department for
culture, media
and sport

Participation, engagement and intensity of sports participants

Analysis of the Taking Part Survey

Babatunde Buraimo (University Of Central Lancashire), Helen Jones and Pete Millward (Knight, Kavanagh and Page)

August 2010

improving
the quality
of life for all

Our aim is to improve the quality of life for all through cultural and sporting activities, support the pursuit of excellence, and champion the tourism, creative and leisure industries.

Contents

Executive summary 4
1. Introduction 6
2. Data and models for sports participation 8
3. Modelling the decision to participate in sport 10
4. Modelling the decision to participate by heterogeneous activities 19
5. Modelling engagement and intensity of sports participation 25
6. Conclusions 34
References 35
Appendix 1 Sport clusters 36
Appendix 2 Probit regression model 37
Appendix 3 Tobit regression model 41

Executive summary

This paper analyses various measures of involvement in sport and their relationship with a number of demographic and economic covariates. The analysis provides a detailed analytical appraisal of involvement in sport by investigating three dimensions. The first is **the individual's decision to participate in sport**. Second, for those who have made a positive decision to participate in sports, is **the level of engagement** measured by the number of days in the previous four weeks prior to the interviews. This is used to provide a detailed insight into the levels of commitment that individuals have to sport. This second dimension is important and allows decision-making to take into account not just whether individuals participate but the frequency with which they do so. The third dimension further enhances the analysis by examining the degree of intensity with which individuals engage. All factors being equal the second dimension apportions equal importance to two individuals with the same levels of engagement. However, in practice one may spend longer participating hence demonstrating a higher level of intensity and the third dimension focuses on this by analysing the **duration of participation in sport**.

For each of the dimensions noted, the paper uses different regression analyses to establish models of sports participation, engagement and intensity. Using data from the Taking Part Survey, probit regressions are used to develop models of sports participation. From these models, parameter estimates can be used to evaluate the probability of participation given a series of demographic, economic and geographic covariates. Similarly, tobit regressions are used to develop models of engagement and intensity. Tobit models are used to reflect the nature of the dependent variables. In the case of engagement; the number of days in the previous four weeks on which the respondent had participated in sport, the data does not fully conform to the properties of normal distribution; there are upper and lower truncations. The use of standard regression procedures in such instances are likely to produces biased estimates of the covariates' parameters. The use of tobit regression allows such potential biases to be corrected. Indeed a comparison of the standard regression and tobit regression procedures and an analysis of the dependent variables revealed that standard regression procedures were inappropriate. This was the case for measures of engagement.

Key findings of the various analyses are:

- In all three dimensions of sports, involvement by female participants was less than that of their male counterparts.
- The probability of participation by females was 8% lowers than that of males'; they participated in 1.8 fewer days; and spent one hour less when participating.
- In all the three dimensions but with varying degrees, age has a negative relationship with sport.
- Those with higher educational attainment were more likely to be involved in sports participation and have a higher level of engagement. The time spent participating was not significantly different to that of other groups.
- Sport overall is a normal good and increases in participation are associated with increases in the various dimensions of involvement in sport.

- Asian or Black respondents are less likely than White respondents to be involved in sports, when it comes to participation and engagement. With respect to intensity of participation, Black people spent less time than White people participating in sport.
- Lifestyle factors and health factors have a strong contributory effect with general positive health having a positive effect.
- Smoking had a negative effect on the dimensions of involvement in sport.
- The various dimensions of sports participation are strongly correlated with access to private transport and the lack of access to this clearly constrains involvement in sports participation
- While the number of children in a household had negative impacts on sport, having participated in sport when growing up had a positive effect on the various dimensions.

A series of parameter estimates for the covariates are presented and discussed. As well as treating sports as a homogeneous group, further analyses are presented on different types of sport using groupings from a cluster analysis by Williams (2010). The results complement those reported on sport as a homogenous group although caution must be exercised, as the low numbers of observations in the disaggregated analyses are likely to result in less reliable parameter estimates, however, the results are useful in that they provide insights into the behaviour of individuals participating in different types of sport.

1. Introduction

Participation in sport is of interest to a wide audience; sports and leisure practitioners, academics and policy makers. There have, as a result, been numerous studies investigating participation. Downward et al. (2009) presents the results of a survey of numerous studies on sports participation highlighting that many such studies use regression analysis to examine the relationships between measures of sports participation and a series of independent variables or covariates deemed to influence these.

The findings of many of these, although with different rationales, support one another in that the relationships between the dependent and independent variables conform to prior expectations. For example, the relationships provide evidence that sports participation is male dominated and that participation declines with age. Further to this and in instances where income data is available, positive income effects are reported. Many studies can be characterised by the nature of the independent variables, which generally offer information on the decision to participate in sport, or not. Consequently, models derived tend to be limited in that very little information on the quality of the participation can be derived. Often, little distinction is made between those who participated once over a specific period, normally the previous four weeks, and those who may have participated on say eight occasions. Both are regarded as sports participants.

It has not always been possible for empirical studies to distinguish between frequent and non-frequent participants. The availability of more detailed information on participation has aided this; principally, data on the frequency of participation. Data on frequency has meant that better inferences can be drawn about sports participation (for examples, see Gratton and Tice (1991) and Downward and Riordan (2007). The use of frequency of participation in specific studies complements those that have previously been limited to the core definition of participation. More recently, Downward and Rasciute (2010) using data from the Taking Part Survey took a novel approach focusing on the relative number of sports to other leisure activities. In this, the focus is not on whether individuals participate or their frequency of participation but on the number of sports activities relative to leisure (non-sport) activities in which they engage. Even more limited in empirical studies of sports participation are measures that provide a detailed view of the level of intensity of sports participation. There is an acknowledgement that analyses and policy decisions about participation should go beyond consideration of whether an individual is a sports participant or not and focus in more depth upon measures which calibrate levels of involvement in sport. While evaluation of frequency of participation has performed such duty, a more detailed approach to evaluation of the intensity of participation is valuable. For example, Gratton and Tice (1991) use a measure based on the frequency of participation and the energy expended.

The following report examines participation in sport and in doing so considers three dimensions. The first of these is the decision to participate. According to the

literature, once the decision to participate is made, the individual's focus shifts to decisions with regard to the frequency of participation (Gratton and Taylor, 2000). This two-stage process is affected by a series of determinants, one being cost. Gratton and Taylor (2000) theorise that the decision to participate in sport is determine by the overall cost while the decision on participation frequency is determined by the marginal cost of participation. There is no direct empirical analysis of this proposition given the absence of relevant data on the different cost of sports participation.

The second dimension is participation frequency. This function applies to those who have made a positive decision to participate in sport. While information and analysis about this is insightful, it stops short of fully informing decision makers on what they need to know about participants. For example, positive associations between sport and health are often hypothesised. The sport-health relationship, however, cannot be adequately examined by focusing on whether or not individuals participate in sport or, although more informative, the number of times they take part. Attention should focus on more detailed and informative measures that incorporate intensity. For this reason, the analysis in this study considers a third dimension; time spent by the individual in the course of his/her participation. The remainder of the report is structured as follows. Section 2 considers the data used to develop the various models covered. Section 3 models and presents the results of the decision to participate. The section that follows examines the decision to participate. However, unlike the preceding section that treats sport as a homogenous activity, in this, sport is treated as a series of heterogeneous activities. Section 5 models the engagement and intensity of sports participation. The final section contains relevant conclusions.

2. Data and models for sports participation

The central focus in this section is the various measures of participation in and demand for sport. Traditionally participation and demand for sport have centred on participation rates (for example, Coalter, 1999). General Household Surveys (GHS), one of the main sources of participation data, provide data on participation rates or the proportion of the population that takes part in sport both during the last four weeks and in the 12 months prior to interview. The extent to which formal econometric and statistical analyses could be used to evaluate participation and demand has, as a result, been limited. As the quality of data has improved, research on participation and demand has sought to provide formal explanations using econometric analysis. In this context, attention has generally focused on individual decision-making. Farrell and Shield (2002) use data from the 1997 Health Survey of England to analyse the propensity to participate in sport. Their analyses provide a detailed view of the (marginal) effects of a series of factors such as gender, income and educational attainment on the probability that an individual will participate.

This approach to analysing participation is limited in that whilst it provides an insight into the likelihood of involvement, it is unable to provide a view on other useful information such as the frequency and duration of participation; information that would be of value to decision-makers and those who formulate policies. The absence of such analysis is often due to data limitation. Influences that cause reduced or even non-participation are 'the holy grail' of participation studies and are, arguably, the key to informing policy makers and enabling them to better understand and consider the actions required to initiate or increase participation.

Developments in data collection over recent years have removed some of the constraints that previously limited the extent to which participation could be analysed. For example, previous analyses of participation using data from the GHS were limited to descriptive statistics appertaining to participation rates. However, Downward and Riordan (2007) use data from the GHS to examine sports participation and provide a more sophisticated analysis than had featured in previous studies. Downward and Riordan use the data to model the participation decision of individuals across three different clusters of activities: sport; recreation and leisure. Their modelling approach is innovative and makes use of the Heckman Selection Procedure to determine both the propensity to participate in sport and the frequency of participation conditional on this propensity. Developments in data collection that allow a more detailed treatment of participation have featured in other research, such as Sport England's Active People surveys. Consequently, measures of sports participation are no longer limited to participation rates but allow a variety of modelling approaches to be adopted. Data from the Taking Part survey allows for a more sophisticated treatment of participation and demand. Firstly, information is available on not just whether individuals participated in an activity or not but also on the frequency of participation and the time spent engaging in the activity. Consequently, a variety of demand measures can be configured. The focus of demand can now extend beyond

participation rates and include engagement and levels of intensity that incorporate the duration and time individuals allocate to participation.

The sample of observations comprises a series of 'batches' on a rolling basis. These batches, for convenience and robustness of analysis are aggregated into yearly surveys comprising of 12 months creating four survey instalments. The first of these involves interviews conducted from July 2005 to June 2006. The interviews in subsequent instalments were conducted from: July 2006 year to June 2007; July 2007 to June 2008; and July 2008 to June 2009 respectively. The sample sizes for surveys one to four are 28,117, 24,174, 25,720 and 14,452 respectively. Over the four instalments, various developments in interview questions have meant not only an increase in the number of questions (and corresponding variables) but also changes in their formats. Consequently regression analyses cannot always be based on the entire sample of data across the four years (N = 92,463). As the number of covariates used in regression expands, the sample size contracts reflecting the limitation that some variables only apply to one survey year[1]. A consequence of limiting the analyses to one year's data is that they effectively become cross-sectional rather than panel[2]. Estimates generated from cross-sectional analysis are often less efficient compared to those generated from panel data. A further advantage of panels is that this allows for unobserved characteristics not caught by any of the specified variables to be captured in fixed and random effect estimates. As a result, omitted variable biases are less likely to be a problem and the estimates of coefficients are more robust. These issues and others are considered in the following analyses.

[1] The regressions have generally been based on the fourth instalment of the survey. This instalment comprises a greater number of variables, some of which are not available in previous survey. Therefore, the use of these variables and data mean that observations from previous years are excluded from the analysis due to missing values.

[2] A panel dataset consist of repeated observations across the same unit. The unit could be households, time or individuals. The use of all instalments of the Taking Part Survey would have allowed panels based on months to be established. Panel estimates generally provide more robust parameter estimates (Verbeek, 2000)

3. Modelling the decision to participate in sport

The starting point for the analyses is the individual decision to participate. The dependent variable in this case is taken to be whether an individual had participated in sport in the four weeks prior to the interview. Data on whether interviewees had participated in the last 12 months prior to the interview are available for the first three years of the survey but not the fourth. While information on participation in the previous 12 months would be insightful, there is a sound rationale for the focus on participation in the last four weeks. First, the absence of the 12-month measure from the last survey means that analysis of participation over the longer period is less contemporaneous. Second, the 12-month measure is likely to be less reliable as interviewees are required to recall participation over a much longer period. Consequently, inaccurate recollection or embarrassment with regard to non-participation may result in the over-reporting of participation for the 12-month period. Third, the 12-month period is more likely to capture participation experiences that are both frequent and non-frequent. The advantage of the four-week measure is that it is more likely to capture frequent and, thus, greater commitment to participation. There is, thus, a need clear distinguish between frequent participation and non-frequent participation given that the experience and benefits gained from participation by regular participants are generally considered to be different to those derived by those who take part on an infrequent basis.

