Ipsos MORI

August 2017

Small Area Estimation of Sport Participation and Activity using data from the Active Lives Survey

Prepared for Sport England

Pawel Paluchowski and Sarah Tipping

Contents

1 Background

Sport England required reliable and accurate estimates for participation in sport and physical activity (at least twice in the past month) and inactivity (less than 30 minutes in the past month), for the 6,791 Middle Super Output Areas (MSOAs) in England.

Ipsos MORI used Small Area Estimation (SAE) to produce these estimates using data from the 2017 Active Lives Survey (ALS).

The ALS is a rich data source. It provides the most comprehensive and authoritative picture of sports participation in England and is central to Sport England's measurement of its own strategy; with the survey also being used to provide Official Statistics and to help measure the performance of its own partners. The ALS has a relatively large sample size; ALS2017 contained 198,911 adults, with a minimum sample size of 500 in each English local authority, however the sample size per MSOA is too small to allow direct estimation

SAE is a useful technique used to generate estimates in small geographical areas that are either not covered by a survey or have such low coverage that any direct estimates will have prohibitively low levels of precision, manifested as very wide confidence intervals. There are a range of different approaches, however, most are variations on the same theme – the relationship between survey data and external data is used to predict missing survey data. We used a model-based approach. More details are given in Section2.

The three estimates, and corresponding 95% credibility intervals, were produced for all 6,791 MSOAs in England. MSOAs are small geographical areas with a minimum size of 5,000 residents and 2,000 households. They have an average population size of 7,500 residents and fit within local authority boundaries.

2 Description of method

This section outlines the process used to generate the small area estimates.

2.1 Setting up covariates

The first step was to derive and select the covariates. Potential covariates need to be available at both individual and population-level. This means we are unable to use many individual-level covariates as most population data would not be available at this level (for example, it would require us to have access to individual-level Census data). This means we did not use individual-level variables such as whether or not the respondent has poor general health, but instead use area-level variables such as the proportion of individuals in the MSOA with poor health. The exception to this is age and sex, since many Census data releases are broken down by age and sex, the population information is available.

A number of external data sources were used to identify variables to use as covariates. Potential covariates included Census data, ACORN, the Index of Multiple Deprivation and urban/rural classification aggregated to MSOA level. Additional administrative variables were also taken from the Neighbourhood Statistics website. We also utilised information from the Active Places dataset. A full list of the variables considered is given in Appendix A.

The previous small area estimation carried out by Gibson and Hewson in 2014 included ONS Neighbourhood Statistic Modelled Healthy Lifestyle Behaviours as covariates. We did not use these variables as they are estimates and contain bias and have varying levels of precision.

The collated data set contains missing values so these were imputed using mean values of the observations in the respective MSOA. After the group mean imputation of missing values, four other data transformation processes were applied to make the variables more suitable to an automated estimation process. This data reprocessing was implemented using the ‘caret’ package in R.

First, the variables in the data were transformed using a Box-Cox transformation. This is a parametric power transformation technique used to alleviate issues in the variables such as non-normality or heteroscedasticity. It is a very commonly used approach in forecasting with the aim to increase the predictive qualities of a predictor. The transformed variables were then centred and scaled in preparation of a principle component analysis (PCA).

PCA was used as the data set contains a large number of potential covariates, many of which were highly correlated. PCA extracts the unique information contained in the original variables and creates fewer synthetic variables minimising correlation between these. In this way, as much information as possible contained in the data set can be leveraged whilst minimising the number of variables added to the model. This data preparation process was applied to MSOA level variables but not to the two individual variables age group and gender.

TABLE OF CORRELATIONS HERE

2.2 Modelling the data

The MSOA-level variables were merged to the ALS and were used in a multi-level model to predict the two outcomes; participation in sport and physical activity and inactivity.

The modelling was carried out using the statistical programming language R and used the packages lme4 for multilevel modelling at the variable selection stage. The statistical multilevel modelling package MLwiN was used to establish the final model estimates and credibility intervals.

The modelling and estimation process was carried out in three distinctive steps. First, for each outcome, a different set of predictor variables was estblished using a stepwise selection method. The procedure implemented a multilevel model in which individual responses were nested within MSOA nested in districts. For performance reasons, this model was less complex in terms of the multilevel structure (as compared to the final estimation models) and used maximum likelihood (as opposed to the slow Bayesian estimation technique). All models were logistic regressions.

