOA- rather than MSOA-level data.

This could be used to focus on potentially modifiable contextual factors such as the availability, accessibility and attractiveness of green and natural spaces to different groups in different localities.

Appendix 2: Creation and interpretation of SAE models

Standard variable selection procedures were used to specify the *AnyVisit14* and *ImprovedSpaces* models. Our approach was to define a standard binary logistic regression model with a two-category response variable, where $y_{ijk} = 1$ if individual i within MSOA j and Local Authority k responds positively to the *AnyVisit14* or *ImprovedSpace*s question, and $y_{ijk} = 0$ if they do not. For a logistic regression model, the probability that $y_{ijkt} = 1$ is calculated as:

$$p_{ijk} = \frac{exp(\eta_{ijk})}{1 + exp(\eta_{ijk})}$$

where η_{ijk} is specified as a linear predictor function $f(i)$ comprising individual, MSOA, and LA-level explanatory variables with corresponding coefficients (or parameter weights) *k=1,..,p* which indicate the relative effect on the outcome given the simultaneous effect of all other variables in the model. Thus;

$$f(i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_n x_{ip}$$

The specific variables included in the final models were as follows:

AnyVisit14 f Ageband + CarAccess + DependentChildren + EconActStatus + EducationalQuals + Ethnicity + GenHealthStatus + MaritalStatus + Sex + Month + COVID_Category + SpatialSignatureCategory + Green/BluespaceAcessibility + EducationDep + EmploymentDep + LADummy

ImprovedSpaces f Ageband + DependentChildren + EconActStatus + Ethnicity + GenHealthStatus + MaritalStatus + Sex + Month + SpatialSignatureCategory + IMD2019_GeogBarriers + LADummy