Modelling participation in the four weeks prior to the interview required the use of regression models in which the dependent variable is limited in response; in this instance a dichotomous variable taking the values of zero and one. In such instances probit and logit regression models are more appropriate than the standard linear regression model[3]. The use of linear regression (or ordinary least squares), while insightful, is likely to result in predicated values of the dependent variable falling outside the boundaries of zero and one[4].

To model participation in sport, the dependent variable in this case is $spor4wk$ and takes the value of one if the interviewee participated in sport in the 4 weeks prior to the interview and zero otherwise. The regression model to determine the probability is:

$$P(spor4wk = 1 \mid x) = F(\mathbf{x}, \beta)$$

$$P(spor4wk = 0 \mid x) = 1 - F(\mathbf{x}, \beta)$$

[3] The properties of logit and probit models are similar in that the distributions of the latent variable follow the same pattern. The core difference lies in the distributions of the tails. This difference is marginal and the results of logit and probit regression models are similar in practice.

[4] Another complication is that the disturbance term is heteroscedastic and is, therefore, in violation of the properties of linear regression, which require the disturbance term to have constant variance and zero mean. For a detailed appraisal of heterscedasticity, see Greene (2003).

where $\mathbf{x}$ is vector of covariates and β are the parameters to be estimated. This leads to the latent variable representation of the model that can be written as follows:

$$spor4wk^{*} = \mathbf{x}'\beta + \varepsilon$$

In the model, $spor4wk$ is an unobserved latent variable with a value that is to be determined and ε is the disturbance term. The following is therefore observed:

$$spor4wk = \begin{cases} 1 & if \quad spor4wk^{*} > 0 \\ 0 & if \quad spor4wk^{*} \leq 0 \end{cases}$$

In developing the participation model, the vector comprises a series of covariates. These variables are a combination of demographic and economic factors that are likely to influence the decision to participate. Given the data-gathering instrument, all data (besides age) is ordinal. For example, respondents were not asked for their actual income (as a ratio data) but to indicate their income on a scale, which has a ranking from one to 12; each ordinal increment representing an increase of £2,500 at the lower income ranges and £5,000 at the higher end. For modelling purposes and where possible, ordinal ranks have been converted to ratio data. In other instances in which there are no ranks between the categories, the data have been converted into categorical or nominal data. The vector of covariates and their corresponding summary statistics are noted in Table 1.

An additional set of covariates that are used are indices of deprivation[5]. The indices provide a relative ranking of the lower super output areas (LSOAs) across England. There are 32,482 areas with an average population of 1,500 people. Each of the LSOAs are ranked across seven domains, namely:

(1) Income:
(2) Employment
(3) Health Deprivation and Disability
(4) Education, Skills and Training
(5) Barriers to Housing and Services
(6) Crime
(7) Living Environment

In addition to the above domains, a composite index based in the seven domains is also available. For the purpose of this analysis, the ordinal rankings in each of the domains are used to produce deciles for each of the domains with decile 1 being the most deprived and decile 10 being the least deprived. The additional benefit of using these domains is that they provide additional information about the areas in which the respondents reside and how ranked measures of deprivation influence decisions to participate and engage in sport. The data range for each of the index is from 1 to 10; however, as they are ordinal data[6], their inclusion in the regression analysis should be as dummy variables. For example, if two areas had income deprivations of 2 and 4,

[5] See http://www.communities.gov.uk/corporate for a detailed appraisal of the indices of deprivation.

[6] As ordinal data, the differences between the ordinal points are not necessarily equal. For example, the difference in deprivation between decile *n* and decile *n+1* is not necessarily the same that between decile *m* and decile *m+1*. Consequently, if treated as ratio or interval data in regression analysis, the parameter estimates will be misleading.

this does not mean that the later has a measure that is twice as much (or has half the level of deprivation) as the former.

The final set of covariates is a geographical one. The regions of each of the respondents in the sample are also used in the analysis; nine in total. It there is a regional dimension to participation, engagement and intensity of sports participants, the dichotomous variables for these regions should be significant in relation to a nominated reference group, in this case London.

Table 1: Summary statistics of covariates

Covariate	Mean	Standard deviation
Female	0.555	0.497
Male	0.445	0.497
Age	49.898	18.726
GCSE/O Level grade A* - C(< 5 A*- C) and L1 equivalents	0.072	0.258
5 or more GCSE/O Level grades A* -C and L2 equivalents	0.193	0.395
A levels, vocational level 3 & equivalents	0.205	0.404
Higher Education & professional/vocational equivalents	0.289	0.453
Other Higher Education below degree level	0.128	0.334
Other qualifications: level unknown	0.050	0.218
Trade Apprenticeships	0.063	0.243
Income (mid-point)	£19,977	£17,992
Large employers and higher managerial	0.040	0.196
Higher professional	0.052	0.222
Lower managerial and professional	0.203	0.402
Intermediate occupations	0.160	0.367
Small employers and own account workers	0.084	0.277
Lower supervisory and technical	0.100	0.300
Semi routine	0.151	0.358
Routine	0.123	0.328
Never worked or long term unemployed	0.031	0.173
Full time	0.419	0.493
Long term sick or disabled	0.044	0.205
Looking after family/home	0.068	0.251
Part time	0.153	0.360
Retired from paid work	0.282	0.450
Student	0.030	0.170
Temporarily sick or injured	0.004	0.067
Asian	0.049	0.216
Black	0.033	0.180
Mixed	0.010	0.099
Other ethnicity	0.008	0.089
White	0.899	0.301
Single, never married/registered same sex civil partnership	0.291	0.454
Married and living with husband/wife	0.458	0.498
Married and separated from husband/wife	0.033	0.179
In registered same-sex civil partnership living with partner	0.003	0.053
Separated, but still legally in a same-sex civil partnership	0.001	0.024
Surviving same-sex civil partner, partner since died	0.000	0.008

Widowed	0.115	0.320
Divorced	0.099	0.299
Drinking: Don't drink	0.219	0.413
Drinking: Less often than once a week	0.277	0.448
Drinking: 1-3 days a week	0.324	0.468
Drinking: 4-6 days a week	0.085	0.279
Drinking: Every day	0.095	0.293
Smoking: No	0.777	0.416
Smoking: Yes	0.223	0.416
General health: Very bad	0.016	0.127
General health: Bad	0.063	0.243
General health: Fair	0.205	0.404
General health: Good	0.401	0.490
General health: Very good	0.315	0.465
Live sport coverage on TV	0.504	0.500
No live sport coverage on TV	0.496	0.500
Sports when growing up : No	0.193	0.394
Sports when growing up : Yes	0.807	0.394

Notes
The summary statistics are for the four year of the survey and each covariate includes zero values.

Empirical results of sports participation

The results of the probit analysis are presented in Table 2. Two models outline the marginal effects: the first focuses on sports participation in the four weeks prior to the interview and the second on a similar measure but places its emphasis on sports described as being of moderate intensity in which participation lasted at least 30 minutes. Additionally, it should be noted that in these models, sports are treated as homogenous and no distinctions are made between different types. Models treating sport as heterogeneous activities are presented at a later point in this study.
In all results, parameter estimates for the indices of deprivation were surprisingly insignificant. This suggests that various domains and levels of deprivation in respondents' LSOAs did not play any significant role in their decision to participate in sport nor did it play any role in the levels of engagement in and intensity of sport. One interpretation is that there is a degree of correlation between the covariates and the various indices. For example, the respondents' incomes are in themselves instrumental in constructing these indices and are likely to be correlated. Furthermore, the indices are likely to be endogenous and themselves influence by say the demographic characteristics of the respondents in the survey. As a result of the lack of significance in the parameter estimates, these indices were excluded from the regression analysis[7].

As with indices of deprivation, we find cause to exclude the region covariates from the models. Relative to the reference group (London), the regional covariates were not significantly different. This implies that the probability of participation in sport was not dependent on the regions of the respondents. This also applies to engagement and intensity of participation.

[7] Retaining variables that are insignificant is sometimes appropriate, however, this was considered inappropriate in the in the case of the indices of deprivation given their volume (63 in total). Retaining them reduces the degrees of freedom of the model.

The impact of sex and age are as *a priori*[8]*.* Controlling for other factors, female respondents are 8.1% less likely to participate in sports compared with male respondents. With respect to sports participation of moderate intensity, the gap is somewhat smaller reducing to 4.0%. The gender gap within sports participation is prevalent in survey data related to participation and is still evident in data from the most recent instalment of the Taking Part survey. Age has a negative impact on the propensity to participate – both general and moderate intensity. For every additional year, the probability of participation in sport reduces by 0.8% for sport (and 0.4% for sport of moderate intensity).

Educational attainment, income, economic activities, and socio-economic categories are all factors which might be expected to show some level of co-linearity. In such instances, the inclusion of these in a regression model can lead to biased and inconsistent parameters. The presence of co-linearity itself does not mean that any (group of) covariates that are co-linear should be excluded from the analysis. In order to determine such potential impacts a variable inflation factor test was used to test for co-linearity among the covariates. The statistical parameters suggest that it was not a problem and consequently education, income, economic status, socio-economic categories as well as others can justifiably remain as covariates within the models. For educational attainment, the reference group is those with five GCSE/O level grade at grades A* - C. Those with higher education and professional or vocational qualifications were 7.4% more likely to participation in sport. In contrast, those with trade apprentice qualifications were less likely to participate. The propensity to participate among those whose highest qualifications are less than five or more passes with grades A* - C in GCSE/O level was not significantly different from the reference group. As is to be expected, sports participation is positively correlated with income. From the analysis, a one per cent increase in income increases the probability of participation by 0.2% generally. The covariates capturing socio-economic status indicate that, relative to those who are long-term unemployed, there were no differences between the groups when it comes to sports participation. Similarly, economic activity does not influence the decision to participate after controlling for other factors.

As might be expected, access to personal transport has a positive effect on the likelihood of participation in the four weeks prior to the interview; it increases by 8.8% for sport in general. Not surprisingly for participation of moderate intensity, there is no discernable difference in the probability between those with or without cars. This suggests that for people without a car, the underlying motivation to participate with a higher degree of intensity overrides the necessity to have access to personal transport.

Ethnic background has an impact on the participation decision. The various ethnic categories from the survey were reduced to five and the reference group is White. For Asian and Black categories, the propensity to participate in sport is 8.2% and 16.4% respectively less, compared with those categorised as White when viewing sport in general. The participation decision of those categories as mixed or others was not significantly different from the reference group. In assessing participation of moderate intensity, the propensity to participate was 4.9% (at the 10% level) less for those categorised as Asian compared with the other categories. This result suggests that ethnicity does constrain sports participation.

---

[8] Participation in sport is predominately male biased and prior expectation is generally that female participation in sport is likely to statistically lower than male's.

On the influence of household characteristics on propensity to participate in sport, focal categories are marital status, number of adults (age 16 and over) in the household and the number of children. Marital status, on the whole, has limited influence on the probability of sports participation. Small noticeable influences are for moderate intensity sports participation in which those categories as In registered same-sex civil partnership living with partner have a higher propensity and those categorised as Widowed a lower propensity. There is no significant influence when it comes to the number of adults in the household. However, the number of children – those under the age of 16 years – does have a statistically significant effect. For every additional child in the household, the probability of sports participation by reduces by 1.9% for sport in general and 3.8% for participation of moderate intensity.