The selection process was set up in the following way. Initially, a model only with a constant would be run. Then, the routine would generate a set of models that each includes one of all the available predictor variables in the data set. Comparing the Akaike Information Criterion (AIC), the routine was able to determine the best of the models and hence, the best variable to be added. Then, this process would be repeated but now assessing all remaining variables by adding them to the better model established in the previous step. The selection of variables stopped when adding variables was not able to further improve the AIC value.

Once the selection process completed, the model is rerun in a more sophisticated multilevel approach that leverages spatial relationships and applies Bayesian estimation. This yields better estimates and enables the generation of credibility intervals around the central estimate.

Regular multilevel models account for spatial correlation between individuals and areas by grouping them together. However, this also creates boundary effects. Outcomes of all individuals within a region are correlated but are not correlated with outcomes of individuals in other regions. Such an assumption may be valid for countries or relatively large regions but is unlikely to apply to small areas such as MSOAs. This is the case because boundaries are to an extent artificial and because social behaviours are “contagious”. Therefore, individuals in a MSOA neighbouring a high-participation MSOA are also more likely to participate in physical activity. By fitting a multiple membership model, such neighbourhood relationships were included. This final model specification using the selected variables in a multiple membership framework used Markov Chain Monte Carlo (MCMC) estimation. MCMC is a Bayesian estimation tequnique which runs a large number (in this case, 5000) of models

and tries to find the most credible model parameters. MCMC estimations are shown to be more accurate than non-Bayesian approaches.

2.3 Generating small area estimates

The MSOA estimates are generated by applying the observed data for each MSOA in the final models which yield a predicted value. As individual level age group and gender variables were used, predictions for each age group by gender in the MSOAs were created first and then aggregated to a final MSOA level estimate. In a Bayesian framework, these central estimates are the most likely values for each MSOA.

Credibility intervals indicate the range of credible estimates around the central estimates. To generate these intervals, the 5000 MCMC model runs are used to generate 5000 different outcomes. The 5th and 95th percentile values provide a range of likely model outcomes for each prediction. To properly reflect the variation in data, the model standard error is then added yielding the final estimate. The intervals can be used to compare the likely ranges of estimates for each area. However, they should not be used in absolute terms.

Appendix A: Potential covariates

Variable name	Variable label	Source
Gend3	Gender - three bands	ALS survey
AGE	Age of respondent	ALS survey
Age8	Age - eight bands	ALS survey
DVAgeGroup	Age group	ALS survey
WTAgeGend	WTDV: Gender by age groups with imputation	ALS survey
IMDscore	Index of Multiple Deprivation (IMD) Score	Deprivation indices 2015, and sources
INCscore	Income Score (rate)	Deprivation indices 2015, and sources
EMPscore	Employment Score (rate)	Deprivation indices 2015, and sources
EDUscore	Education, Skills and Training Score	Deprivation indices 2015, and sources
HLTscore	Health Deprivation and Disability Score	Deprivation indices 2015, and sources
CRIscore	Crime Score	Deprivation indices 2015, and sources
HOUscore	Barriers to Housing and Services Score	Deprivation indices 2015, and sources
ENVscore	Living Environment Score	Deprivation indices 2015, and sources
IDACIscore	Income Deprivation Affecting Children Index (IDACI) Score (rate)	Deprivation indices 2015, and sources
IDAOPIscore	Income Deprivation Affecting Older People (IDAOPI) Score (rate)	Deprivation indices 2015, and sources
CYPSDscore	Children and Young People Sub-domain Score	Deprivation indices 2015, and sources
SKILLSDscore	Adult Skills Sub-domain Score	Deprivation indices 2015, and sources
GEOSDscore	Geographical Barriers Sub-domain Score	Deprivation indices 2015, and sources
BARSDscore	Wider Barriers Sub-domain Score	Deprivation indices 2015, and sources
INDSDscore	Indoors Sub-domain Score	Deprivation indices 2015, and sources
OUTSDscore	Outdoors Sub-domain Score	Deprivation indices 2015, and sources
midpop2012_0_15	Dependent Children aged 0-15: mid 2012 (excluding prisoners)	Deprivation indices 2015, and sources
midpop2012_16_59	Population aged 16-59: mid 2012 (excluding prisoners)	Deprivation indices 2015, and sources
midpop2012_60pl	Older population aged 60 and over: mid 2012 (excluding prisoners)	Deprivation indices 2015, and sources
inddep_indiv	Number of Income Deprived individuals	Deprivation indices 2015, and sources
incdep_child	Number of income deprived children	Deprivation indices 2015, and sources
incdep_older	Number of income deprived older people	Deprivation indices 2015, and sources
EmpDomNum	Employment Domain numerator	Deprivation indices 2015, and sources
educationpost16	Staying on in education post 16 indicator	Deprivation indices 2015, and sources
entryhighered	Entry to higher education indicator	Deprivation indices 2015, and sources
adult_english_prof	Adult skills and English language proficiency indicators - combined	Deprivation indices 2015, and sources
lifelostindicator	Years of potential life lost indicator	Deprivation indices 2015, and sources
disabilityindicator	Comparative illness and disability ratio indicator	Deprivation indices 2015, and sources
morbidityindicator	Acute morbidity indicator	Deprivation indices 2015, and sources