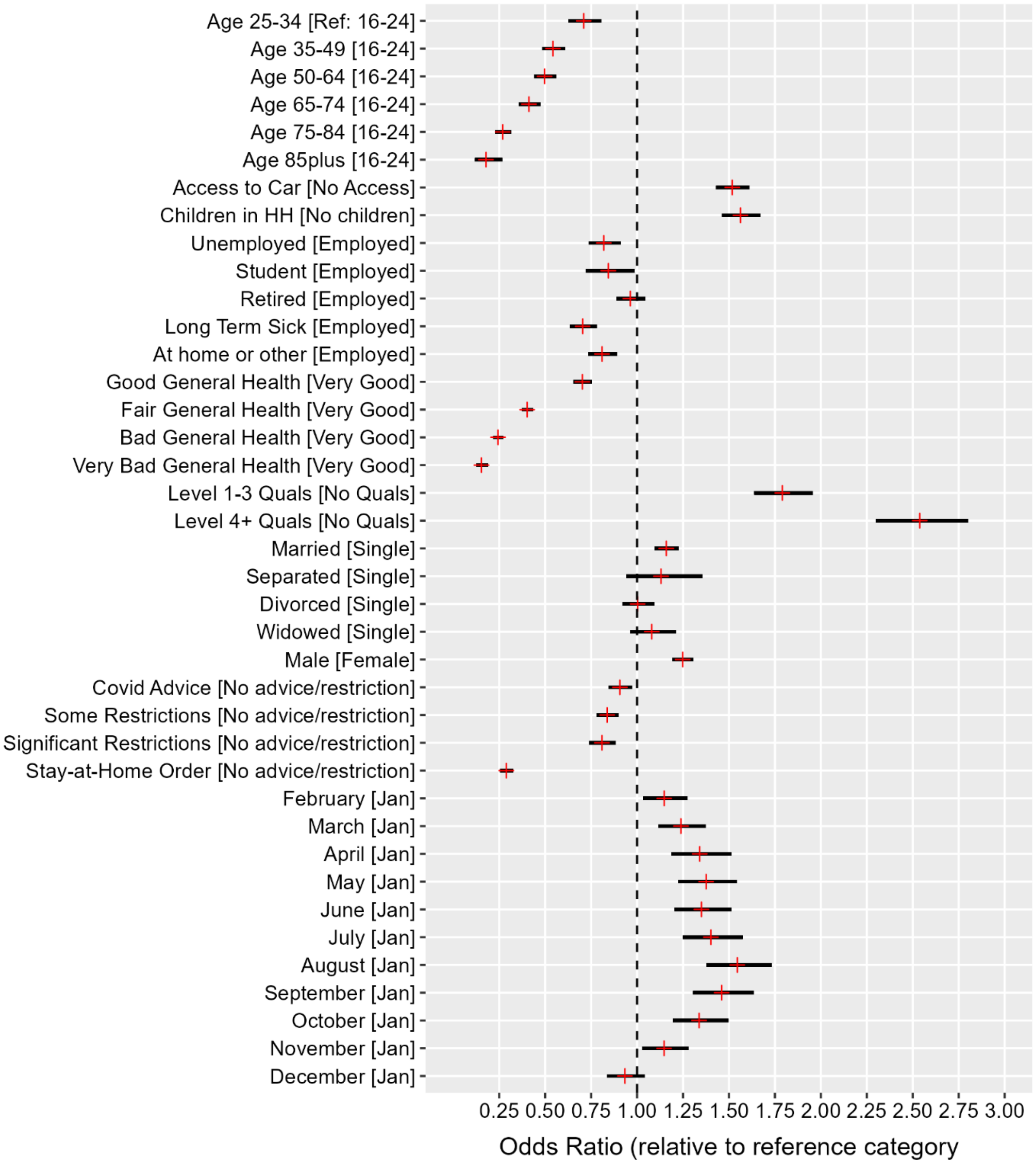
The *AnyVisit14* model

As illustrated in Figures 6 and 7 below, it is clear that many individual-level variables have a strong and well-defined impact on whether individuals visit green and natural spaces.

These graphs plot Odds Ratios (OR), which is the ratio between the odds of someone in a particular group having visited a green and natural space relative to the odds of someone in the reference group having visited. For example, the OR that someone with Very Bad General Health will have visited a green and natural space over the past 14 days relative to someone with Very Good General Health (the reference category) is 0.133. We are, moreover, very confident of the size of the effect [OR 95%CI: 0.125 – 0.188]. Similarly, the

