

# Heterogeneous sports participation and labour market outcomes in England

Michael Lechner and Paul Downward

Michael Lechner, Professor of Econometrics, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbuelstrasse 14, CH-9000 St. Gallen, Switzerland, [Michael.Lechner@unisg.ch](mailto:Michael.Lechner@unisg.ch), [www.michael-lechner.eu](http://www.michael-lechner.eu) .

Paul Downward, Professor of Economics, School of Sport Exercise and Health Sciences, Loughborough University, LE11 3TU, UK, [Paul.Downward@lboro.ac.uk](mailto:Paul.Downward@lboro.ac.uk), <http://www.lboro.ac.uk/departments/ssehs/staff/professor-paul-downward-.html>.

# Heterogeneous sports participation and labour market outcomes in England

**Abstract:** Based on a unique composite dataset measuring heterogeneous sports participation, labour market outcomes and local facilities provision, this paper examines for the first time the association between different types of sports participation and employment and earnings in England. Clear associations between labour market outcomes and sports participation are established through matching estimation whilst controlling for some important confounding factors. The results, which are supplemented and supported by a formal sensitivity analysis, suggest a link between different types of sports participation to initial access to employment and then higher income opportunities with ageing. However, these vary between the genders and across sports. Specifically, the results suggest that team sports contribute most to employability, but that this varies by age across genders and that outdoor activities contribute most towards higher incomes.

**Keywords:** Sports Participation, Human Capital, Labour Market, Matching Estimation, Sensitivity Analysis

**JEL classification:** Z2, I12, I18, J24, L83, C21.

## 1 Introduction

The role that sport plays in society is of a resurgent policy concern. Historically, sports policy aimed to raise physical fitness levels in the population to meet the needs of the military (Houlihan, 1997; Green, 2004; Green and Houlihan, 2005). In the period after the Second World War, sport became a branch of social welfare (McIntosh, 1980; Coalter, 2007; Downward *et al* 2009), ultimately becoming manifest in Europe in the European Sports Charters of 1991 and 2001. Symbiotically, interest in the policy promotion of hosting major sports events, like the Olympics, and promoting elite sports success at them has occurred across a diverse range of governments (DCMS/Strategy Unit, 2002; Carter, 2005; DCMS, 2010). Most recent sport policy emphasises a need to counter falling levels of physical activity to improve health and reduce the chance of succumbing to diseases (e.g. Department of Health, 2004; WHO, 2002, 2007; DCMS, 2015).<sup>1</sup> Surprisingly, what has received much less attention in policy discussion is the general role that sport participation can play in improving an individual's labour market outcomes.

This paper contributes to filling a gap in the literature by answering the following related research questions. How does participation in different types of sport affect the earnings and employment of the working-age population? Are these effects for the different types of sports heterogeneous with respect to genders and age? These currently unanswered questions are important since it is well known that different types of sporting activities are expected to have different organisational features and aspects of practice, which may affect their impact on the individual (Downward *et al* 2009). It is also well known that the type and intensity of sports related activities vary considerably over age and across genders (Van Tuyckom, *et al* 2010; Breuer and Wicker, 2009; Breuer and Wicker, 2008; Kay, 2003). To address these questions a

---

<sup>1</sup> See also the 2007 White Paper on Sports by the European Commission ([http://ec.europa.eu/sport/white-paper/white-paper\\_en.htm](http://ec.europa.eu/sport/white-paper/white-paper_en.htm); accessed 12/11/2014).

matching analysis is undertaken based on a unique composite dataset synthesised from different UK sources in a research design that gets as close as currently possible to a causal analysis.

The results of the empirical analysis show that there is considerable heterogeneity with respect to the type of participants in the different sports. Furthermore, the results indicate large positive associations of sports participation with earnings, which are largest for fitness and outdoor sports. There is a negative relation with unemployment, particularly for males, which goes together with higher employment rates for younger males and higher retirement rates for older males. Comparing the different sports against each other shows that participation in team sports is more associated with increased employability (but that this varies by age and gender) and that fitness and outdoor activities have higher associations with income. The results of a sensitivity analysis provided in an online appendix do not challenge these findings in general.

The paper proceeds as follows. In Section 2, literature examining the effects of sports participation on labour market outcomes is briefly reviewed. In Section 3 details of the datasets, which are synthesised, are provided. Section 4 outlines the measures of sports participation that are employed in the analysis and describes their allocation over individuals. Section 5 presents an evaluation of alternative empirical strategies before outlining the preferred matching estimation. Section 6 outlines the results from the matching estimation. Conclusions follow in Section 7.<sup>2</sup>

## **2 Literature review**

### **2.1 Theoretical issues**

The main theoretical premises by which sports participation may affect labour market outcomes are through the development and signalling of human capital through cognitive skills

---

<sup>2</sup> Appendix A provides further descriptive statistics. Appendix B also provides the details and results of the extensive robustness analysis, which includes a variation in the sample definitions, the already mentioned formal sensitivity analysis, as well as an alternative instrumental variable identification strategy. Both appendices are provided on the internet.