anxietyindicator	Mood and anxiety disorders indicator	Deprivation indices 2015, and sources
kmto_post	Road distance to a post office indicator (km)	Deprivation indices 2015, and sources
kmto_school	Road distance to a primary school indicator (km)	Deprivation indices 2015, and sources
kmto_store	Road distance to general store or supermarket indicator (km)	Deprivation indices 2015, and sources
knto_gp	Road distance to a GP surgery indicator (km)	Deprivation indices 2015, and sources
overcrowing	Household overcrowding indicator	Deprivation indices 2015, and sources
homelessness	Homelessness indicator	Deprivation indices 2015, and sources
houseafford	Housing affordability indicator	Deprivation indices 2015, and sources
housepoor	Housing in poor condition indicator	Deprivation indices 2015, and sources
housenoheat	Houses without central heating indicator	Deprivation indices 2015, and sources
airquality	Air quality indicator	Deprivation indices 2015, and sources
roadaccidents	Road traffic accidents indicator	Deprivation indices 2015, and sources
RUC11CD	Urban rural indicator - code	ONS
Child_development	Children achieving a good level of development at age 5	Public Health England website[1] ()
Crude_fertility_rate	Crude fertility rate, 2010-14	Public Health England
GCSE_achievement	GCSE achievement (5 A*-C incl. Eng & Maths)	Public Health England
Low_birth_weight	Births with birth weight less than 2500g as a proportion of live and still births with valid weight, 2010-14	Public Health England
Obese_Y6_kids	Percentage of measured children in Year 6 who were classified as obese, 2012/13-2014/15	Public Health England
OutOfWork_claimants	% of the working age population who are claiming out of work benefit, 2015/16	Public Health England
English_proficiency	Proficiency in English (% of people who cannot speak English well or at all)	Public Health England
Provide_care	Provision of Unpaid Care - 1 or more hours per week	Public Health England
healthy_life_m	Healthy life expectancy at birth (male)	Public Health England
Hospital_stays_alcohol	Hospital Admissions for Alcohol Attributable Harm (narrow definition)	Public Health England
Admissions_all	Emergency Admissions, All Causes	Public Health England
Incidence_cancer_all	Incidence of all cancers	Public Health England
x001	Count of the usual Residents	Census 2011 data
x002	Count of the number of Households	Census 2011 data
census_0_15	proportion population age 0-15	Census 2011 data
census_16_24	proportion population age 16-24	Census 2011 data
census_25_34	proportion population age 25-34	Census 2011 data
census_35_44	proportion population age 35-44	Census 2011 data
census_45_54	proportion population age 45-54	Census 2011 data
census_55_64	proportion population age 55-64	Census 2011 data
census_65_74	proportion population age 65-74	Census 2011 data
census_75plus	proportion population age 75+	Census 2011 data
census_males	proportion males	Census 2011 data
census_white	proportion from white ethnic background	Census 2011 data
census_ukborn	proportion born in UK	Census 2011 data
census_christian	proportion christianity	Census 2011 data
census_noreligion	proportion with no religion	Census 2011 data
census_working	proportion working - Ft, PT or self employed	Census 2011 data