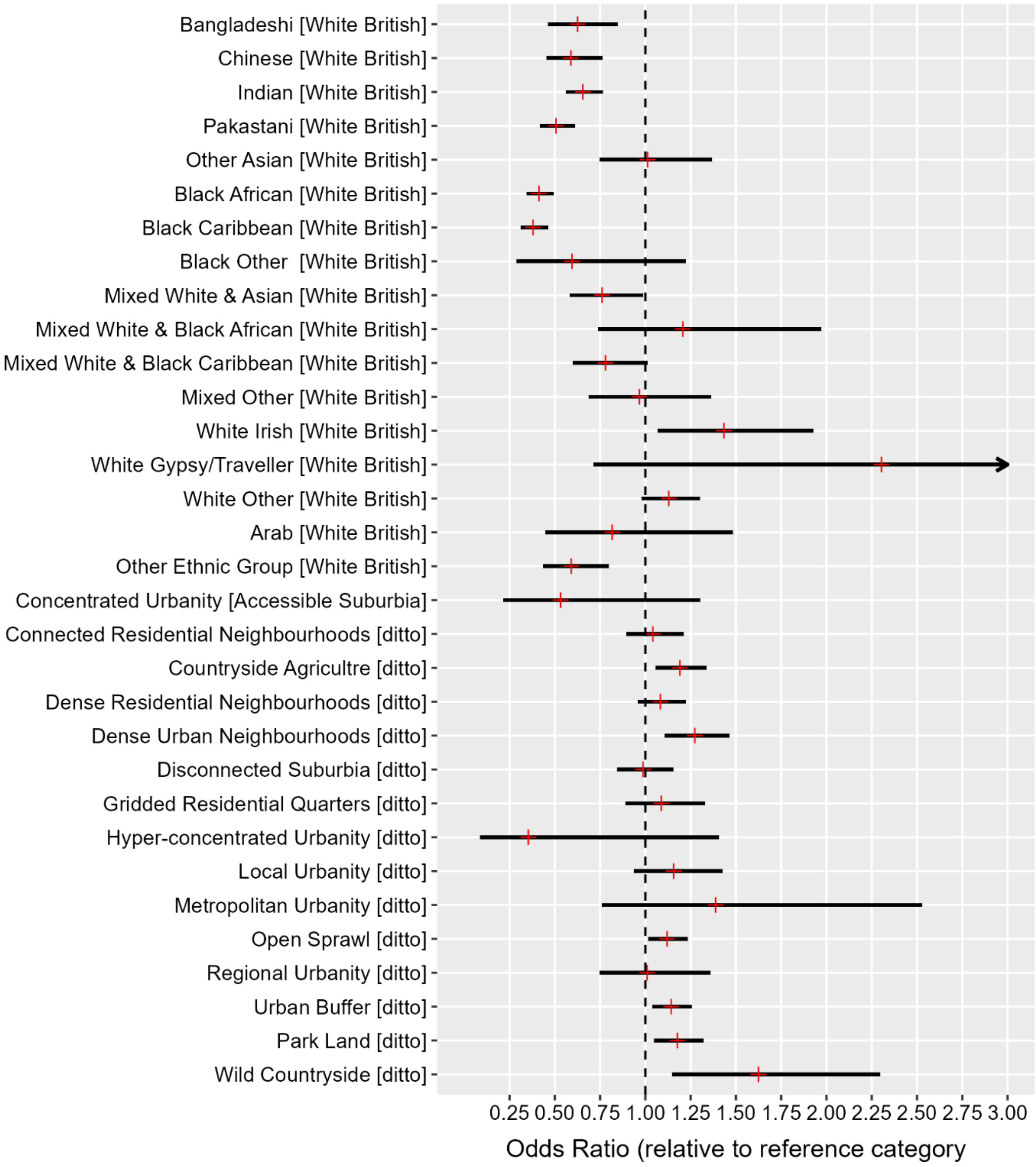
OR of someone aged 85+ having visited relative to someone aged 16-24 is 0.178 [OR 95%CI: 0.119-0.266].

Figure 6: Odds Ratios from the *AnyVisit14* model (Part 1)



Odds Ratios (detailed in Appendix 3), are useful for illustrating the relative impact of different factors and, as in these two figures, can be used to graphically illustrate the level of confidence we have in their effect on individual's visiting behaviour, but what really matters in SAE is how the model as a whole allows us to predict the *likelihood* that individual person-types living in different types of MSOAs and different LAs will have visited a green and natural space in the last 14 days.

Figure 7: Odds Ratios from the *AnyVisit14* model (Part 2)



For instance, leaving aside for the moment the additional impact of where people live, the lowest likelihood estimate (of just 0.6%) is for an *85plus* year-old *single female* of *Black Caribbean* ethnicity, who has *no educational qualifications*, is in *very bad general health*, is *long-term sick*, has *no access to a car* and *no children in her household*, and who responds in *December* when *Stay-at-Home COVID restrictions* are in place.

This contrasts with the highest likelihood estimate (of 98.9%) which refers to a *16–24-year-old married* male of *white gypsy or traveller* ethnicity, who has *degree-level educational qualifications*, who is *retired* but in *very good general health*, has *access to a car* and *has children in their household*, and who responds in *August* when there are *no COVID restrictions or advice about travelling or social mixing*.

These are intuitively reasonable estimates – the former almost certainly not having visited a green and natural space but the latter almost certainly having done so – but neither person-type, for various reasons, are likely to exist in the real world. The former actually represents an impossible combination of characteristics, not least because Stay-at-Home COVID restrictions were never in place in December. Nevertheless, the model predicts the likelihood associated with all theoretical person-types and these, adjusted to account for the nature of the MSOA and LA in which they live, can be applied to those individuals who, through the microsimulation of 2021 Census data, are estimated to comprise local populations.