(Mincer, 1958; Becker, 1964 and Spence, 1973); and non-cognitive skills such as social adaptability and informal sources of learning through work experience and learning-by-doing. Participating in sport might, therefore, entail investment in, for example, a teamwork ethic which is then rewarded in the labour market directly (Rees and Sabia, 2010), or perhaps indirectly through access to networks of employment (Jackson, 2010). Lechner (2009) and Rooth (2011) also argue that externalities could occur through the impact of sports participation on health and fitness, which also signals greater future productivity.<sup>3</sup>

## **2.2 Labour market outcomes**

The impact of sport on labour market outcomes is examined empirically in two strands of literature. The first strand explores the links between single measures of young people's sports participation, and typically higher educational performance, attainment and earnings. Early studies such as Long and Caudill (1991), and Ewing (1995, 1998, and 2007) explicitly attempt to identify causal effects. Whilst Maloney and McCormick (1993) argue that participation in college athletics reduced scholarly success, they nonetheless recognise that the results may be due to sample selection effects, given the lower overall standardised test scores achieved by athlete entrants to high school.

Later studies have broadly tried to distinguish between selection and causal effects using a variety of econometric strategies. Barron *et al* (2000) make use of an instrumental variable approach to identify positive impacts on the wages and educational attainment of student athletes, though they interpret the impacts as stemming from signalling the greater ability of ath-

---

<sup>3</sup> In this way, Rooth (2009) finds that physical attractiveness, which in part is an outcome of physical activity, might improve the chances of employability, such that females might be judged more harshly when connected with obesity.

letes, and their having less preference for leisure (indicating industriousness), rather than impacts due to human capital.<sup>4</sup> Their theoretical model, which is used to interpret the results, however, explicitly rules out an impact of sport on human capital through productivity enhancement, and, instead, focusses on the time-allocation consequences of sport as crowding-out education. The instruments employed in the analysis such as school size, and other school characteristics, as indicators of supply opportunities, as well as the income of parents and the health, height and weight of the student could be affected by sports participation however.

In contrast, Eide and Ronan (2001) show that lower attainment was achieved for white male sport participants, but white females and black males participating in sport experienced increases in academic success. They also used the height of students as an instrumental variable. Lipscomb (2007) also finds general increases in educational attainment for those participating in extra-curricular school-sponsored sports. A fixed effects modelling strategy and information on the joining and quitting of clubs by individuals are used to try to control for selection effects and to identify causal effects respectively. In the latter case, it is clear that parental choice could confound the relationships between participation and club activity. Pfeifer and Cornelissen (2010) also argue that outside-school sport has a strong effect on a variety of levels of educational achievement. School characteristics and city size are used as instruments to measure the supply of sporting opportunities, whilst student's height is also used to measure student propensity for sport. Rees and Sabia (2010) also use the height of a student as an instrumental variable, but argue that sports participation does not have an effect on academic performance for indicators such as grade-point averages, paying attention in class and college aspirations. However, Stevenson (2010)<sup>5</sup> makes use of changes in US law formalising the need for equality in male and female student sports provision, as a natural experimental shock, interacted with

---

<sup>4</sup> Stevenson (2010) also discusses this possibility

<sup>5</sup> The legal change involved Title IX of the Educational Amendments to the 1964 Civil Rights Act. This banned gender discrimination in federally funded educational institutions.

pre-legislative male participation in sport as an instrument, in an analysis of post-legislation female education and earnings. It is shown that increases in female sports participation raise female subsequent college attendance and labour force participation and wages. Finally, Felfe *et al* (2016) make use of both instrumental variables measuring distance from sports facilities, and lags to control for reverse causality, in an analysis that suggests that sports club participation enhances measures of children's cognitive and non-cognitive development. On balance, thus, the literature is suggestive of positive effects of sports participation on educational achievement and earnings, but earlier studies rely on questionable instruments for identification.

The second strand of literature focusses on the potential for sport to have labour market outcomes for the wider population. A comprehensive analysis is undertaken by Lechner (2009) who uses a matching approach on panel data that has been restructured to group individuals that previously had the same sports participation behaviour in the period before analysis as part of controlling for confounding effects. It is found that sports activities have positive long-term impacts on earnings and wages (as well as on health and subjective well-being). Subsequently, Cabane (2011) parametrically analyses panel data to conclude that sports participation in the previous year has a positive association with greater job autonomy and higher wages in the next year. In this way, lags are used to try to control for reverse causality and for confounding effects, with the aspiration that signalling effects are isolated from health, human capital and networking effects. Rooth (2011) presents results based on a randomised field experiment that job candidates demonstrating experience of sports are more likely to be called for interview than candidates who are similar in other respects but have not participated in sport. Rooth (2011) is also the only study to investigate the impact of different sports. It is argued that soccer and golf have the greatest impacts with the implication that this is because of their signalling the social rather than the health spillovers from sport. In a separate analysis of secondary data associated with

military enlistment, Rooth (2011) also presents (non-experimental) evidence that physical fitness is associated positively with earnings, controlling for cognitive and non-cognitive skills for siblings, thereby controlling for particular unobservable family effects. Hyytinen and Lahtonen (2013) extend this work drawing upon three waves of cohort data for both identical and non-identical twins that records their physical activity, which is then matched at the individual level to longitudinally collected data that records labour market behaviour as well as provides socio-economic controls. It is established that physical activity has a long-run effect on taxable annual income of approximately 14-17 percent.<sup>6</sup> Finally, Lechner and Sari (2014) use a matching analysis to show that changes from moderate to higher intensity activity increases earnings between 10-20% in Canada over the long run.