There are a series of health related covariates. These relate to drinking, smoking and self-reported health. Overall drinking and sports participation have a positive association. This relationship is more than likely reflecting the social dimension that accompanies many sports. This is likely to be particularly significant for those that are either team or partner-sports. Contrastingly, smoking's influence is negative and reduces the probability of participation by 7.7% for general participation in sport and 5.8% for participation of moderate intensity. Relative to those who report their health as being fair, those who report their general health as being bad are 12.0% less likely to participate. For those whose general health is categorised as good and very good, the propensity to participate generally increases by 4.3% and 11.4% respectively relative to those reporting fair, all other factors held constant. These findings are as to be expected particularly when moderate participation in sport is considered as greater levels of intensity are likely to require greater levels of health; as a consequence the marginal effects become more pronounced. In the context of sport participation in the four weeks prior to interview, this is not a serious issue as sport participation captures a whole variety of intensities from those likely to positively influence health to those which will have no impact on health. The participation-health relationship is examined in more detail when assessing intensity of engagement in sport (as measured in this report by participation duration).

Live sport coverage on TV and Sport when growing up, have the expected signs and are statistically significant at conventional levels. Watching live coverage of sport on television has a positive effect and the marginal impact is an increase in the probability of participation by 9.7% for participation in general and 3.6% for participation of moderate intensity. As to be expected, those who participated in sport when growing up have a greater probability of continuing their participation into adulthood, 16.6% for overall sports participation. With respect to the covariate Sports when growing up, some caution must be exercised. Conceivably, most adults participated in sport when growing up even if just exclusively at school. What is not clear is whether this variable captures PE led/based school sport and extracurricular sport or whether it just refers to the latter. In any case, those who reported participation in sport when growing up have a higher propensity to participate. The analysis is also robust in replacing the dichotomous variable Sports when growing up with a measure of the frequency of sports participation when growing up. The result of this analysis implies that participation when growing up only influences the probability of adult participation when the reported frequency has been at least once per week. This is an important point to note for policy formulation. Where self reported sports participation when growing up refers to sport outside school, the initiatives that cause the increased frequency of sports participation to the level of at least once per week among children are the ones likely to be effective is making such participation a sustainable feature, where all other factors remain constant.

The final set of covariates control for months. The reference group is January and relative to this, participation is likely to occur in August to November inclusive. The months of June and July were excluded because of (perfect) colinearity.

Given the above commentary on the results in Table 2, the likelihood of an individual take part in sport or, put another way, the probability of that a person will not take part in sport, can be established given a set of factors. The person least likely person to participate in sport has the following characteristics:

(1) Female
(2) Advancing in age
(3) Low educational attainment
(4) Low income
(5) No access to a car
(6) Asian or Black ethnic origin
(7) General health is bad and smokes
(8) Did not participate in sport when growing
(9) In a household with a large number of children

The individual with the highest propensity to participate in sport is simply the reverse: male; young with high educational attainment and high income; white with access to a car living in a household with no children; in very good health; this person drinks occasionally but does not smoke; and he participated in sport with a high frequency when growing up.

Table 2: Probit model of propensity to participate in sports

	Sport in last 4 weeks		Sport of moderate intensity in last 4 weeks	
COVARIATE	Marginal effect	T statistic	Marginal effect	T statistic
Female	-0.081***	5.13	-0.040***	3.08
Age of respondent	-0.008***	10.60	-0.004***	7.00
Education: *reference group is 5 or more GCSE/O Level grades A*-C*				
GCSE/O Level grade A* - C< 5 A* - C and L1 equivalents	0.017	0.60	-0.028	1.16
A levels, vocational level 3 & equivalents	0.048**	2.32	0.003	0.15
Other Higher Education below degree level	0.041*	1.73	0.004	0.20
Higher Education & professional/vocational equivalents	0.074***	3.52	0.031*	1.73
Other qualifications: level unknown	0.025	0.70	0.029	0.88
Trade Apprenticeships	-0.058*	1.78	-0.075***	2.78
Natural logarithm of **income**	0.023***	2.59	0.010	1.31
Socio-economic group: *reference group is long-term unemployed*				
Large employers and higher managerial	0.080	1.56	0.029	0.69
Higher professional	0.027	0.55	-0.015	0.37
Lower managerial and professional	0.033	0.75	0.025	0.71
Intermediate occupations	0.031	0.69	0.017	0.48
Small employers and own account workers	0.057	1.20	0.011	0.27
Lower supervisory and technical	0.026	0.57	0.006	0.17
Semi routine	-0.015	0.34	-0.019	0.54

Routine	-0.025	0.51	-0.017	0.46
Long term unemployed	-0.026	0.30	0.031	0.39
Economic inactivity: reference group is full-time				
Looking after family/home	0.025	0.72	0.018	0.56
Part time	0.031	1.56	0.007	0.42
Retired from paid work	-0.019	0.73	0.026	1.13
Student	0.097*	1.66	0.010	0.21
Long term **sick or disabled**	-0.030	0.58	-0.061	1.24
Temporarily sick or injured	-0.149	1.21	-0.072	0.67
Car use: Yes	0.088***	4.29	0.020	1.13
Ethnicity: reference group is White				
Asian	-0.082**	2.36	-0.049*	1.77
Black	-0.164***	4.47	-0.025	0.84
Mixed	0.023	0.34	0.057	0.96
Other ethnicity	-0.119*	1.66	-0.073	1.31
Marital status: reference group is Single				
Married and living with husband/wife	-0.006	0.32	0.008	0.52
Married and separated from husband/wife	0.024	0.64	-0.026	0.82
In registered same-sex civil partnership living with partner	0.130	1.34	0.120	1.56
Separated, but still legally in a same-sex civil partnership	0.239	1.19	-0.042	0.22
Widowed	-0.026	0.73	-0.072**	2.34
Divorced	0.000	0.00	0.031	1.41
Number of adults in household	0.003	0.27	-0.003	0.37
How many children under 16 live in this household?	-0.019**	2.31	-0.038***	5.53
Drinking: reference group is don't drink				
Drinking: Every day	0.068**	2.45	-0.004	0.17
Drinking: 1-3 days a week	0.099***	4.74	0.031*	1.70
Drinking: 4-6 days a week	0.118***	4.31	0.060**	2.43
Drinking: Less often than once a week	0.048**	2.28	0.021	1.10
Smoking: Yes	-0.077***	4.61	-0.058***	4.26
General health: reference group is Fair				
General health: Very bad	-0.085	1.05	0.004	0.05
General health: Bad	-0.116***	2.83	0.004	0.11
General health: Good	0.043**	2.24	0.094***	5.19
General health: Very good	0.114***	5.80	0.185***	9.85
Live sport coverage on TV	0.097***	6.84	0.036***	3.11
Sports when growing up	0.166***	8.91	0.074***	4.81
Month: reference group is January				
February	0.038	1.51	0.019	0.84
March	0.037	1.28	0.063**	2.43
April	0.099**	2.51	0.027	0.77
May	0.089	0.64	0.140	1.08
August	0.112***	4.18	0.109***	4.35
September	0.114***	4.97	0.076***	3.64
October	0.087***	3.36	0.083***	3.47
November	0.055**	2.40	0.023	1.11
December	0.005	0.18	0.011	0.42

Observations	6290		6290	
Ll	-3707		-3218	

Notes

(1) Dependent variable is participation in sport in the four weeks prior to interview

Estimates are marginal effects. T statistics are absolute values. *** p<0.01, ** p<0.05, * p<0.1

4. Modelling the decision to participate by heterogeneous activities

The above analysis is useful and provides a general view of the demand for sport as measured by the probability of participation based on doing so in the four weeks prior to the interview. In this approach all sports are treated as homogeneous, however, the propensity to participate in some sports is likely to be different to others. At the most detailed micro level, parameter estimates can be generated for every single sport and even "sub-markets" or segments within sports based on appropriate segmentation factors. The analysis in this section, however, disaggregates sport into 10 different groups:

(1) Athletics & rugby
(2) Dancing & yoga
(3) Outdoor sports
(4) Swimming, cycling & gym
(5) Racquet sports & running
(6) Bowling
(7) Cricket, football, pub sports & ten-pin bowling
(8) Boxing, martial arts & weightlifting
(9) Minor team sport
(10) Water sports

The groups are based on the 14 devised by Williams (2010); this used cluster analysis to group activities that are most commonly associated with each other. This approach grouped all 112 culture and sport activities, however, for the purpose of this analysis, the 69 sports covered in the study have been used to form 10 clusters. Some sports do not feature in the clusters, for example, Croquet, Pilates, Frisbee, Trampolining, Rambling and walking and Skittles. These are duly allocated to the appropriate clusters; croquet to group 7, pilates to group 2, frisbee to group 7, trampolining to group 2, rambling and walking to group 3, and skittles to group 7. Probit models on the dichotomous variable participation in sport in the four weeks prior to interview are developed for each of the 10 clusters. The advantage of this disaggregation is that the parameter estimates are able to reflect more precisely the groups of sports and their characteristics. For example, the impact of economic variables on participation in water sport is likely to be different from that of swimming. Similarly, the impact of demographic factors such as the number of children is likely to pose different marginal effects on the cluster containing swimming compared with the team sports cluster, as swimming intuitively tends to be a sport that many adults can participate in with children in contrast to a team sport. The full results of the probit models are reported in Appendix 2. The more prominent and significant impacts are annotated in Table 3 and discussed below.

Table 3: Summary of strength and direction of covariates' relationships with likelihood of sports participation, by heterogeneous sports groups

	Athletics & rugby	Dancing & yoga	Outdoor sports	Swimming, cycling & gym	Racquet sports & running
Female	-***	***	-***	***	-***
Age of respondent	-***		-***	-***	-***
Education: reference group is 5 or more GCSE/O Level grades A*-C					
A levels, vocational level 3 & equivalents				**	
Other Higher Education below degree level	*	*			*
Higher Education & professional/vocational equivalents		**	***	**	***
Trade Apprenticeships	-*				
Income	*			***	**
Socio-economic group: reference group is Long term unemployed					
Large employers and higher managerial			-*		
Higher professional			-**		
Lower managerial and professional			-***		
Intermediate occupations			-***		
Semi routine			-*		-**
Routine			-**		-*
Economic activity: reference group is Full time					
Never worked or long term unemployed			-**	-*	
Looking after family/home				**	
Part time	**	*		*	
Retired from paid work		-*	-**	-*	
Student	**				**
Car use: yes		**	***	***	
Ethnicity: reference group is White					
Asian	-*	***	-***	-*	
Black			-***	-*	-***
Other ethnicity			-**		
Marital status: reference group is Single					
Married and living with husband					-**
Married and separated from husband		-**			
Divorced	**				
Number of adults in household			-*		
How many **children under 16 live in this household**?		-***			
Drinking: reference group is don't drink					
Drinking: Less often than once a week				***	

Drinking: 1-3 days a week				***	
Drinking: 4-6 days a week			*	***	**
Drinking: Every day				*	
Smoking: Yes		-**		-***	-***
General health: reference group is Fair					
General health: Good			**	***	***
General health: Very good	*		***	***	***
Live sport coverage on TV	***			***	***
Sports when growing up		**	***	***	***

Table 3 (cont.): Summary of strength and direction of covariates' relationships with likelihood of sports participation, by heterogeneous sports groups

	Bowling	Cricket, football, pub sports & tenpin bowling	Boxing, martial arts & weight lifting	Team sport	Water sports
Female		-***	-***		-*
Age of respondent		-***	-***	-***	
Education: reference group is 5 or more GCSE/O Level grades A*-C					
A levels, vocational level 3 & equivalents			**		
Other Higher Education below degree level			*		
Higher Education & professional/vocational equivalents		-*	**	*	
Other qualifications: level unknown	**				
Trade Apprenticeships		-***	-*		
Income					**
Socio-economic group: reference group is long term unemployed					
Higher professional		-*			
Semi routine				-*	
Routine		-*			-*
Economic activity: reference group is full time					
Long term sick or disabled				**	
Part time		**		***	
Retired from paid work	***	*			
Student				**	*
Car use: Yes	**	**			**
Ethnicity: reference group is White					
Asian					-**
Black		-***		**	
Mixed			*		
Other ethnicity	*				
Marital status: reference group is Single					
Married and living with husband		-**	-***		

In registered same-sex civil partnership living with partner				**	
Separated, but still legally in a same-sex civil partnership	*				
Number of adults in household		*			
Drinking: reference group is don't drink					
Drinking: Less often than once a week				-**	*
Drinking: 1-3 days a week		***			*
Drinking: 4-6 days a week		**	*		**
Drinking: Every day	-*	**			
Smoking: Yes	-**	***	-*		
General health: Bad		-***			
General health: Very good			***		
Live sport coverage on TV	***	***		***	
Sports when growing up		***	**	***	

Notes
Estimates are marginal effects. T statistics are absolute values. *** p<0.01, ** p<0.05, * p<0.1

The marginal effects of being female, like the estimates presented in Table 2 in which sport is treated as a homogenous group are also pronounced in the disaggregated estimates. For nearly all the groups, the marginal effects of female participation, where statistically significant, are negative (athletics and rugby; outdoor sports; racquet sports and running; cricket, football, pub sports and ten-pin bowling; boxing, martial arts and weightlifting; and water sports[9]). However, for the two groups dancing and yoga, and swimming, cycling and gym, the marginal effects of female participation are positive and significant as to be expected. From the parameters, the group of sport with the highest propensity for female participation is swimming, cycling and gym (0.107); that of dancing and yoga was 0.051 or 5.1% more likely when compared with male participants.