MSOA-level effects and the LA Dummy Variable

With respect to the MSOA-level categorical variable illustrated in **Error! Reference source not found.** above (the Spatial Signatures Classification), people are, unsurprisingly, significantly more likely to have visited green and natural spaces if they live in areas classified 'Wild Countryside', 'Countryside Agriculture' or 'Park land' rather than in areas of 'Accessible Suburbia' (the reference category). Other outcomes are less obvious, with people living in 'Dense Urban Neighbourhoods' and 'Open Sprawl' also being more likely to have made such visits than people living in 'Accessible Suburbia'.

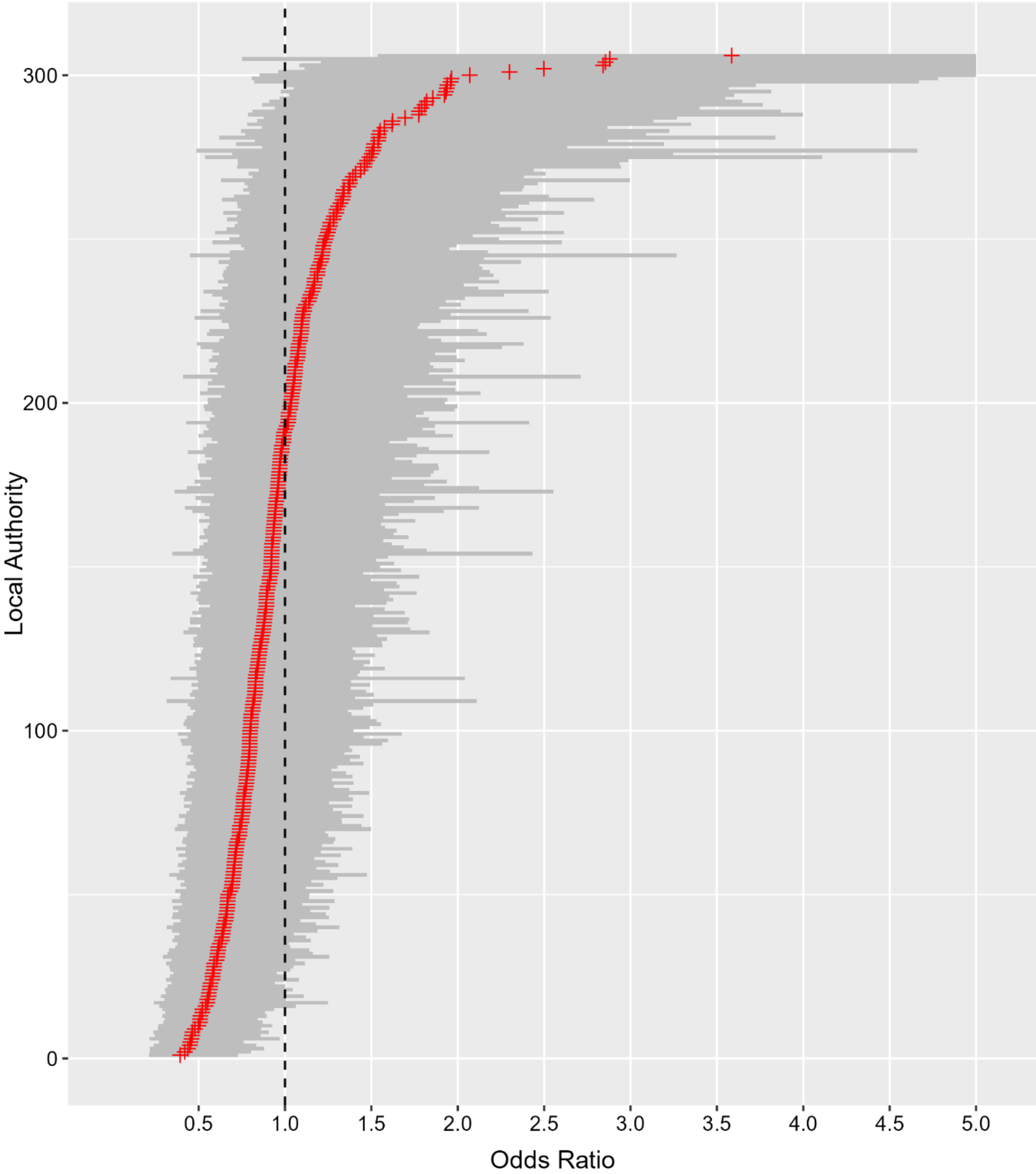
The *AnyVisit14* model also includes two continuous variables drawn from the Index of Multiple Deprivation 2019 (IMD2019), namely the *Employment* and *Education* deprivation sub-domain scores, as well as the *Access to Green/Blue space Domain Score* from the CRDC Access to Healthy Assets and Hazards (AHAH) dataset. The associated parameter estimates for each of these are given in Appendix 3, although they are interpretatively difficult as they describe the change in outcome associated with a unit change in the predictor variable, and each has a different scale and range.