Overall, the implication of this discussion is that this strand of literature more strongly reveals causal effects running from sports participation to labour market outcomes, with some debate about the mechanisms. The current paper seeks to add to this literature by investigating the labour market impacts from sport participation in the UK for the first time, but also, following Rooth (2011), examining heterogeneous sports, whilst accounting as much as possible confounding effects.

### **3 Data**

#### **3.1 Data sources**

The current research uses data that are synthesised from three major surveys. The main source of data is the ongoing Active People Survey (APS) commissioned by Sport England and

---

<sup>6</sup> The robustness of the analysis is checked by identifying if the 1990 data for income measurements is a source of bias because it reflects the employment status of the physically active at the time. Earlier income before physical activity is used to check for reverse causality. Differences in the health status of twins are also used to check for any subsequent specific differences in health to those that are measured at the time that physical activity is measured.



initially conducted by Ipsos MORI, but subsequently TNS BRMB.<sup>7</sup> It began with the first wave with a sample size of 363,724 between mid-October 2005 and mid-October 2006. Subsequent waves have then taken place annually and because of cost the initial sample size was reduced from approximately 1000 to approximately 500 respondents per local authority. APS 2 was undertaken between mid-October 2007 and mid-October 2008 and has a sample size of 191,325. APS 3 took place between mid-October 2008 and mid-October 2009 and has a sample size of 193,947 and APS 4 was conducted between mid-October 2009 and mid-October 2010 and has a sample size of 188,354. Currently, Wave 9 of the Survey has been published covering the periods from mid-October 2014 to mid-October 2015, with wave 10 due to be published in June 2016. The survey is conducted by telephone on the national sample with households identified by random digit dialling. Household respondents are then selected according to the next birthday rule.

In the current research, APS 4 provides the core individual data for sports participation, labour market outcomes and some individual characteristics of participants. This is because APS 4 is matched to data on the number of sports facilities at the local authority level made available from the Active Places Survey (APLS), which was made available to the researchers for 2008, and which adds exogenous information to the dataset. The APLS is a 'live' industry management tool with no public access to the raw data. There is no historic evolution of the number of facilities recorded in the data as it is continually updated. Consequently, the data is viewed as relevant for 2008, coincident with APS3. APS 3 is then also used to contribute variables that might control for other confounding impacts of participation on labour market outcomes by providing lagged information on aspects of sports participation aggregated to local authority level.

---

<sup>7</sup> This survey is due to be replaced by the Active Lives Survey because of the most recent UK government policy (DCMS, 2015).

The other major source of data was the Annual Population Survey (APopS). First conducted in 2004, the APopS combines results from the Labour Force Survey (LFS) with enhanced data on social and socio-economic variables. The survey is based on approximately 55,000 households generating approximately 360,000 respondents per dataset, and covering a target sample of at least 510 economically active respondents for each local authority. Local authority data can be accessed through special license and in the current research the APopS from the year corresponding to APS 3 is also used to provide variables for a variety of household composition and health and sickness variables which, when also aggregated to a local authority level, control for potential influences on sports participation that could confound outcomes. Finally, data on local authority populations was obtained from the Office for National Statistics.<sup>8</sup>

### **3.2 Sample selection**

As the APS measures sports participation in England for adults aged 16 years or older,<sup>9</sup> and the APLS also covers England, whilst the APopS covers UK household members of all ages, the focus is upon respondents living in England. This generated a sample of 169,460 observations (age 16-80). Some observations were dropped for the local authorities where it was not possible to get reliable population data and where respondents were classified as either having a disability or severe long-term illness. The potential lack of employability of such individuals is not the main focus of the research. For the same reasons the focus is on males and females in the age categories of 26 to 45 years of age and 46 to 64 years of age. This produced a working sample size of 79,561 observations.

---

<sup>8</sup> <http://www.ons.gov.uk/ons/datasets-and-tables/index.html>. (accessed 12/11/14) APopS data are generally available at the level of Government Office Region but can be accessed at local authority level by special license, which is the case here.

<sup>9</sup> From July 2012, which covers the end of APS 6 and onwards, the sample covers respondents aged 14 years.

## **4 Sports participation**

### **4.1 Aggregation of different sports**

The questions investigating sport participation in the APS4 captures data on 415 activities. For each activity, questions are asked about whether or not the respondent participated in the activity in the last four weeks; on how many days in the last four weeks the respondent participated in the activity; the usual length of time in minutes in which the activity is undertaken; if participating in the activity raised the respondent's breathing rate; and, if participating in the activity made the respondent out of breath or sweat. A positive response to the penultimate question identifies valid 'moderate' activity, and a positive response to the last question valid 'vigorous' activity (Ipsos Mori, 2007).

These questions along with the question probing participation or not, required a binary response. Scoring each of these variables as a '1' and '0' for 'yes' and 'no' respectively means that the product of responses to all of these questions identifies the minutes in the last four weeks in which moderate or vigorous sport was undertaken or not, or sport which does not reach either of these thresholds. In the current research, the focus is upon participation that achieves at least moderate intensity, assuming that for sport to have an impact on labour market outcomes though, for example, human or health capital acquisition requires at least some degree of organisation and intensity.