The impact of age for many of the groups is similar to that shown for sport as a homogenous group. In nearly all of the sports, the relationship between age and the probability of participation is negative; the exceptions are dancing and yoga, bowling and water sport where age did not have any discernable effect on the probability of participation.

The relationship between educational attainment and participation is as previously noted, although across the groups of sports, the pattern is not systematic. The dominant group for educational attainment with respect to participation are those with higher education and professional/vocational equivalents and other higher education below degree level. A notable marginal effect of educational attainment on participation is the impact of the category higher education and professional/vocational equivalents on participation in the group cricket, football, pub sports and ten-pin bowling. The impact, although only significant at the 10% level is negative. This suggest that these activities do not attract those with higher educational attainments and this group is less likely to participate in the major team sports of cricket and football as well as pub games and ten-pin bowling.

[9] While these sports are significant at the 1% and 5% levels, water sport is significant at the 10% level.

The role of income is significant for only some of the groups, namely outdoor sports, swimming, cycling and gym, racquet sports and running and water sports. In these instances, the relationship between income and the propensity to participate are positive.

While there are correlations between socio-economic group and income, there are some significant impacts of the socio-economic categories. The most notable of these is the impact on the probability of participating in outdoor sport. Relative to the reference group (Never worked or long term unemployed), participation by those in large employers and higher managerial, higher professional, lower managerial and professional, and intermediate occupations were negative. There is a distinct division in the socio-economic groups with respect to outdoor sport which is not present in the other groups. Supplementing these variables of an economic nature is access to car. Car ownership or access to private transport is a strong determinant of sports participation. In only three of the 10 groups was there no significant effect. In the other seven, the relationship is always positive.

The negative marginal effects noted for non-whites is prevalent in some of the groups particularly those in the category Black. The impact of Asian on participation in dancing and yoga, however, is positive and likewise is the impact of Black on participation in team sports. This would seem to suggest that while those whose ethnicity can be described as black have a lower propensity to participate in sport compared with their white counterparts, they have a stronger preference for team sports.

With reference to household composition, the number of adults in the household does not have a systematic effect on the probability of participation, although there are notable exceptions at statistical margins. The number of children under the age of 16 has a strong negative impact on sports participation. Once sport is disaggregated into groups, it would seem that this negative impact predominately affects two clusters: athletics and rugby, and dancing and yoga.

On lifestyle matters, the effects of smoking are as to be expected and across five of the groups, the impact is negative, having controlled for other factors. For the other groups the impact is not significantly different from zero with the exception of the group Cricket, football, pub sports & tenpin bowling. The main driver for this positive relationship between smoking and participation in sport is likely to be the sub-group pub sports. The evidence pertaining to drinking, however, is that drinking and participation in sport are positively correlated. This is particularly notable in cricket, football, pub sports and ten-pin bowling, perhaps as to be expected, and in swimming, cycling and gym. While the relationship between the categories of self-reported general health are not always significantly different from zero, in instances when they are, they have the right sign: good and very good health having a positive marginal effect on the probability of sports participation while bad and very bad have a negative effect. Very good and good health, however, seems to have a greater positive impact on participation than the corresponding negative impact that poor or ill health has.

A very striking impact is that of sport when growing up. With the exception of bowling which is a sport that does not generally attract a young audience, and water sport, the impact of sport when growing up was positive at the 1% and 5% levels. The probability of participating in sport, relative to those who responded no to sport when growing up, ranges from 0.7% for team sport and 8.8% for the group cricket, football, pub sports and ten-pin bowling. This serves to highlight the impact and importance of sport among youth and the transition from youth to adult when it comes to sports participation. Finally, the link between live sports coverage on TV is overwhelming for a number of groups (or possibly vice versa). The evidences suggest that there is a

strong association between watching live sport on television and participating in sport. The groups for which there are no significant relationships are boxing, martial arts and weight lifting, water sport, dancing and yoga and outdoor sports. For many of these groups (or sport), their prominence on television is low (boxing, water sport) or they cannot readily be classed as spectator sports (dancing, yoga or mountaineering) hence, it is not surprising that the relationship with television is not statistically different from zero.

5. Modelling engagement and intensity of sports participation

The focus of this section is sport engagement and intensity. As with the preceding analysis, the 2008/09 Taking Part Survey is used. By engagement, emphasis is on the number of days an individual had engaged in sport, lasting at least 30 minutes, in the four weeks prior to the interview. This particular measure of engagement, which focuses on days, aggregates all sports activities and does not give additional weighting to those who engage in sport on numerous occasions per day. Furthermore, no additional weighting is given to participants whose engagements are across different sports. For example, a participant who took part in running and tennis on the same day has the same engagement measure as a participant who only participated in one activity.

As would be expected, the minimum and maximum values of this simple measure engagement are 0 and 28 respectively. The values of engagement are constrained between these limits and the analysis must consider this. If non-participants, zero values, are excluded from the data, the mean value of engagement is 11.6 days with a standard deviation of 9 days. Put another way, those who participated in sport in the last four weeks did so on 2.9 days per week. If the measure of engagement is modified to reflect the number of sport-days to take account of those participants who engage in more than one sport on given days, the mean engagement increases to 11.9 days with a standard deviation of 13.2 days. This measure does not account for the duration of participation whereas the simple measure of engagement is based on those whose duration of participation lasted at least 30 minutes. The main focus in this part of the analysis is to examine the number of days that an individual devotes to sport and less so to the structure of how these days are distributed across sport. For this reason the dependent variable is the simple measure of engagement, number of days an individual had engaged in sport, lasting at least 30 minutes, in the four weeks prior to the interview.

Given the characteristics of the engagement measure, ordinary least squares (OLS) regression can be used, however, as noted above, the values of the engagement lie between 0 and 28, effectively constrained at the lower and upper ends. For this reason, using OLS regression may result in biased estimates. An alternative approach would be to use a tobit model. Using a tobit model allows the estimation of the parameters to take into account that the data are censored and that the dependent variable has a truncated normal distribution. This truncated distribution is such that the tails of the distribution are unobserved and if this peculiar characteristic of the data is not taken into account, parameter estimates are likely to be biased. Essentially the distribution of the data is a mixture of discrete and continuous distribution representing the probability that an individual, firstly, participates in sport and, secondly, the number of days within a four-week period in which participation lasted at least 30 minutes.

This mixed distribution of the dependent variable by definition is a random variable $sportdays$ which is derived from the ‘true’ measure of engagement which take the

value 0 if participation is nil. For a detailed appraisal of the tobit model, see Greene (2003). The general formulation is given by:

$$sportdays^{*} = \mathbf{x}'\beta + \varepsilon$$

$$sportdays = 0 \quad if\ sportdays^{*} \leq 0$$

$$sportdays = sportdays^{*} \quad if\ 0 < sportdays^{*} < 28$$

$$sportdays = 28 \quad if\ sportdays^{*} \geq 28$$

If censoring in the data is not problematic, the parameter estimates from OLS will not be much different to those generated from a tobit regression. The test scores from a Hausman test showed that there were significant differences between the estimates of OLS and tobit models at appropriate levels. Using OLS is therefore not appropriate to model engagement.

The vector $\mathbf{x}$ comprises a series of covariates. As with the participation model, these covariates are a combination of demographic and economic factors that are likely to influence engagement.

The modelling approach is similar for intensity of sports participation. The main difference is that the dependent variable now focuses on the duration of participation measured in hours and is based on the most recent engagement in sport. This measure of sports participation captures the intensity and the time that participants devote to sports. One should acknowledge that for some sport there are very few degrees of freedom attached to the duration. For example, participants of team sport, say football (outdoors 11 a side) are both involved in time dating and the modal time spent on participation is the duration of a match; this is likely to be the case for many team sports or formal and organised sports activities that are part of sports leagues.

Given the dependent variable is measured in hours and its distribution has no upper constraint but is constrained at zero and therefore required a censored regression approach. As with the analysis of engagement, a tobit regression models is used but this time, only a lower limit constraint value is applied. The covariates remain the same. The mean value of the dependent variable, excluding zero values, is 2.25 hours with a standard deviation of 3.78 hours. As can be noted, a standard deviation to the left of the mean value produces a negative value and given that the minimum value of the dependent variable is zero, OLS is inappropriate.

Empirical results of sports engagement

Table 4 displays the results of the tobit model. Unlike the estimates of the probit model, the tobit (and the OLS) estimates exhibited multi-collinearity. This was tested for using the variable inflation factor test. Consequently, the group of variables representing socio-economic categories was excluded at the expense of income and educational attainment. The test statistic for the remaining estimates was comfortably within acceptable parameters to mitigate against collinearity between the covariates.

Considering the covariates for sex, age and educational attainment in turn, all other factors controlled for, the number of days of sport engagement is two days less for female participants compared with male in the four weeks prior to the interview. As with the propensity to participate in sport, engagement in sport is a male preserve. The marginal effect of age is negative and for every additional year of age above 16, engagement declines by 0.2 days in a four-week period. Put another way, an increase

in age by four years reduces the number of days of sports engagement in a four-week period by almost a day. In the case of educational attainment, those with less than five GCSE/O Level grades A* - C at the mean have 1.5 fewer participation days in the four weeks prior to the interview compared with the reference group (those with five or more GCSE/O Level grade A* - C); significant at the 10% level. The estimates of all other educational attainment variables were not significantly different from zero with the exception of those with higher education and professional qualifications. These individuals had a higher engagement rate participating in 2.1 more days in the four-week period. The education bias is prevalent in engagement too.

Turning our attention to income, economic status and car use, a 1% increase in income is at the mean associated with an increase in 0.8 days of engagement, again emphasising the fact that from an economics' perspective, sports engagement is a normal good. The impact of economic status to some extent reinforces the effects of income as those in the category long term sick or disabled, who are likely to be lower down the income scale, report 4.9 days fewer (over the four-week period) than those in full-time employment. The other significant covariate in this group is those looking after the family and home. The effect of this covariate is positive. In the absence of variables capturing the precise levels of leisure time available, it can be inferred that those in this group have disproportionately more leisure time than those in the other groups and this causes greater engagement in sport by 2.3 days. For car use, the effects are as to be expected allowing 2.3 more days of engagement.

Those from Asian and Black ethnic backgrounds have a lower engagement rate compared with White; 4.0 and 3.2 days respectively. While being Asian or Black are unlikely in themselves to impede sports engagement, there are underlying social and economic characteristics that are being picked up by these variables.

With respect to household composition (marital status; number of adults and number of children), those whose marital status is reported as divorced have a higher engagement rate of 2.0 days compared with those who are single. Parameter estimates of other marital statuses are not different to those of the reference group single. The number of adults in household does not have any statistically significant effect on engagement, however, the number of children does. For every child, engagement drops by 1.5 days in the four-week period. Children in the household, and increase in the number of children, seem to cause a substitution effect in which time originally devoted to sport is substituted with other forms of time expenditure.