[1] http://www.localhealth.org.uk/#v=map7;l=en (accessed 22/06/2017)

census_unemployed	proportion unemployed	Census 2011 data
census_retired	proportion retired	Census 2011 data
census_student	proportion student ft	Census 2011 data
census_homemaker	proportion home/family	Census 2011 data
census_SIC_A	proportion SIC_A	Census 2011 data
census_SIC_B	proportion SIC_B	Census 2011 data
census_SIC_C	proportion SIC_C	Census 2011 data
census_SIC_D	proportion SIC_D	Census 2011 data
census_SIC_E	proportion SIC_E	Census 2011 data
census_SIC_F	proportion SIC_F	Census 2011 data
census_SIC_G	proportion SIC_G	Census 2011 data
census_SIC_H	proportion SIC_H	Census 2011 data
census_SIC_I	proportion SIC_I	Census 2011 data
census_SIC_J	proportion SIC_J	Census 2011 data
census_SIC_K	proportion SIC_K	Census 2011 data
census_SIC_L	proportion SIC_L	Census 2011 data
census_SIC_M	proportion SIC_M	Census 2011 data
census_SIC_N	proportion SIC_N	Census 2011 data
census_SIC_O	proportion SIC_O	Census 2011 data
census_SIC_P	proportion SIC_P	Census 2011 data
census_SIC_Q	proportion SIC_Q	Census 2011 data
census_llti	proportion long term illness	Census 2011 data
census_noQual	proportion no qualifications	Census 2011 data
census_level1Qual	proportion level 1 qualifications	Census 2011 data
census_level2Qual	proportion level 2 qualificaiton	Census 2011 data
census_ApprenticeQual	proportion apprenticeships	Census 2011 data
census_level3Qual	proportion level 3 qualifications	Census 2011 data
census_level4Qual	proportion level 4 qualifications	Census 2011 data
census_nssec_1_2	proportion NSSEC 1&2 (professiona/managerial)	Census 2011 data
census_NSSEC_3	proportion NSSEC_3	Census 2011 data
census_NSSEC_4	proportion NSSEC_4	Census 2011 data
census_NSSEC_5	proportion NSSEC_5	Census 2011 data
census_NSSEC_6	proportion NSSEC_6	Census 2011 data
census_NSSEC_7	proportion NSSEC_7	Census 2011 data
census_NSSEC_8	proportion NSSEC_8	Census 2011 data
census_ownerocc	proportion owner occupiers/buying with mortgage	Census 2011 data
census_rent_socal	proportion social renters	Census 2011 data
census_rent_private	proportion private renters	Census 2011 data
census_Dwelling_Detached	proportion individuals in Detached	Census 2011 data
census_Dwelling_Semidetached	proportion individuals in Semidetached	Census 2011 data
census_Dwelling_Terrace	proportion individuals in Terrace	Census 2011 data
census_Dwelling_Flat_Purpose	proportion individuals in Flat_Purpose	Census 2011 data
census_Dwelling_Flat_Converted	proportion individuals in Flat_Converted	Census 2011 data
census_Dwelling_Flat_Commercial	proportion individuals in Flat_Commercial	Census 2011 data

census_Dwelling_Mobile	proportion individuals in Mobile	Census 2011 data
census_depchildren	proportion families without dependent children	Census 2011 data
census_nocars	proportion households with no car	Census 2011 data
census_TTW_home	proportion_mode of travel to work_home	Census 2011 data
census_TTW_underground_tram	proportion_mode of travel to work_underground_tram	Census 2011 data
census_TTW_train	proportion_mode of travel to work_train	Census 2011 data
census_TTW_bus_coach	proportion_mode of travel to work_bus_coach	Census 2011 data
census_TTW_taxi	proportion_mode of travel to work_taxi	Census 2011 data
census_TTW_motorbike	proportion_mode of travel to work_motorbike	Census 2011 data
census_TTW_car_van	proportion_mode of travel to work_car_van	Census 2011 data
census_TTW_passenger	proportion_mode of travel to work_passenger	Census 2011 data
census_TTW_bike	proportion_mode of travel to work_bike	Census 2011 data
census_TTW_foot	proportion_mode of travel to work_foot	Census 2011 data
census_TTW_Other	proportion_mode of travel to work_Other	Census 2011 data
census_TTW_No_Work	proportion_mode of travel to work_Do not work	Census 2011 data
Acorn_Category	Acorn_Category	ACORN 2016
Acorn_Group	Acorn_Group	ACORN 2016
Acorn_Type	Acorn_Type	ACORN 2016
Acorn_pCat1	Proportion in ACORN cat A	ACORN 2016
Acorn_pCat2	Proportion in ACORN cat B	ACORN 2016
Acorn_pCat3	Proportion in ACORN cat C	ACORN 2016
Acorn_pCat4	Proportion in ACORN cat D	ACORN 2016
Acorn_pCat5	Proportion in ACORN cat E	ACORN 2016
Acorn_pCat6	Proportion in ACORN cat F	ACORN 2016
popdens	Population density of MSOA (households/area in hectares)	Census/Geog info
SupergroupCode	ONS area classification - Supergroup Code	ONS
SupergroupName	ONS area classification - Supergroup Name	ONS
GroupCode	ONS area classification - Group Code	ONS
GroupName	ONS area classification - Group Name	ONS
SubgroupCode	ONS area classification - Subgroup Code	ONS
SubgroupName	ONS area classification - Subgroup Name	ONS