Sports were aggregated into five main analytical groups; team sports, keep fit activities, racquet sports, leisure activities and outdoor activities. This aggregation of activities was based on an extension to an empirical classification already used in UK research. For example, Downward and Riordan (2007), drawing on Rodgers (1977, 1978), identified an empirical classification of sports in the UK associated with teams, fitness activities, and leisure activities. These distinctions have also proved to have relevance in assessing the well-being of sports participants in the UK such that Downward and Rasciute (2011) show that team or group sports yielded

additional value to individuals because of their social interactions. Rooth (2011) supports this result.

Consequently, team sports were defined to include versions of football, cricket, rugby, netball, basketball etc.; fitness activities were defined to include individual sport and fitness activities like field and track athletic activities, cycling, martial arts, weight training and lifting, that would typically take place in fitness or leisure centres and clubs; racquet sports such as badminton, tennis, and squash were included as a separate category because, whilst they are offered at fitness centres there are also distinct clubs and leisure outlets that cater for them and they inevitably take place in the context of groups; leisure activities were defined to include versions of swimming, bowling and dancing etc.

The outdoor activities category is the most eclectic. This included elements of outdoor pursuits such as hunting, horse riding and hill walking; as well as motor sports involving cars and bikes, and winter sports such as skiing and bobsleighbing; and water sports such as canoeing and some leisure activities like golfing. These activities were included as a separate group in part to recognise their distinctiveness from the existing categories that have been used in UK research of purely leisure pursuits, fitness activities and team sports. However, the category is also defined to recognise their connection with lifestyle, possible vacation activity, and their typically longer duration, either intrinsically or connected with tourism, as well as their requirement for equipment. Naturally, other classifications are possible, such as provided by Andreff (2006) for France. It is clear that further investigations on the robustness of the chosen categories of sport is an area for future research. However, notwithstanding these arguments, in the current context it is important to recognise that particular, specific activities dominate the behaviour within the groups. Consequently, one can also view the groups as to an extent representing these specific activities.

## 4.2 Descriptive analysis of sports participation

Table 1 shows participation shares for the higher and lower intensity activities in the full sample, as well as the participation rates for higher intensity activities in the respective subsamples. The main activities that empirically underpin the behaviour in the Team, Fitness, Racquet, Outdoor and Leisure sports are also presented.<sup>10</sup> The disaggregation reveals that football, indoor fitness activities, including badminton but also tennis, swimming and golf are the main drivers of behaviour in the team, fitness, racquet, outdoor and leisure groups. In the light of the comments earlier concerning the classification of sport it is comforting to note that the dominance of these activities is in accord with the literature (Downward *et al* 2009).

With the exception of leisure sports, most sporting activity takes place at moderate or vigorous intensity for these age cohorts. Just over 40% of this population performs at least some sports related activities of this degree of intensity. Nonetheless, activity levels are highly age and gender specific as indicated in the literature. Ageing lowers general activity levels. In addition, there is generally higher activity level for males compared to females, but this difference declines with age.

The importance of allowing for an analysis of the heterogeneity of sporting experience is clear. While team sports are important for males and, in particular young males, their role for females is considerably smaller. In contrast, leisure sports are generally more likely to have female participants. Overall, it appears that participation in fitness, racquet, outdoor and leisure activities can carry through more easily into middle age, compared to team sports. Moreover, female participation in fitness activities can come to exceed that of males in middle age.

---

<sup>10</sup> Activities with participation rates below 1% for all subsamples are omitted.

Table 1: Sports activities in the full sample (shares in %)

Subsample	Participation rates in activities of higher intensity					... of low intensity
	All	Men		Women		All
		26-45	46-64	26-45	46-64	
Type of sport						
All	43.7	54.7	36.6	41.2	31.9	13.5
Team	11.2	19.4	5.5	2.7	0.9	1.6
Fitness	24.9	30.6	17.5	29.1	20.5	3.8
Racquet	5.7	7.9	6.5	3.8	3.7	0.7
Outdoor	6.4	8.4	10.2	3.2	3.5	1.0
Leisure	9.5	9.1	6.4	13.0	9.5	7.3
Football	8.5	16.1	4.0	0.9	0.2	0.9
Rugby	1.1	1.5	0.4	0.1	0.0	0.1
Netball	0.7	0.3	0.1	1.2	0.2	0.1
Basketball	0.7	0.5	0.1	0.1	0.1	0.1
Cricket	0.6	1.1	0.7	0.1	0.0	0.5
Hockey	0.5	0.4	0.2	0.2	0.2	0.0
Baseball	0.3	0.3	0.1	0.2	0.1	0.0
Fitness Indoor	16.5	17.7	10.9	20.7	16.7	2.5
Running	8.2	12.9	6.1	9.3	3.2	0.0
Combat	1.5	2.5	0.8	1.0	0.2	0.1
Bodybuilding	1.1	2.5	0.9	0.4	0.2	0.4
Asian sports	1.1	0.4	0.3	2.3	2.1	0.6
Athletics	0.4	0.3	0.2	0.3	0.1	0.1
Badminton	3.9	6.0	4.7	2.3	2.1	0.4
Tennis	2.1	2.4	2.2	1.6	1.7	0.3
Golf	3.7	5.0	8.1	0.4	1.5	0.0
Outdoor act.	1.3	2.4	1.3	0.8	0.5	0.2
Horse riding	0.8	0.1	0.1	1.5	1.1	0.4
Swimming	8.7	8.7	6.0	12.2	8.7	6.1
Gymn., dancing	0.6	0.1	0.1	0.8	0.7	0.2
Bowling	0.4	0.2	0.4	0.1	0.3	1.1

Note: *Intensity high*: Only sports with moderate or rigorous intensity. *Intensity low*: Any participation in this type of sports that is not intensive enough to be classified as *high intensity*. Ice hockey, cycling, rowing, hunting, motor sports, skiing and bob sleighing are omitted from this table, because the participation rates are below 1% for all groups. The category 'all' includes all individuals not older than 80. Sample weights used.