Table 4: Tobit regression model for sports engagement in the last 4 weeks prior to interview

COVARIATE	Marginal effect	T statistic
Female	-1.807***	3.84
Age of respondent	-0.232***	10.64
Education: reference group is 5 or more GCSE/O Level grades A*-C		
GCSE/O Level grade A* - C(< 5 A* - C) and L1 equivalents	-1.532*	1.67
A levels, vocational level 3 & equivalents	0.791	1.22
Other Higher Education below degree level	0.457	0.62
Higher Education & professional/vocational equivalents	2.100***	3.36
Other qualifications: level unknown	1.053	0.90
Trade Apprenticeships	-2.437**	2.34
Natural logarithm of **income**	0.764***	2.79
Socio-economic group: reference group is		

Long term unemployed		
Long term sick or disabled	-4.859***	2.60
Looking after family/home	2.290**	2.08
Part time	0.933	1.53
Retired from paid work	0.016	0.02
Student	0.707	0.52
Temporarily sick or injured	-6.210	1.44
Car use: Yes	2.257***	3.49
Ethnicity: reference group is White		
Asian	-4.026***	3.72
Black	-3.172***	2.72
Mixed	0.335	0.16
Other ethnicity	-4.250*	1.86
Marital status: reference group is Single		
Married and living with husband	0.940	1.61
Married and separated from husband	-1.298	1.08
In registered same-sex civil partnership living with partner	3.926	1.49
Separated, but still legally in a same-sex civil partnership	-5.351	0.73
Widowed	-1.512	1.27
Divorced	2.070**	2.57
Surviving same-sex civil partner, partner since died	2.593	0.18
Number of adults in household	-0.312	1.10
How many children under 16 live in this household?	-1.455***	5.82
Drinking: reference group is Don't drink		
Drinking: Less often than once a week	0.733	1.07
Drinking: 1-3 days a week	1.814***	2.68
Drinking: 4-6 days a week	3.310***	3.78
Drinking: Every day	0.387	0.42
Smoking: Yes	-2.797***	5.37
General health: reference group is Fair		
General Health: Very good	7.117***	11.14
General Health: Good	3.398***	5.43
General Health: Bad	-3.224**	2.26
General Health: Very bad	-0.606	0.22
Live sport coverage on TV	1.944***	4.43
Sports when growing up	4.199***	7.01
Month: reference group is January		
February	1.593**	2.00
March	2.617***	2.88
April	3.042**	2.48
May	5.479	1.26
August	3.793***	4.41
September	2.956***	4.02
October	2.599***	3.11
November	1.433*	1.95
December	0.402	0.43
Constant	-6.984**	2.26
	6291	
	-15009	

Notes
Estimates are marginal effects. T statistics are absolute values. *** p<0.01, ** p<0.05, * p<0.1

From a lifestyle and heath perspective, drinking has a positive effect on engagement and the evidence seems to be suggesting that for the most part, participation and engagement in sport and drinking have some degree of complementarity attached[10]. While earlier, smoking was noted to reduce the propensity for sports participation, for those smokers who participate in sport, their levels of engagement is 2.8 days less compared with those who are non-smokers. The effects of self-reported general health are strong and as to be expected. Those who report their general health to be very good and good have 7.1 and 3.4 more days of engagement compared with those whose general health was reported to be fair. For those whose general health was reported to be bad, the engagement in sport is 3.2 days less compared with the reference group. A surprising result that is difficult to explain is that those whose general health is very bad have an engagement not statistically different to those whose health is reported to be fair.

Those who participated in sport when growing up have higher a level of engagement, 4.2 more days in the four-week period. In addition, there is a strong association between live coverage of sports on television and engagement reinforcing the relationship between sports participation and sports viewing. The final group of covariates, months, indicate that there is no strong pattern to engagement. The main finding is that relative to January, all other months show a greater degree of engagement with the exceptions of May and December. If a pattern is to be drawn out, it is that engagement from January increases reaching a peak in August, which is then followed by a gradual decline.

The results presented in the engagement model support those presented in the participation model and provide further substantiating evidence of the relationships between the demand for sport and demographic, economic, lifestyle and health factors. For policy makers, a notable objective is not just to increase sports participation but also to increase the levels of engagement in sport. The mean of engagement in the 28-day period is 11.6 days, however, the median is 8 and the mode, excluding zeros, is 4. This skewness suggests that engagement is not as optimistic as the mean of 2.9 days per week suggests and the central tendency for engagement is lower. How might engagement be readily increased? One of the substantive marginal effects within the covariates is the parameter of sports when growing up. Youth policies and sport are therefore important areas of policy development. Those responding positively to this factor have an increased engagement of over four days. Similarly, another substantive and positive covariate is that of health and while the relationship between general health and sport is complex, those with good and very good health have in excess of three and seven days additional engagements over and 28-day period. The effects of education are also substantive and progression within the education system has positive returns to sport; one could argue that this contributes to sports literacy. Consequently, an individual with a higher education degree who participated in sport when growing up and with very good general health on average will have an engagement rate of 11 days.

Empirical results of sports engagement by cluster groups

Replicating the tobit analysis for each of the sport cluster groups provides further insight on engagement in different types of sports. The main challenge in this

[10] It should be noted here that the issue of endogeneity is raised in that whilst it is noted that causality runs from drinking to engagement, the covariates drinking could be endogenous in that causality may run from engagement to drinking. This is an area for further detailed empirical research.

disaggregated analysis is the reduced degrees of freedom due to the smaller samples. Consequently, the values of the parameter estimates may not be precise. The only resolution is to increase the number of observations in each of the categories. Given the selected covariates, this is not possible. The number of observations could be increased by using data across all four instalments of the Taking Part survey. However, since some of the covariates are only available in the fourth survey, such variables would have to be omitted. Given the significance of these variables, the empirical results are likely to suffer from omitted variable bias. Estimates that are less precise are preferable to those that are biased therefore the analysis is based on the fourth instalment of the survey and the disaggregated analysis by various sports groups may not be as precise and this should be taken into account when interpreting the estimates.

The marginal effects of the covariate female are in most cases is negative and significant. The sign on female for outdoor sports is positive, however, the value is not significantly different from zero given the magnitude of the t statistic. The other sport clusters in which female is not significantly different from zero are bowling, boxing, martial arts and weightlifting. Similarly, age of respondent has the expected sign in the many cases in which it is significant. An increase in age is associated with reduction in engagement. Although the marginal effects of the various categories of educational attainment are not in the main significant across the sport cluster, in cases where they are significant, they have the appropriate sign. For example, the impact of higher education and professional is positive on outdoor sport cricket, football, pub sports and ten-pin bowling, and boxing, martial arts and weightlifting. The effects of income is somewhat compromised and the parameter estimates across the various sport clusters is not significant. The key observation is that the number of observation used in estimating the parameters in many of these cases are relative low given the nature of the model being used.

The only models in which there are significant numbers of observations and consequently high degrees of freedom are outdoor sports, racquet sport and running, and cricket, football, pub sports and ten-pin bowling. For the other models, the estimates are unreliable and should be viewed as exploratory and not as a definitive means of assessing the relationships between the covariates and the dependent variable. For suitable advances to be made, increases in the number of observations by individual sports or sport clusters are necessary.

The engagement of female participants in outdoor sport was not significantly different from that of male's. For racquet sports and running, and the group comprising cricket and football, female engagement was 2.6 and 2.7 hours less that male's respectively. In the case of outdoor sports, and to a very small extent of the group that comprises cricket, higher educational attainment resulted in a greater level of engagement. This was not the case with racquet sport and running and there was no discernable pattern regarding the relationship between education and engagement in sport.

The dominant relationship emerging from the economic activity covariates were those of students, looking after family and home, and retired from paid work with regards outdoor sport. Those who had retired had an increased engagement over and above the reference group (full time) of 4.3 hours in the previous 4 weeks. For students this was a reduced engagement of 4.3 hours (significant at the 10% level). For those looking after family and home, the engagement was higher than the reference group by 4.6 hours. This to some extent emphasises the availability and the relative cost of leisure time. In the case of students, the relative lack of engagement, across many sports groups, may be a substitution effect between leisure time and study time, the evidence in Appendix 3 suggesting that the latter is more costly than the former.

Other significant covariates of note for outdoor sport are that drinking and engagement negatively related, the is a positively strong health association with smoking having a negative effect on engagement and general health described as good or very good having a strong positive association with levels of engagement. For the group comprising cricket, the category Asian has a negative impact on the number of hours devoted to those sports. This is interesting in itself as 69%[11] of respondents in this group reported not to drink and from the earlier probit model, drinking and participation are positively correlated.

Empirical results of intensity of sport participation

The general associations and patterns that were established in the participation and engagement models are prevalent in the duration of sports participation model in Table 5. For female participants, the level of intensity measured in hours is 1 hour less than male respondents. Given that the mean value is 2.25 hours, this represents 43% of the time spent by male participants on sports. For age, the negative correlation implies that the older the participant, the less time they spend participating in sport. For educational attainment, there are no differences in the parameter estimates of the covariates. This is interesting in that, at the first and second stages in which an individual firstly decides whether to participate in sport and how frequently that participation should occur, educational attainment plays a role in decision making, however, once the decision to participate and engage had been made, education does not influence the time spent. In part and as noted earlier, there are limited degrees of freedom on how much time a participant spends participating in sport, particularly sports which involve partners and teams.

Given the low degrees of freedom involved in time spent on sport, income still has a significant effect and a 1% increase in income increases the duration by 0.2 hours or 12 minutes. Other economic indicators are not significant. No categories in economic status or socio-economic status were significantly different from zero, the categories in the later group were subsequently dropped as test statistics indicated they were the source of multi-colinearity. Quite interesting but not surprising is the positive marginal effect of car use which was significant at the 1% level and increase duration by 37 minutes. The relative ease of access to facilities resulting from access to private transport plays an important role in the length of time participants participate in sport. The significant marginal effects noted on the ethnic minority covariates Black and Mixed is difficult to explain. Why would those whose ethnicity is Black spend less time participating in sport relative to White having controlled for a series of other influences? Similarly, why would an individual of Mixed ethnic background spend more time participating in sport? It is difficult to imagine that these two groups have disproportionately less and more access to leisure time.

For every additional child in the household, the duration of participation reduces by 10 minutes (significant at the 10% level). Although marginal, this implies that participants in households with children have less leisure time to devote to sport. For some this diminishing amount of sport time results in non-participation as shown in earlier analysis of the probability of participation.

[11] The proportion of non-drinking in other ethnic groups are 17% for White, 49% for Black, 34% for mixed and 47% for others; the mean proportion for non-drinking is 22%.

With respect to lifestyle and health covariates, drinking per se and the frequency of drinking has positive associations with sport and the duration of time spent on sport. Interestingly, it is the low frequency of drinking – drinking less than once a week and drinking: 1-3 days a week – that are significant. Greater drinking frequencies have no significant effect on participation. Those with moderate drinking frequencies spend more time participating in sport than others. The patterns that were established earlier with general health prevail. The general health of respondents influences how long they spend participating in sport. The most substantive estimate in this category was that of general health: very good with a marginal effect of 1.2 hours compared with those of fair general health. Those who reported general health to be good spent 27 minutes more on sport compared with the reference group. Those with bad general health had a negative marginal effect while those whose general health was reported to be very bad spent a similar amount of time to that of the reference group.

Watching live coverage of sport on television is associated with increase sport time. Live coverage of sport on television can be inferred to promote all dimensions of sport: participation, engagement and intensity. Similarly participating in sport when growing up implies that positive sports youth policy that encourage young participants to engage in sport are key to adult involvement in sport.

Table 5: Duration of sports participation in hours.