Other things being equal, however, adults living in areas with lower access to green/blue space are, as one might expect, less likely to have visited a green and natural space over the previous 14 days, as are adults living in MSOAs with a higher education deprivation score (which reflects a lower levels of skills in the adult population and poorer educational outcomes among children). There is, on the other hand, a greater likelihood that individuals living in areas with higher Employment Deprivation (and thus a higher proportion of working-age people involuntarily excluded from the labour market) will, other things being equal, have visited green and natural spaces.

It is important, however, not to over-interpret these *predictive* relationships as these summary variables may well proxy a number of disparate local characteristics.

The final component of the *AnyVisit14* model is the LA-level dummy effect. Although differences between LAs can be marked, with relatively small sample sizes the 95% CIs around LA-level effects are, as illustrated in **Error! Reference source not found.** below, relatively wide. This figure uses North Norfolk LA as an arbitrarily chosen, broadly mid-point, reference category.

Figure 8: *AnyVisit14*: LA-level Odds Ratios (& 95% CIs) (base = North Norfolk)

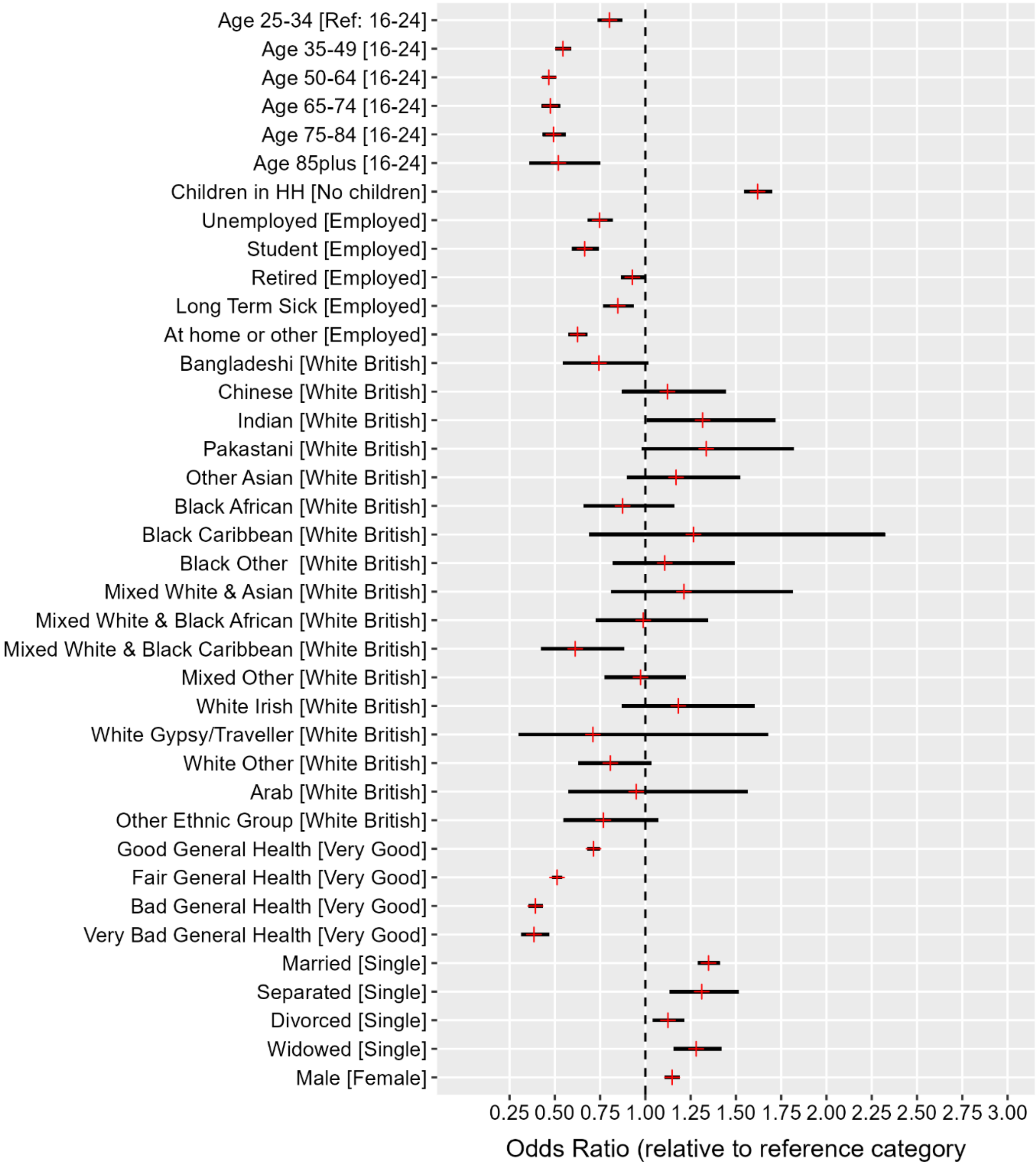


The ImprovedSpaces model

The Odds Ratios associated with individual-level characteristics are not as large as with respect to the *Any_Visits_14* model (Appendix 3 and Figures 9 and 10 below), but it is perhaps surprising that they are as large as they are. For instance, the odds that people aged 50 plus will say that local green and natural places have improved over the past five years is about half that of people aged 16-24 (the reference category), whilst the OR for people in Very Bad General Health relative to those in Very Good General Health is 0.385 [95%CI: 0.316 – 0.468]. This does seem to undermine the idea that responses to the M1_Q3 (*ImprovedSpaces)* question says very much at all about actual changes to local green and natural spaces.

Similarly indicative of the care that needs to be taken interpreting this response is that people are more likely to think that such spaces have improved if asked in April through November than if asked in January (the reference Month), as are people with dependent children compared to those without children [OR: 1.620; 95%CI: 1.544 – 1.699].

Figure 9: Odds Ratios from the ImprovedSpaces model (Part 1)