These results accord with expectations from the literature not only with respect to age and gender. They are also in line concerning the facts that (i) organized competitive activity traditionally takes place in male team sports, that (ii) fitness activities can be undertaken for middle aged participants, with perhaps less time available to them and whose physique is less able to cope with competitive sport, and (iii) that casual leisure activity can persist more easily over the life course. This heterogeneity, coupled with the major finding in the literature that sports participation varies with income and other socio-economic circumstances, as well as facility supply and regional characteristics (Downward et al 2012), implies that such factors need to be taken into account in any analysis seeking to uncover causal behaviour.

## **5 Towards the causal effects of sports participation and labour market outcomes: Possible identifying strategies**

### **5.1 General considerations**

The above discussions of the literature and data description suggests that there are many individual and regional characteristics that are related to sports participation. In fact, most of them, if not all, are also likely to influence labour market outcomes as well. Such variables, together with those not observable in the data at hand can clearly confound the relationship between sports participation and labour market outcomes and thus have to be dealt with if we want to move closer to a causal interpretation of any empirical analysis.

As indicated in the section 2.2, the literature in the field of sports economics has either ignored this requirement or, more recently, adopted an instrumental variable approach in which the instruments are potentially weak. Exceptions are perhaps Lechner (2009) and Rooth (2011). The increased attempt for more robust identification does, however, reflect the broader recognition of the need for adequate analysis of cause in applied econometric analysis, particularly through the desire to evaluate policy more appropriately (Morgan and Winship, 2007; Angrist and Pischke, 2008). It remains, that the impact of any form of ‘treatment’ (sports participation in our application) on ‘outcomes’ is subject to these concerns. For example, Blundell and Dias (2002, 2009) as well as Imbens and Wooldridge (2009) outline a variety of strategies within the literature that are relevant for analysing non-experimental data with the aim of finding associations that may be causally interpreted with some confidence. It remains, to consider which approach might be best in our setting in coming at least close to obtaining causal effects of sports participation on labour market outcomes.

The various approaches suggested in the literature can be distinguished in whether they are based on removing confounding based on observable variables only (selection-on-observability)

bles) or whether they attempt to deal with confounding by unobservable variables as well. Starting with ‘selection-on-unobservables’ we will now discuss the appropriateness of the various approaches in our setting.

## **5.2 Selection on unobservables**

### **5.2.1 Instrumental variables**

Imbens and Angrist (1994) show that the key conditions for instrumental variables to identify causal effects are that (i) they affect the treatment (‘power’) and do this in a monotonic fashion (‘monotonicity’), and that (ii) they affect the outcomes only by affecting the treatment (‘exclusion restriction’). In our application we may argue that the local supply of facilities may be such a variable. It seems also reasonable to assume that an increase in supply will not lead to a reduction of sport activities of some individuals, so that the ‘monotonicity’ condition is fulfilled as well. Concerning the exclusion condition, the key concern will be that the local supply of facilities is correlated with other features of the local area. If such features are related to labour market outcomes, like the type of employment available and the local economic situation in general, then this condition might be violated. However, if such local factors are observable, the exclusion restriction remains valid conditional on such covariates. Natural parametric estimators, like 2SLS, LIML etc. have been extensively discussed. Recently, for example, Abadie (2003) and Frölich (2007) discussed semi-parametric extensions that can deal with a large number of covariates without becoming subject to curse-of-dimensionality problems, as completely non-parametric methods would be subject to. In this application it turned out that the power of the supply of local facilities is not strong enough. Therefore, our main results are not based on this strategy, but we outline its implementation and results in the Internet Appendix B.3.



There are two strategies related to IV, namely the control function approach and the regression-discontinuity design (RDD). Concerning the former, for which the most famous example is the heckit estimator (Heckman, 1979), Vytlacil (2002) shows that it is conceptually identical to instrumental variables. The latter, RDD, is a local version of IV (see Van der Klaaw, 2008, and Imbens and Lemieux, 2008, for comprehensive surveys). The main idea is that there is a forcing variable influencing the treatment. This variable is supposed to influence the probability of being assigned to treatment. At least at one point this variable jumps and thus the participation variable jumps. Since this jump is assumed not to influence the treatment directly, it essentially acts as a ‘local’ instrumental variable around the jump point. Unfortunately, in our application, we are not aware of any such jump that is large enough to be exploited empirically.

### **5.2.2 Differences-in-differences estimation**

If the data have a time dimension then a difference-in-difference strategy may be an option as well (see Lechner, 2010). Suppose there are only two periods: The main idea is that there is a group of individuals increasing their sports participation from one period to the next, while a ‘control’ group would exhibit a stable participation level. If the control group is indeed ‘comparable’ then their change in the outcome variables could be used to infer how the outcome variables of the ‘treated’ would have changed if they had not increased their sports participation. To use this method on an individual level in our case would require panel data which is not available. Local authorities could be used as the unit of analysis but with an obvious loss of power due to the immense reduction in the number of observations, there is also no indication of a sufficiently exogenous and large change of local sports participation that could be exploited. A more realistic approach is thus to at least control for the confounders available in the data and to assess the potential for bias. Thus, next we discuss an approach that does exactly that.