COVARIATE	Marginal effect	T statistic
Female	-1.003***	-0.05
Age of respondent	-0.045***	0.20
Education: reference group is 5 or more GCSE/O Level grades A*-C		
GCSE/O Level grade A* - C(< 5 A* - C) and L1 equivalents	0.007	0.13
A levels, vocational level 3 & equivalents	0.198	0.01
Other Higher Education below degree level	0.019	-0.05
Higher Education & professional/vocational equivalents	0.131	0.02
Other qualifications: level unknown	-0.050	-0.15
Trade Apprenticeships	-0.154	0.19
Natural logarithm of income	0.193	-0.49
Socio-economic group: reference group is Long term unemployed		
Long term sick or disabled	-0.493	-0.23
Looking after family/home	-0.228	0.04
Part time	0.037	0.00
Retired from paid work	0.003	0.08
Student	0.079	-2.61
Temporarily sick or injured	-2.606	0.62
Car use: Yes	0.620***	-0.53
Ethnicity: reference group is White		
Asian	-0.528	-1.33
Black	-1.327***	1.51
Mixed	1.514**	-1.01
Other ethnicity	-1.014	-0.07
Marital status: reference group is Single		
Married and living with husband	0.197	0.43
Married and separated from husband	0.430	2.55
In registered same-sex civil partnership living with partner	1.087	0.20
Separated, but still legally in a same-sex civil partnership	2.546	2.01

Widowed	-0.075	0.22
Divorced	0.224	1.09
Surviving same-sex civil partner, partner since died	2.005	0.13
Number of adults in household	0.127	-0.16
How many children under 16 live in this household?	-0.157*	0.60
Drinking: reference group is Don't drink		
Drinking: Less often than once a week	0.510**	-0.10
Drinking: 1-3 days a week	0.598**	0.57
Drinking: 4-6 days a week	0.572*	0.72
Drinking: Every day	0.724**	0.51
Smoking: Yes	-0.099	-1.02
General health: reference group is Fair		
General Health: Very good	1.231***	0.70
General Health: Good	0.429*	-0.56
General Health: Bad	-1.018**	0.43
General Health: Very bad	-0.562	1.23
Live sport coverage on TV	0.698***	1.03
Sports when growing up	1.029***	0.34
Month: reference group is January		
February	0.343	0.45
March	0.452	0.88
April	0.880*	0.29
May	0.288	1.09
August	1.090***	1.14
September	1.140***	0.80
October	0.801***	0.80
November	0.800***	0.49
December	0.490	-5.10
-5.101***	-4.53	
Observations	6291	
ll	-9600	

Notes
Estimates are marginal effects. T statistics are absolute values. *** p<0.01, ** p<0.05, * p<0.1

6. Conclusions

The analyses have taken a progressive view of involvement in sport. The starting point for individual decision-making is whether they should participate or not. Given this process, much of the data from the Taking Part survey is effective for analyses, however, analyses are limited to the last instalment of the survey as some questions were only asked in the last edition of the interviews and consequently data from these questions are not available in earlier editions. While the modelling in this paper has been limited to the last edition of the survey, modelling the decision to participate in sports has still been rich in that observations from both participants and non-participants are valuable. From the analysis of participation, female participation in sport is consistently limited compared with male participation. This is the case for sport in general as well as for specific groups of sport. The only sports groups in which female participation dominates that of male are dancing and yoga and swimming, cycling and gym. For a small number of sports clusters there was no significant difference in male and female participation. The evidence regarding age is as to be expected and sport participation follows a general pattern in that as age increase, the probability of sports participation, either as a homogenous group or as sports clusters, declines. There are some exceptions for sport clusters (dancing and yoga, bowling and water sports).

As to be expected, multi-collinearity between groups of variables namely educational attainment, socio-economic groups, economic activity and income, could not be ruled out, however, a test statistic showed that collinearity between the variables was not problematic in the case of the probit model. Not all of the educational and economic covariates were significant; however, income was consistently influential and had a positive marginal effect on the decision to participate. Lifestyle factors were also considered: drinking, smoking, live sport on television and whether participants took part in sport when growing up. As shown in previous studies, there is a positive association between drinking and participation, although this may well be a manifestation of the social aspect of sports participation. Smoking in contrast has a negative effect on participation in sport. As one would expect, those who participated in sport when growing up had a higher probability of continuing this participation into adulthood.

While the analyses of participation used observations for participants and non-participants, the number of observations for engagement and intensity of participation was lower as these were based on participants only. The nature of the data is such that standard regressions models were inappropriate; the dependent variables being censored and not conforming to the properties of normal distribution. For both engagement and participation, the relationship between the covariates and the dependent variables were generally as expected. While the models that treat sport as homogenous are robust in their estimates, those of the sports clusters are likely to lack precision as the number of observations across some of the clusters was limited; in some instances, the number of observations was just over 100. For this reason, the disaggregated models of sport engagement should be viewed with some degree of caution.

References

Coalter, F. (1999) Sport and recreation in the United Kingdom: flow with the flow or buck the trends? *Managing Leisure*, 4(1), 24 – 39

Downward, P. and Riordan, J. (2007) Social interactions and the demand for sport: an economic analysis, *Contemporary Economics Policy*, 25(4), 518-537

Downward, P. and Rasciute, S. (2010) The relative demands for sports and leisure in England, *European Sport Management Quarterly*, 10(2), 189-214

Downward, P., Dawson, A. and Dejonghe, T. (2009) Sports Economics: Theory, Evidence and Policy, London, Butterworth-Heinemann

Farrell, L. and Shield, M. A. (2002) Investigating the economic and demographic determinants of sporting participation in England, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(2), 335-348

Gratton, C. and Taylor, P. (2000) Economics of Sport and Recreation, London, Routledge

Greene, W. H. (2003) Econometric Analysis, New Jersey, Prentice Hall

Verbeek, M. (2000) A Guide to Modern Econometrics, Chichester, Wiley

Williams, J. (2010) Alternative methods of measuring participation in culture and sport: analysis of the Taking Part survey, Department for Culture Media and Sport

Appendix 1 Sport clusters based Williams' (2010) cluster analysis of culture and sport activities

Group 1: Athletics & rugby

- American football, archery, curling, Gaelic sports, gymnastics, lacrosse, orienteering, rugby league, rugby union, track and field athletics, triathlon

Group 2: Dancing & yoga

- Keep fit, aerobics, dance exercise, pilates, trampolining, yoga

Group 3: Outdoor sports

- Angling or fishing, BMX, cyclo-cross, mountain biking, climbing/mountaineering, cycling [for health, recreation, training], cycling [to get to places, i.e. work, shops], hill trekking or backpacking, motor sports, rambling/walking for pleasure, shooting

Group 4: Swimming, cycling & gym

- Health, fitness, gym or conditioning activities, swimming or diving [indoors or outdoors], cycling for pleasure or fitness

Group 5: Racquet sports & running

- Badminton, horse riding, ice skating, jogging, cross-country, road running, squash, table tennis, tennis

Group 6: Bowling

- Bowls (lawn) [outdoor], Bowls [indoors]

Group 7: Cricket, football, pub sports & ten-pin bowling

- Cricket, croquet, darts, football [indoors or outdoors], golf, pitch and putt, putting, skittles, snooker, pool, billiards (excluding bar billiards), tenpin bowling

Group 8: Boxing, martial arts & weightlifting

- Boxing, judo, karate, other Martial Arts (include self defence), taekwondo, weight training (include body building), weightlifting

Group 9: Minor team sport

- Baseball/softball, basketball, hockey, netball, rounders, volleyball

Group 10: Water sports

- Any other water sport, canoeing, rowing, skiing, waterskiing, windsurfing or boardsailing, yachting or dingy sailing

Appendix 2 Probit regression model for sports participation by heterogeneous groups

	Athletics & rugby		Dancing & yoga		Outdoor sports		Swimming, cycling & gym		Racquet sports & running	
COVARIATE	ME	T statistic	ME	T statistic	ME	T statistic	ME	T statistic	ME	T statistic
Female	-0.007***	3.70	0.051***	9.90	-0.137***	11.90	0.105***	7.19	-0.027***	2.80
Age of respondent	-0.000***	4.86	0.000	0.08	-0.002***	4.22	-0.003***	5.04	-0.004***	7.96
Education: reference group is GCSE/O Level grade A* -C(< 5 A*-C)										
GCSE/O Level grade A* - C< 5 A* - C and L1 equivalents	-0.004	1.14	-0.007	0.80	-0.012	0.56	0.017	0.62	0.007	0.37
A levels, vocational level 3 & equivalents	-0.001	0.36	0.010	1.50	0.025	1.58	0.041**	2.05	0.021	1.49
Other Higher Education below degree level	0.006*	1.87	0.014*	1.81	0.029	1.61	0.024	1.06	0.030*	1.87
Higher Education & professional/vocational equivalents	0.001	0.47	0.014**	2.13	0.062***	3.90	0.050**	2.51	0.056***	4.02
Other qualifications: level unknown	-0.001	0.15	-0.018	1.60	-0.001	0.04	-0.007	0.18	-0.015	0.52
Trade Apprenticeships	-0.006*	1.90	0.001	0.04	-0.024	1.05	-0.014	0.44	-0.026	1.18
Natural logarithm of income	0.002*	1.93	0.001	0.39	0.011	1.63	0.023***	2.73	0.013**	2.26
Socio-economic group: reference group is long term unemployed										
Large employers and higher managerial	-0.005	1.51	0.030	1.55	-0.057*	1.94	0.016	0.34	0.023	0.76
Higher professional	0.002	0.41	0.018	1.11	-0.066**	2.38	0.014	0.32	-0.015	0.58
Lower managerial and professional	-0.001	0.16	0.011	0.87	-0.079***	3.01	0.015	0.37	0.001	0.04
Intermediate occupations	-0.002	0.57	0.018	1.26	-0.074***	2.88	-0.018	0.44	-0.003	0.12
Small employers and own account workers	-0.003	0.68	0.002	0.12	-0.040	1.41	-0.031	0.73	0.002	0.08
Lower supervisory and technical	0.006	1.09	-0.014	1.16	-0.016	0.56	-0.033	0.80	-0.028	1.16
Semi routine	-0.003	0.76	0.002	0.19	-0.050*	1.87	-0.038	0.95	-0.046**	2.02
Routine	-0.001	0.37	-0.006	0.42	-0.066**	2.43	-0.052	1.23	-0.042*	1.75
Economic activity: reference group is Full time										
Never worked or long term unemployed			0.000	0.00	-0.118**	2.10	-0.135*	1.75	0.044	0.69
Long term sick or disabled	-0.001	0.11	0.014	0.66	0.044	1.02	0.006	0.12	-0.048	1.05
Looking after family/home	-0.003	0.55	0.016	1.46	0.031	1.10	0.067**	1.96	0.023	0.94
Part time	0.008**	2.52	0.011*	1.86	0.012	0.84	0.033*	1.74	0.014	1.13
Retired from paid work	0.004	0.94	-0.011*	1.68	-0.042**	2.26	-0.049*	1.96	-0.013	0.73
Student	0.019**	2.01	-0.013	0.87	-0.051	1.45	0.036	0.67	0.096**	2.53