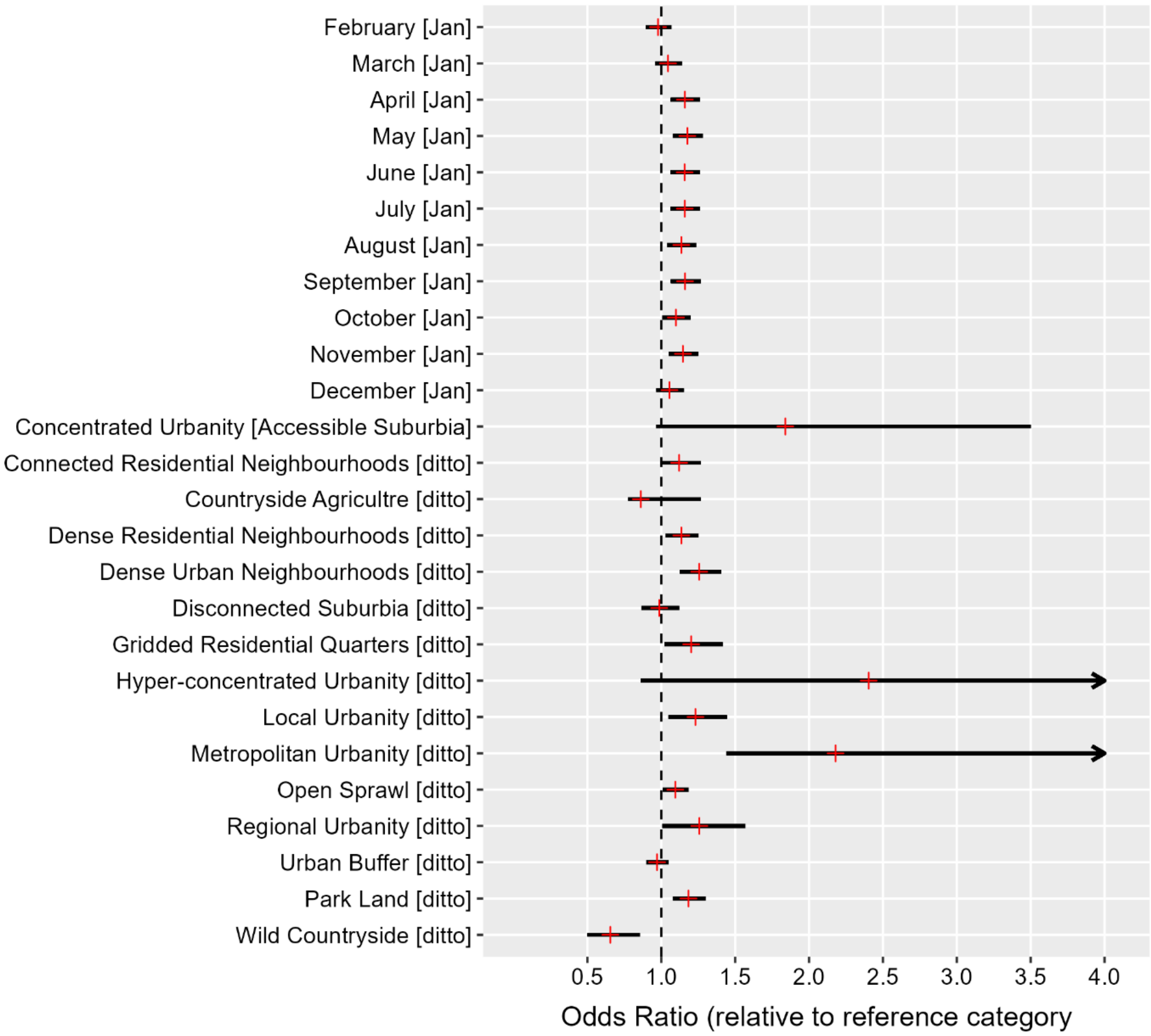


MSOA-level effects and the LA Dummy Variable

At MSOA-level, the *ImprovedSpaces* model includes the categorical Spatial Signatures Classification (**Error! Reference source not found.** below) and the numeric (and thus not p lotted) Geographical Barriers Sub-Domain Score from the 2019 Index of Multiple Deprivation (not plotted).

Interestingly, the only Spatial Signatures category with a significant <u>negative</u> effect on individuals' propensity to say that local green and natural spaces have improved is 'wild countryside'. People in many of the urban and residential categories are more inclined, other things being equal, to say that such spaces have improved. Reinforcing this urban-rural pattern is the fact that the Geographical Barriers Score, which measures the proximity of a variety of key services (post office, primary school, supermarket and GP surgery), suggests that people living close to such services (and thus with low levels of 'Geographical Barriers' deprivation) are also rather more likely to say that local green and natural spaces have improved (Appendix 4).