### 5.3 Selection on observables: The matching approach

When the confounding factors, i.e. the variables that are jointly related to sports participation and labour market outcomes, are observable in the data, then their influence can be removed by controlling for them. If all these factors are observable, then we obtain causal effects from such a strategy. If the list of observable confounders is incomplete, then at least the confounding influence of such variables is removed. The various parametric, semi-parametric and non-parametric procedures that do this are extensively reviewed in Imbens (2004).

In the next section, an analysis is provided of how labour market outcomes might be influenced by participation in the different types of sports activities controlling for the observable confounding influences in the data.<sup>11</sup> The estimator used for this purpose is a matching estimator. Matching estimators are particularly attractive when analysing the impact of discrete variables of interest, like the sports participation variables in this context, on outcomes. Matching estimators can be thought of as semi-parametric generalisations of linear or non-linear regression estimators. They obtain estimates whilst allowing for the heterogeneity of individual effects and with no need to impose tight functional form restrictions as with parametric models.

For example, in case of a binary variable of interest, like participating in a particular sport or type of sport, matching estimators ensure measurement, for example, of income differences for participants and non-participants that have the same distribution of covariates. The current matching estimator draws from the large-scale comparison of matching estimators in Huber, Lechner, and Wunsch (2013) and, in particular, is a version of propensity-score radius matching with regression adjustment, as suggested by Lechner, Miquel, and Wunsch (2011). This has been shown to deliver robust results with high precision.<sup>12</sup> Under the condition that all variables

---

<sup>11</sup> The variables used in this section as part of the so-called propensity score used in the estimator include the number of children in the household, ethnicity, age and age-squared, education, region, local authority characteristics, and numbers of different types of sports facilities. Appendix A presents the variables and results from the probit analysis.

<sup>12</sup> See also Huber, Lechner, and Steinmayr (2015) for operational details of this estimator. The particular version of this estimator used is the RAD\_MATCH Gauss package version 3.1.1. It has the feature that sampling weights are accounted

that jointly influence (confound) the outcome variables as well as the participation variables are included in the matching exercise, the resulting effect would have a causal interpretation. In our case the aim is to keep as constant the ‘exogenous’ confounding factors that can be identified in the data available and at least move towards a causal analysis from one of simply association. Controlling for the impact of counterfactual influences on these outcomes is a natural feature of matching estimators. To improve the credibility of our results by better understanding the effects of important missing confounders, we conduct a formal sensitivity analysis along the lines of Ichino, Mealli, and Nannicini (2008). Based on the available outcome variables in APS, an estimate of the relationship between the different types of sports participation noted above on various labour market outcomes such as individual incomes and on whether the individual is working, unemployed, or retired is undertaken.

## **6 Results of matching estimation**

Table 2 contains the results of the comparisons of the different types of sports with not participating in any of them, while Table 3 presents the results for a direct comparison between the different types of sports.<sup>13</sup>

Table 2 shows that most types of sports participation have positive associations with the individual income of working age males and females and, likewise, on the likelihood of working for both sexes for those aged 26 to 45 years of age. In the former case, team sports are not associated with higher earnings for older females, and in the latter case, outdoor sports not with the chance of employment. This is evidence of the possibility that sports participation enhances labour market outcomes. These results are strongly echoed in the negative associations of sports

---

for in general. Furthermore, bootstrap inference as described in Huber, Lechner and Steinmayr (2015) is based on weights that are combinations of sampling weights, matching weights as well as regression weights. Furthermore, the improved bootstrap smoother as proposed by Racine and MacKinnon (2007) is used to economise on the required bootstrap replications. In addition, the variable *degree* and the sample weight are used as additional variables in the Mahalanobis step, in which the propensity score is overweighed by a factor of 5. The distance measure is set to 150%.

<sup>13</sup> To relate these effect estimates to the appropriate levels of the outcome variables, refer to Table A.1 in the appendix.

participation with unemployment for males of all ages, with the exception of team sports for younger males. The latter result is likely to represent less incidence of the practice of team sports at older ages. There is much less statistically significant evidence of effects on unemployment for females. This only appears in the case of racquet sports. Finally, and consistent with these results, participation in sports is more likely to be associated with a greater chance of being retired for males generally, with the exception of team sports, but only fitness sports for females.