Temporarily sick or injured					-0.074	0.80	0.106	0.90		
Car use: Yes	0.000	0.05	0.007	1.27	0.029**	1.99	0.062***	3.26	0.043***	3.38
Ethnicity: reference group is White										
Asian	-0.005*	1.69	0.054***	4.18	-0.104***	4.83	-0.056*	1.78	-0.030	1.59
Black	-0.002	0.59	-0.016	1.54	-0.080***	3.40	-0.057*	1.72	-0.067***	3.45
Mixed	-0.002	0.39	0.005	0.26	-0.003	0.06	0.081	1.22	0.080*	1.75
Other ethnicity			0.003	0.15	-0.109**	2.44	0.007	0.11	0.006	0.16
Marital status: reference group is Single										
Married and living with husband	-0.002	0.75	-0.002	0.30	-0.003	0.20	0.004	0.24	-0.025**	2.21
Married and separated from husband	-0.004	0.83	-0.018**	2.00	0.011	0.39	-0.033	0.94	0.030	1.21
In registered same-sex civil partnership living with partner			0.011	0.41	-0.063	1.15	-0.024	0.30	-0.023	0.50
Separated, but still legally in a same-sex civil partnership			0.077	0.95	0.090	0.46	-0.157	0.79	0.101	0.66
Widowed	-0.002	0.41	-0.012	1.33	-0.027	0.97	-0.047	1.35	-0.038	1.44
Divorced	0.010**	2.34	-0.001	0.16	0.020	1.03	0.015	0.62	-0.009	0.54
Number of adults in household	-0.000	0.01	0.000	0.17	-0.011*	1.70	-0.003	0.38	0.001	0.10
How many children under 16 live in this household?	-0.000	0.21	-0.009***	3.66	0.001	0.19	-0.003	0.41	-0.005	1.03
Drinking: reference group is Don't drink										
Drinking: Less often than once a week	0.003	1.07	0.006	0.88	0.012	0.71	0.081***	3.84	0.006	0.39
Drinking: 1-3 days a week	0.002	0.71	0.009	1.45	0.024	1.49	0.088***	4.23	0.016	1.16
Drinking: 4-6 days a week	-0.001	0.30	0.006	0.66	0.035*	1.66	0.086***	3.10	0.048**	2.48
Drinking: Every day	0.002	0.53	0.011	1.14	0.026	1.17	0.049*	1.70	-0.008	0.41
Smoking: Yes	0.001	0.73	-0.012**	2.47	-0.008	0.68	-0.094***	6.14	-0.042***	4.28
General health: reference group is Fair										
General health: Very bad	0.030	1.21			-0.058	1.01	0.017	0.20	-0.058	0.78
General health: Bad	-0.003	0.56	-0.000	0.01	-0.047	1.55	-0.027	0.66	-0.025	0.77
General health: Good	-0.001	0.23	0.006	1.01	0.034**	2.28	0.058***	3.04	0.036***	2.59
General health: Very good	0.005*	1.84	0.010	1.59	0.058***	3.78	0.106***	5.43	0.086***	5.92
Live sport coverage on TV	0.005***	3.01	0.004	1.07	-0.014	1.36	0.042***	3.21	0.028***	3.19
Sports when growing up	0.003	1.12	0.011**	2.32	0.053***	3.97	0.106***	6.19	0.064***	5.73
Month: reference group is January										
February	0.012**	2.02	0.005	0.67	0.000	0.02	0.015	0.62	0.012	0.72
March	0.010	1.59	-0.012	1.61	0.029	1.30	0.048*	1.71	-0.006	0.34
April	0.044***	3.51	-0.003	0.27	0.012	0.40	0.083**	2.13	0.056**	2.13
May					0.265**	2.21	0.043	0.30	-0.004	0.04
August	0.006	1.03	-0.005	0.70	0.092***	4.05	0.075***	2.78	0.011	0.61
September	0.020***	3.13	0.001	0.10	0.108***	5.60	0.075***	3.31	0.022	1.48
October	0.013**	2.04	-0.001	0.16	0.086***	3.95	0.076***	2.95	0.015	0.86
November	0.014**	2.43	-0.003	0.43	0.048***	2.59	0.038*	1.70	0.003	0.17
December	0.026***	3.06	-0.002	0.22	0.019	0.83	-0.008	0.28	-0.018	0.99
Observations	6115		6199		6290		6290		6269	
ll	-456.3		-992.4		-2688		-3726		-2292	

Estimates are marginal effects. T statistics are absolute values. *** p<0.01, ** p<0.05, * p<0.1

	Bowling		Cricket, football, pub sports & ten-pin bowling		Boxing, martial arts & weightlifting		Minor team sport		Water sports	
COVARIATE	ME	T statistic	ME	T statistic	ME	T statistic	ME	T statistic	ME	T statistic
Female	0.001	0.22	-0.179***	15.15	-0.049***	8.28	0.000	0.01	-0.006*	1.87
Age of respondent	-0.000	0.32	-0.006***	10.91	-0.001***	4.58	-0.001***	6.42	-0.000	1.29
Education: reference group is GCSE/O Level grade A* -C(< 5 A*-C)	0.001	0.27	0.002	0.10	0.009	0.74	-0.003	0.68	-0.010	1.62
GCSE/O Level grade A* - C< 5 A* - C and L1 equivalents										
A levels, vocational level 3 & equivalents	0.003	0.76	0.001	0.10	0.017**	2.01	0.003	0.80	0.000	0.09
Other Higher Education below degree level	-0.001	0.32	0.005	0.28	0.017*	1.72	0.001	0.22	-0.005	1.13
Higher Education & professional/vocational equivalents	-0.005	1.25	-0.027*	1.79	0.020**	2.33	0.006*	1.69	0.007	1.53
Other qualifications: level unknown	0.017**	2.20	0.010	0.35	-0.001	0.09	0.005	0.68	-0.008	0.83
Trade Apprenticeships	-0.001	0.23	-0.056***	2.64	-0.020*	1.71	0.004	0.60	0.002	0.21
Natural logarithm of income	0.001	0.86	0.002	0.31	-0.001	0.33	-0.001	0.55	0.004**	2.19
Socio-economic group: reference group is Long term unemployed										
Large employers and higher managerial	0.019	1.04	0.016	0.47	-0.009	0.64	0.006	0.82	0.006	0.56
Higher professional	0.005	0.38	-0.053*	1.79	-0.019	1.61	-0.001	0.11	-0.001	0.09
Lower managerial and professional	0.005	0.43	-0.020	0.71	-0.011	0.94	0.004	0.89	-0.004	0.60
Intermediate occupations	0.003	0.23	-0.030	1.08	-0.009	0.69	-0.002	0.46	-0.007	1.08
Small employers and own account workers	0.008	0.56	-0.040	1.38	-0.017	1.42	-0.003	0.69	-0.005	0.61
Lower supervisory and technical	0.013	0.83	-0.010	0.34	-0.005	0.40	0.002	0.43	-0.004	0.49
Semi routine	0.004	0.35	-0.041	1.49	-0.012	0.95	-0.006*	1.67	-0.003	0.38
Routine	0.031	1.49	-0.047*	1.66	-0.017	1.48	0.000	0.09	-0.011*	1.70
Economic activity: reference group is Full time										
Never worked or long term unemployed			0.108	1.39			0.004	0.23		
Long term sick or disabled	0.011	0.83	-0.060	1.48	-0.023	1.14	0.045**	2.28		
Looking after family/home			-0.031	1.14	0.005	0.34	-0.006	1.22	-0.002	0.20
Part time	-0.002	0.44	0.036**	2.28	-0.006	0.84	0.011***	3.12	0.005	1.15
Retired from paid work	0.029***	4.06	0.041*	1.94	0.007	0.66	0.005	0.79	-0.006	1.25
Student	0.009	0.50	-0.045	1.29	-0.012	0.86	0.018**	2.03	0.026*	1.66
Temporarily sick or injured										
Car use: Yes	0.007**	2.14	0.036**	2.45	0.005	0.72	0.003	0.98	0.011**	2.42
Ethnicity: reference group is White										
Asian	-0.003	0.44	-0.023	0.95	-0.001	0.08	0.002	0.47	-0.013**	2.10
Black			-0.078***	3.24	-0.003	0.23	0.013**	2.39	-0.011	1.57
Mixed			-0.045	0.97	0.052*	1.88	-0.004	0.54	-0.003	0.25
Other ethnicity	0.035*	1.89	-0.041	0.83	-0.019	0.82	0.006	0.57	0.008	0.52
Marital status: reference group is Single										
Married and living with husband	0.001	0.28	-0.028**	2.09	-0.019***	2.91	-0.002	0.64	-0.001	0.17

Married and separated from husband	-0.006	0.79	-0.001	0.05	-0.007	0.50	-0.001	0.27	0.007	0.82
In registered same-sex civil partnership living with partner			-0.021	0.37	0.007	0.28	0.050**	2.56	0.003	0.18
Separated, but still legally in a same-sex civil partnership	0.170*	1.88	0.081	0.46						
Widowed	0.004	0.70	-0.023	0.76	-0.014	0.99	-0.003	0.33	-0.005	0.52
Divorced	-0.001	0.18	-0.002	0.10	-0.002	0.16	-0.002	0.53	0.005	0.84
Number of adults in household	-0.001	0.55	0.011*	1.73	0.004	1.29	-0.001	0.57	0.002	0.94
How many children under 16 live in this household?	-0.001	0.54	0.007	1.22	-0.003	0.90	0.001	0.74	-0.002	1.20
Drinking: reference group is Don't drink										
Drinking: Less often than once a week	-0.003	0.91	0.002	0.10	0.012	1.33	-0.005**	2.10	0.013*	1.90
Drinking: 1-3 days a week	-0.001	0.25	0.059***	3.46	0.012	1.40	-0.004	1.51	0.012*	1.88
Drinking: 4-6 days a week	-0.003	0.67	0.052**	2.25	0.020*	1.66	-0.003	0.89	0.022**	2.35
Drinking: Every day	-0.006*	1.68	0.050**	2.13	0.003	0.27	-0.002	0.45	0.014	1.57
Smoking: Yes	-0.007**	2.27	0.036***	2.97	-0.010*	1.83	-0.000	0.13	-0.002	0.73
General health: reference group is Fair										
General health: Very bad	0.000	0.03	-0.058	0.91	0.046	0.98				
General health: Bad	-0.005	1.04	-0.076***	2.58	0.005	0.23	-0.007	1.31	-0.004	0.43
General health: Good	0.001	0.25	-0.005	0.34	0.011	1.39	0.002	0.62	-0.004	1.04
General health: Very good	-0.001	0.31	0.003	0.23	0.028***	3.33	0.005	1.53	0.004	1.02
Live sport coverage on TV	0.007***	2.69	0.099***	9.56	0.005	1.00	0.008***	3.97	-0.001	0.52
Sports when growing up	0.000	0.12	0.088***	6.43	0.016**	2.31	0.007***	2.67	0.006	1.64
Month: reference group is January										
February	0.002	0.32	0.012	0.65	0.012	1.20	-0.001	0.39	0.002	0.33
March	0.001	0.19	0.025	1.12	0.020*	1.77	-0.005	1.38	0.001	0.11
April	0.009	0.87	0.024	0.81	0.022	1.48	0.002	0.39	0.004	0.46
May			-0.084	0.89	0.021	0.36			0.048	1.19
August	0.010	1.50	0.037*	1.77	0.003	0.31	-0.002	0.46	0.001	0.22
September	0.010*	1.87	0.031*	1.75	-0.000	0.03	0.003	0.80	0.002	0.40
October	0.021***	2.92	-0.011	0.54	0.005	0.50	0.002	0.41	0.007	1.13
November	-0.000	0.10	0.011	0.64	-0.006	0.76	-0.002	0.70	-0.004	0.99
December	0.005	0.73	0.004	0.18	-0.002	0.21	0.006	1.35	-0.008	1.45
Observations	5664		6269		6214		6194		6048	
ll	-440.9		-2579		-1184		-546.3		-621.4	

Estimates are marginal effects. T statistics are absolute values. *** p<0.01, ** p<0.05, * p<0.1

Appendix 3 Tobit regression model for sports engagement in the last 4 weeks prior to interview, by sport cluster