totalsites	total number of active places sites in MSOA	ACTIVE PLACES DATA
totalfacilities	total number of active places facilities in MSOA	ACTIVE PLACES DATA
ave_fac_per_site	average number of facilities per site in MSOA	ACTIVE PLACES DATA
proprefurb	proportion of facilities in MSOA that have been refurbished	ACTIVE PLACES DATA
propdisabaccess	proportion of facilities in MSOA that have disabled access	ACTIVE PLACES DATA
accessfree_per000	number of free access facilities per 1000 residents	ACTIVE PLACES DATA
accesspay_per000	number of pay to play access facilities per 1000 residents	ACTIVE PLACES DATA
accessclub_per000	number of club access facilities per 1000 residents	ACTIVE PLACES DATA
accessmember_per000	number of members only access facilities per 1000 residents	ACTIVE PLACES DATA
facpitch_per000	number of pitches (grass or astroturf) per 1000 residents	ACTIVE PLACES DATA
factrack_per000	number of tracks per 1000 residents	ACTIVE PLACES DATA
facpool_per000	number of pools per 1000 residents	ACTIVE PLACES DATA
facgym_per000	number of gyms (studios or health centres) per 1000 residents	ACTIVE PLACES DATA

facsportshall_per000	number of sports halls per 1000 residents	ACTIVE PLACES DATA
facgolf_per000	number of golf courses per 1000 residents	ACTIVE PLACES DATA
factennis_per000	number of tennis courts (outdoors and in) per 1000 residents	ACTIVE PLACES DATA
disabfacs_per000	number of disabled access facilities per 1000 residents	ACTIVE PLACES DATA
mgtocalauth_per000	number of facilities managed by the LA per 1000 residents	ACTIVE PLACES DATA
mgtcommercial_per000	number of facilities managed by a commercial company per 1000 residents	ACTIVE PLACES DATA
mgteducation_per000	number of facilities managed by an education establishment per 1000 residents	ACTIVE PLACES DATA
MYPE2015_allages	Total MSOA population	2015 mid-year population estimates
lake_count	Total number of lakes in the MSOA	lake geo file

There were a large number of potential covariates, many of which were highly correlated. Principal Component Analysis was used to summarise the variables.

Sarah Tipping
Statistical Consultant
Ipsos MORI
Sarah.Tipping@ipsos.com

Pawel Paluchowski
Statistical Consultant
Ipsos MORI
Pawel.Paluchowski@ipsos.com

For more information

Ipsos MORI
3 Thomas More Square
London E1W 1YW

t: +44 (0)20 7347 3000
f: +44 (0)20 7347 3800

www.ipsos-mori.com
www.twitter.com/IpsosMORI

About Ipsos MORI

Ipsos MORI, part of the Ipsos Group, is a leading UK research company with global reach. We specialise in researching Advertising (brand equity and communications); Loyalty (customer and employee relationship management); Marketing (consumer, retail & shopper and healthcare); MediaCT (media and technology); and Social & Political Research and Reputation Research. Over the past 60 years, the UK market research industry has grown in stature and in global influence. The companies that formed Ipsos MORI were there from the very beginning. In the Ipsos MORI story we trace the history of the firm, through its founders and luminaries, to celebrate how we have helped shape the research sector as well as the influences that have made Ipsos MORI what it is today.