*Table 2: Conditionals associations of sport activities and labour market variables – the comparison to being non-active*

Variables	Subsample	Average effects			
		Men		Women	
		26-45	46-64	26-45	46-64
		Team or fitness or outdoor or leisure vs. no sports			
Household income (annual, in GBP)		<b>4917</b>	<b>4326</b>	<b>4014</b>	<b>2998</b>
Employment (in %)		<b>3.8</b>	0.5	<b>5.2</b>	0.8
Unemployment (in %)		<b>-2.8</b>	<b>-2.3</b>	<b>-1.5</b>	-0.6
Retirement (in %)		0.0	<b>2.4</b>	-0.1	<b>2.1</b>
		Team vs. no sports			
Household income (annual, in GBP)		<b>4292</b>	<b>2575</b>	<b>4562</b>	2171
Employment (in %)		<b>3.9</b>	2.7	<b>3.0</b>	5.4
Unemployment (in %)		<b>-3.5</b>	1.9	-1.8	4.1
Retirement (in %)		-0.1	-0.8	-0.1	-3.9
		Fitness vs. no sports			
Household income (annual, in GBP)		<b>6215</b>	<b>4635</b>	<b>3883</b>	<b>2986</b>
Employment (in %)		<b>4.1</b>	-0.1	<b>4.6</b>	-0.3
Unemployment (in %)		<b>-2.9</b>	<b>-2.6</b>	<b>-1.1</b>	-0.9
Retirement (in %)		-0.1	<b>3.2</b>	0.0	<b>1.3</b>
		Racquet vs. no sports			
Household income (annual, in GBP)		<b>5836</b>	<b>3807</b>	<b>5280</b>	<b>2871</b>
Employment (in %)		<b>4.9</b>	1.6	<b>6.5</b>	-3.8
Unemployment (in %)		<b>-3.6</b>	<b>-4.2</b>	<b>-3.1</b>	1.6
Retirement (in %)		<b>-0.1</b>	<b>3.5</b>	0.0	1.2
		Outdoor vs. no sports			
Household income (annual, in GBP)		<b>6528</b>	<b>5355</b>	<b>3402</b>	<b>5508</b>
Employment (in %)		<b>3.7</b>	0.6	-1.4	-3.0
Unemployment (in %)		<b>-4.3</b>	<b>-3.5</b>	-0.8	-0.2
Retirement (in %)		0.0	<b>3.6</b>	-0.1	<b>5.0</b>
		Leisure vs. no sports			
Household income (annual, in GBP)		<b>4456</b>	<b>4611</b>	<b>3393</b>	<b>2151</b>
Employment (in %)		<b>5.5</b>	1.5	<b>4.3</b>	0.6
Unemployment (in %)		<b>-4.3</b>	<b>-2.7</b>	-1.2	-0.9
Retirement (in %)		<b>-0.1</b>	<b>2.0</b>	0.0	2.7

Note: Inference: 999 bootstrap replications; bootstrapping p-values, smoothed version, linear bias adjustment, symmetric p-values used. **Bold italics**: Variable is significant at 1% level; **Bold**: Variable is significant at 5% level; *Italics*: Variable is significant at 10% level.

Overall, the results of the comparisons of participation in one of the sports compared to nonparticipation suggest that sports participation is associated with earnings increases for males and females of all ages of more than 10%,<sup>14</sup> and greater chances of employment for younger males and females. For males particularly, and to a much lesser extent for females, sports participation is also associated with a general reduction in the chance of unemployment, which for the younger cohort is commensurate with a larger employment rate, while for the older cohort it goes together with an increased share of people in retirement. These results are similar across different sports, despite the different kind of participants (as discussed in section 4).

Each of these associations is based on a comparison of undertaking a different type of sport relative to none at all. Comparing the impact of different types of sports against each other reveals greater heterogeneity in the labour market outcomes from sports participation. For males, moving away from team sports to fitness sports for younger males and towards outdoor sports for older males is associated with earnings increases. Increases are also associated here with moves from racquet sports and, for younger males, from leisure sports to fitness and outdoor activities. In contrast, for the older age group, movements towards team sports from fitness and outdoor activities are associated with greater chances of employment (and reduced retirement and sometimes reduced earnings). This would suggest some distinction in emphasis between access to work and then earnings.

---

<sup>14</sup> Note however that the sensitivity analysis in Appendix B.2 suggests that it might be somewhat smaller than 10%.

*Table 3: Conditionals associations of sport activities and labour market variables – the comparison among the activities*