	Athletics & rugby		Dancing & yoga		Outdoor sports		Swimming, cycling & gym		Racquet sports & running	
COVARIATE	ME	T statistic	ME	T statistic	ME	T statistic	ME	T statistic	ME	T statistic
Female	-5.360**	2.10	-3.419	1.20	1.280	1.38	-2.299***	4.15	-2.649***	3.04
Age of respondent	-0.066	0.44	-0.202**	2.18	-0.217***	4.93	-0.132***	5.00	-0.104**	2.40
Education: reference group is GCSE/O Level grade A* -C(< 5 A*-C)										
GCSE/O Level grade A* - C< 5 A* - C and L1 equivalents	3.260	0.53	-2.545	0.50	-2.270	1.20	-0.755	0.68	-0.691	0.34
A levels, vocational level 3 & equivalents	-4.625	1.38	1.933	0.61	1.547	1.24	-0.609	0.79	-2.186*	1.67
Other Higher Education below degree level	1.290	0.34	-3.327	1.00	2.624*	1.85	-0.162	0.19	-2.386	1.64
Higher Education & professional/vocational equivalents	-0.721	0.21	-0.066	0.02	2.482**	2.08	0.810	1.10	-0.244	0.20
Other qualifications: level unknown	-3.227	0.43	11.895	1.02	-2.750	1.03	1.521	0.93	-0.857	0.26
Trade Apprenticeships	12.661	1.22	9.509	1.26	-1.369	0.68	-2.100	1.48	-3.071	1.16
Natural logarithm of income	-2.669*	1.76	0.357	0.30	1.198**	2.27	-0.336	1.02	-0.376	0.71
Economic activity: reference group is Full time										
Long term sick or disabled	-45.662	.	-0.349	0.04	-1.401	0.38	-3.266	1.29	-3.731	0.49
Looking after family/home	66.329	.	-2.070	0.51	4.581**	1.99	-0.869	0.69	1.231	0.55
Part time	0.248	0.07	0.553	0.22	0.457	0.38	-0.430	0.62	-0.238	0.21
Retired from paid work	-1.772	0.20	4.022	1.08	4.325**	2.44	1.037	0.95	2.725	1.31
Student	-10.483**	2.29	-5.522	0.63	-4.243*	1.65	-3.425**	2.28	-2.384	1.23
Temporarily sick or injured					-1.369	0.11	-4.639	1.03		
Car use: Yes	4.472	1.30	1.782	0.50	-0.955	0.74	2.567***	3.11	2.854**	2.06
Ethnicity: reference group is White										
Asian	-8.623	1.03	-7.064**	2.01	-1.933	0.70	-0.606	0.47	-3.603*	1.79
Black	11.919**	2.00	-2.610	0.27	0.071	0.03	0.544	0.37	-1.376	0.48
Mixed	16.010	1.27	-2.601	0.26	3.362	0.86	-0.929	0.41	-5.056*	1.66
Other ethnicity			-10.929	1.32	1.879	0.30	-1.246	0.50	3.592	0.95
Marital status: reference group is Single										
Married and living with husband/wife	-8.024**	2.49	2.383	1.00	-0.097	0.09	-0.094	0.14	-0.428	0.39

Married and separated from husband/wife	26.594***	2.09	6.459	0.77	-0.796	0.36	-2.330	1.57	-4.166**	1.99
Separated, but still legally in a same-sex civil partnership			7.611	0.51	-87.455	.	-76.810	.	-5.206	0.43
Surviving same-sex civil partner,partner since died							-6.272	0.58		
Widowed			-1.808	0.32	-3.747	1.32	-3.220**	1.99	-5.061	1.40
Divorced	-4.552	1.03	4.003	1.14	0.802	0.53	0.641	0.68	2.654	1.56
In registered same-sex civil partnership living with partner			6.548	0.66	5.911	1.13	1.153	0.37	-1.091	0.24
Number of adults in household	-2.041*	1.77	0.427	0.37	0.256	0.51	-0.248	0.77	0.192	0.40
How many children under 16 live in this household?	0.409	0.30	-1.949	1.58	-1.154**	2.44	-1.201***	4.13	-1.001**	2.19
Drinking: reference group is Don't drink										
Drinking: Less often than once a week	6.010	1.10	-2.372	0.77	-2.468*	1.74	-0.755	0.89	1.722	1.22
Drinking: 1-3 days a week	5.587	1.06	-2.794	0.93	-2.840**	2.05	0.098	0.12	-0.462	0.33
Drinking: 4-6 days a week	5.863	0.89	-0.902	0.23	-2.457	1.48	0.735	0.70	0.140	0.09
Drinking: Every day	22.556***	3.06	-3.021	0.74	-5.009***	2.82	-1.733	1.48	0.715	0.35
Smoking: Yes	-3.997	1.42	-2.417	0.81	-1.881*	1.96	-0.865	1.32	-0.517	0.48
General health: reference group is Fair										
General health: Very bad					5.365	0.82	5.237	1.34	-5.502	0.49
General health: Bad	74.041	.	-2.905	0.37	5.439*	1.69	4.337**	2.26	-1.095	0.23
General health: Good	11.033**	2.38	-0.392	0.13	2.848**	2.24	1.690**	2.12	2.330	1.61
General health: Very good	14.674***	3.31	5.347*	1.81	6.407***	5.04	4.248***	5.33	6.758***	4.75
Live sport coverage on TV	2.181	0.84	0.496	0.26	2.460***	3.01	1.310**	2.56	-0.253	0.30
Sports when growing up	14.847***	2.76	-1.880	0.68	3.871***	2.96	1.795**	2.35	3.899***	2.70
Month: reference group is January										
February	-5.813	0.85	-0.110	0.03	2.554	1.54	-0.384	0.40	0.816	0.56
March	-0.750	0.11	-0.800	0.18	2.399	1.34	1.465	1.36	1.327	0.78
April	-2.943	0.39	0.098	0.02	4.461*	1.79	-1.139	0.81	-0.147	0.07
May					2.098	0.38	2.716	0.51	0.236	0.03
August	0.107	0.01	-0.353	0.09	1.434	0.88	0.675	0.67	1.075	0.66
September	1.991	0.30	2.605	0.87	2.245	1.60	-0.154	0.18	2.241	1.64
October	2.563	0.36	0.337	0.09	-0.771	0.49	-0.331	0.34	2.643*	1.68
November	0.483	0.07	2.211	0.70	0.642	0.44	-0.884	1.00	1.180	0.85
December	0.654	0.09	-4.683	1.21	2.009	1.07	-0.572	0.50	1.685	0.93
Constant	17.758	1.11	18.394	1.29	-0.372	0.06	16.746***	4.46	13.527**	2.22
Observations	112		281		1139		2118		932	
ll	-309.9		-848.2	.	-3504	.	-6838	.	-2971	.

Estimates are marginal effects. T statistics are absolute values. *** p<0.01, ** p<0.05, * p<0.1

	Bowling		Cricket, football, pub sports & ten pin bowling		Boxing, martial arts & weightlifting		Minor Team sport		Water sport	
COVARIATE	ME	T statistic	ME	T statistic	ME	T statistic	ME	T statistic	ME	T statistic
Female	0.071	0.03	-2.690***	2.88	-1.451	0.83	-1.845	0.94	-1.411	0.72
Age of respondent	-0.149	1.11	-0.145***	3.60	-0.127	1.52	-0.395***	2.88	-0.263**	2.52

Education: reference group is GCSE/O Level grade A* -C(< 5 A*-C)										
GCSE/O Level grade A* - C< 5 A* - C and L1 equivalents	5.927	1.26	2.167	1.39	3.392	1.01	-3.210	0.60	-9.056	1.13
A levels, vocational level 3 & equivalents	-3.338	0.98	1.493	1.37	2.294	1.03	-2.844	0.90	0.947	0.33
Other Higher Education below degree level	2.132	0.56	1.744	1.36	4.109	1.57	-0.672	0.15	4.834	1.30
Higher Education & professional/vocational equivalents	5.578	1.66	2.604**	2.37	3.544	1.62	0.093	0.03	2.751	1.05
Other qualifications: level unknown	10.949***	2.68	3.695*	1.68	8.645	1.55	1.736	0.26	60.667	.
Trade Apprenticeships	4.752	1.11	0.324	0.18	-3.250	0.69	4.847	0.84	3.383	0.65
Natural logarithm of income	0.040	0.03	0.261	0.52	0.840	0.88	-0.126	0.10	-0.331	0.29
Economic activity: reference group is Full time										
Long term sick or disabled	-48.953	.	-4.882	1.14	1.818	0.20	-9.140	1.26		
Looking after family/home			1.823	0.69	2.969	0.60	-9.042	1.13	6.097	0.88
Part time	0.377	0.08	0.597	0.52	2.775	1.19	-1.952	0.66	-1.500	0.54
Retired from paid work	1.146	0.29	2.436	1.42	8.402**	2.31	-1.951	0.28	3.235	0.78
Student	-8.722	0.81	0.752	0.35	0.980	0.28	4.250	1.13	-2.731	0.67
Temporarily sick or injured										
Car use: Yes	8.253	1.43	0.445	0.37	2.236	1.03	0.821	0.24	-9.046*	1.71
Ethnicity: reference group is White										
Asian	-0.332	0.04	-3.649*	1.95	-4.972	1.61	-6.307	1.48	-2.744	0.29
Black			0.497	0.20	-3.856	1.10	-8.901***	2.94	2.146	0.20
Mixed			2.525	0.63	1.162	0.24	56.794	.	9.530	0.93
Other ethnicity	15.268*	1.74	-3.188	0.66	-15.168*	1.79	-4.746	0.66	-13.794	1.49
Marital status: reference group is Single										
Married and living with husband/wife	-1.633	0.41	0.300	0.28	-1.225	0.63	5.341	1.58	0.534	0.21
Married and separated from husband/wife	0.263	0.02	1.858	0.83	-0.719	0.14	-1.019	0.18	-10.180**	2.10
Separated, but still legally in a same-sex civil partnership	-55.954	.	-4.734	0.38						
Surviving same-sex civil partner,partner since died										
Widowed	-2.938	0.66	-5.934**	1.96	-10.762*	1.70	2.942	0.29	-7.898	1.01
Divorced	-6.261	1.29	0.377	0.24	-0.741	0.25	6.372	1.15	-0.687	0.21
In registered same-sex civil partnership living with partner			-2.597	0.62	4.503	0.73	-1.594	0.26	-6.786	0.83
Number of adults in household	-1.574	0.69	0.085	0.18	0.803	1.03	-0.097	0.09	1.086	1.02
How many children under 16 live in this household?	-1.517	0.85	-0.261	0.62	-1.259	1.42	-0.648	0.58	-0.225	0.18
Drinking: reference group is Don't drink										
Drinking: Less often than once a week	-4.513	1.39	2.227	1.57	1.430	0.56	2.681	0.85	-1.057	0.27
Drinking: 1-3 days a week	-2.971	0.96	1.569	1.16	0.786	0.31	3.443	1.07	-3.825	1.00
Drinking: 4-6 days a week	-3.853	0.89	3.105*	1.86	-3.103	1.01	3.827	1.00	-2.979	0.70
Drinking: Every day	-4.813	1.09	-1.293	0.73	0.048	0.01	-1.720	0.37	-1.860	0.40

Smoking: Yes	1.070	0.26	-2.176**	2.58	2.315	1.26	-2.182	0.83	-5.203**	2.17
General health: reference group is Fair										
General health: Very bad	-13.517	.	8.451	1.23	89.373	.				
General health: Bad	6.605	0.90	1.549	0.45	2.964	0.46	-72.104	.	13.271*	1.87
General health: Good	-1.393	0.45	3.608***	3.29	5.975**	2.28	-8.111*	1.95	1.046	0.38
General health: Very good	3.414	1.12	7.079***	6.38	13.322***	5.01	-4.771	1.14	2.747	0.98
Live sport coverage on TV	2.235	0.86	0.922	1.07	0.625	0.42	-0.595	0.28	6.144***	3.30
Sports when growing up	4.844	1.39	3.782**	2.46	4.863*	1.92	6.398	1.44	-6.311**	2.01
Month: reference group is January										
February	-1.455	0.32	1.991	1.43	2.233	0.87	-5.607	1.50	4.707	1.46
March	-4.655	0.77	5.661***	3.61	2.216	0.82	10.593**	2.15	8.401**	2.33
April	-4.001	0.59	1.313	0.66	0.892	0.27	2.091	0.48	-5.329	1.21
May			-8.105	0.86	-13.205	1.12			2.420	0.26
August	-3.755	0.80	3.269**	2.21	1.913	0.67	-8.269*	1.94	-3.331	0.94
September	1.032	0.26	4.173***	3.24	3.998	1.58	-5.154	1.63	-1.758	0.61
October	0.876	0.23	2.381	1.55	2.090	0.76	3.489	0.92	-4.315	1.41
November	0.310	0.07	1.685	1.27	2.326	0.88	-1.560	0.43	5.273*	1.69
December	-4.998	1.03	1.811	1.10	0.059	0.02	-3.021	0.84	-4.443	0.91
Constant	8.634	0.44	-2.065	0.37	-7.525	0.71	31.778**	2.02	36.546***	2.85
Observations	102		1259		350		139		149	
ll	-304.6	.	-3804	.	-1026	.	-408.8	.	-448.4	.