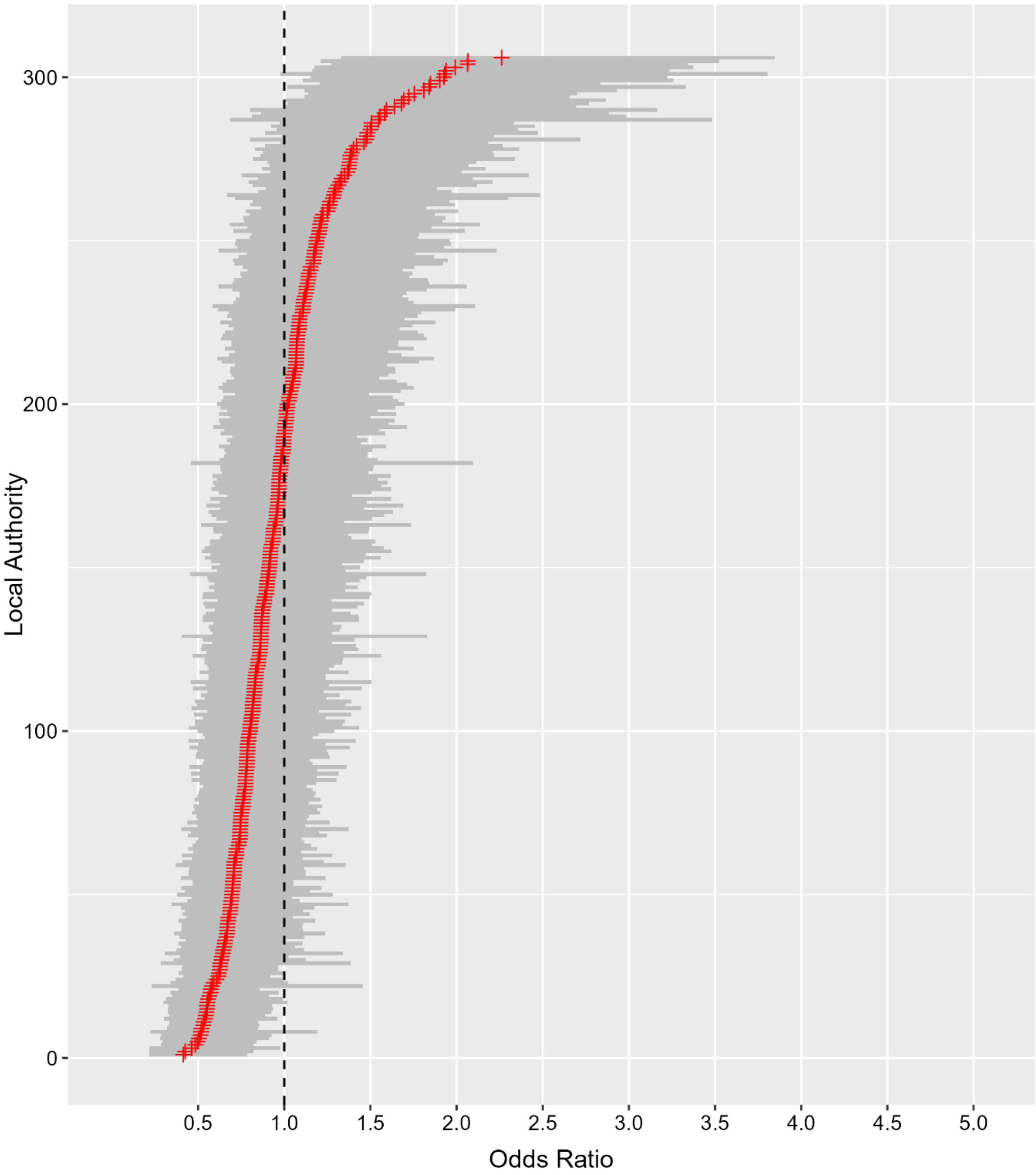
Figure 10: Odds Ratios from the ImprovedSpaces model (Part 2)



Turning, finally, to the role of the Local Authority dummy variable in the *ImprovedSpaces* model, these, as illustrated in **Error! Reference source not found.** below, are again a ccompanied by relatively wide 95% CIs although, as with respect to the *AnyVisit14* model, there are clearly significant differences between LAs.

In this case, Darlington LA has been chosen as a broadly mid-point reference category and, once again, the wide 95% CIs reflect the relatively small number of cases available for many LAs.

Figure 11: *ImprovedSpaces*: LA-level Odds Ratios (& 95% CIs) (base = Darlington)



Appendix 3 – 6 contain the date used to inform this report. See the [report page] to download the appendix data Excel file for full details.

Content:

www.gov.uk/natural-england