Variables	Subsample	Average effects			
		Men		Women	
		26-45	46-64	26-45	46-64
Team vs. fitness					
Household income (annual, in GBP)		<b>-2168</b>	<b>-2397</b>	-42	-1374
Employment (in %)		-0.3	<b>5.2</b>	<b>4.6</b>	2.9
Unemployment (in %)		-0.8	-0.8	-1.0	4.2
Retirement (in %)		0.0	<b>-4.3</b>	0.0	-2.5
Team vs. racquet					
Household income (annual, in GBP)		-1308	506	14	1880
Employment (in %)		2.9	2.0	4.6	<b>13.8</b>
Unemployment (in %)		-3.6	0.8	-0.4	1.0
Retirement (in %)		0.0	<b>1.1</b>	0.0	-5.4
Team vs. outdoor					
Household income (annual, in GBP)		-376 <sup>b</sup>	<b>-2141</b>	361	-1189
Employment (in %)		1.1 <sup>b</sup>	<b>3.9</b>	5.0	<b>16.1</b>
Unemployment (in %)		0.1 <sup>b</sup>	-0.2	0.2	-1.0
Retirement (in %)		-0.1 <sup>b</sup>	<b>-3.6</b>	0.0	<b>-10.1</b>
Team vs. leisure					
Household income (annual, in GBP)		978	247	1445	-253
Employment (in %)		1.3	<b>2.8</b>	3.9	7.3
Unemployment (in %)		0.4	-0.2	<b>-1.8</b>	4.0
Retirement (in %)		0.0	<b>-3.4</b>	-0.1	<b>-7.4</b>
Outdoor vs. fitness					
Household income (annual, in GBP)		-654	484	-1873	<b>3677</b>
Employment (in %)		-1.1	-0.2	-7.5	-3.3
Unemployment (in %)		<b>-1.1</b>	-0.8	-0.3	2.1
Retirement (in %)		0.1	1.2	0.0	0.6
Outdoor vs. racquet					
Household income (annual, in GBP)		1261	<b>1711</b>	-570	<b>2857</b>
Employment (in %)		1.5	-2.1	-3.4	1.1
Unemployment (in %)		-1.9	0.7	0.8	0.8
Retirement (in %)		0.1	1.4	0.0	2.1
Fitness vs. leisure					
Household income (annual, in GBP)		<b>3567</b>	1121	<b>1330</b>	1128
Employment (in %)		1.2	0.0	0.7	0.4
Unemployment (in %)		0.6	0.0	-0.2	0.0
Retirement (in %)		0.0	-0.5	-0.1	-1.4
Racquet vs. leisure					
Household income (annual, in GBP)		984	345	1313	1072
Employment (in %)		-0.9	<b>3.0</b>	0.3	-5.7
Unemployment (in %)		0.9	-2.3	-1.1	4.4
Retirement (in %)		0.0	-0.8	-0.1	-1.7
Fitness vs. racquet					
Household income (annual, in GBP)		1491	906	135	-126
Employment (in %)		1.7	<b>-2.6</b>	-0.8	4.0
Unemployment (in %)		-0.2	<b>1.0</b>	<b>2.4</b>	-1.8
Retirement (in %)		0.0	1.2	0.0	1.8
Outdoor vs. leisure					
Household income (annual, in GBP)		<b>2427</b>	1448	412	<b>4038</b>
Employment (in %)		-0.2	-1.1	-4.0	<b>-4.6</b>
Unemployment (in %)		-0.3	-0.1	0.2	2.0
Retirement (in %)		0.1	1.1	-0.1	2.3

Note: Inference: 999 bootstrap replications; bootstrapping p-values, smoothed version, linear bias adjustment, symmetric p-values used. **Bold italics**: Variable is significant at 1% level; **Bold**: Variable is significant at 5% level; *Italics*: Variable is significant at 10% level.

For females, the primary impact appears to be a greater association with higher earnings for older ages in switching towards outdoor activities, or towards fitness activities from leisure activities for younger females, which is the same for younger males. However, in contrast to males, movements from outdoor or racquet sports to team sports are associated with a higher chance of employment for older females. This could signal that the effects of sport on either human, health or social capital that is typically accrued by younger males in their traditional patterns of participation, needs to be compensated for by females later in their working life.

Overall, the results suggest that team sports can contribute somewhat more to employability but that this varies by age across the genders, and that outdoor activities contribute most to income when sports are directly compared. Broadly in line with Rooth (2011) it may well be that golf is important in this regard, which fits the stylised fact of it being an arena for business. It seems likely to be the case that because these activities are associated more with the younger and middle age groups they are structurally connected to initial access to employment and then higher income opportunities with ageing that are associated with a career ladder. However, these vary between the genders. Clearly, these life course effects are worthy of future research. A full discussion of the sensitivity of these results to the sample, missing confounders and alternative estimation strategies is given in Appendix B.2

## **7 Conclusion**

Sport is now widely discussed in government policy and particularly in connection with its potential impact on social welfare and the health of a nation's citizens. In contrast, the role that sport plays in labour market outcomes, with the related economic benefits to society, is much less researched. This lack of knowledge is also true for the UK, which provides the data-base for this research. This paper provides new insights into the human capital impacts of sports participation, recognising that its effects could be mediated through related health and social capital impacts, by signalling higher potential individual productivity and perhaps related to

team work and networking. A further important innovation and novel feature of this paper is that it investigates these impacts for different types of sports.

The analysis is undertaken based on a unique composite dataset drawing upon three major surveys and supplemented by official statistics on population. A matching estimator, supplemented by a formal sensitivity analysis, is employed, in which several of the important confounding factors connected with sports participation and labour market outcomes are controlled for. In particular those that emanate from social and economic characteristics and sports' facilities. Several of the more detailed health, psychological, and social factors are controlled for at the aggregated level of the local authorities. The results are also disaggregated to examine both men and women, and those aged between 26 to 45 years and those aged 46 to 64 years.

The results indicate large positive associations of sports participation with earnings, which appear to be largest for fitness and outdoor sports. Furthermore, there is a negative relation to unemployment, particularly for men. Interestingly, this reduction goes together with higher employment rates for younger men and higher retirement rates for older men. Comparing the different sports against each other reveals that team sports can contribute most to employability, perhaps by signalling teamwork, but that this varies by age across the genders, such that older women might need to accrue these skills; and that outdoor activities contribute most to income when sports are directly compared. There appears to be, therefore, a link between sports participation and the structure of the labour market connected to initial access to employment and then higher income opportunities with ageing that are associated with a career ladder. However, these effects vary between the genders and they to be further investigated.

## References

- Abadie, A. (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models, *Journal of Econometrics*, 113, 231-263.
- Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic Schools, *Journal of Political Economy*, 113, 151-184.



- Andreff, W. (2006). Sports Accounting, in Andreff, W. and S. Szymanski (eds.), *Handbook on the Economics of Sport*, Edward Elgar, Cheltenham UK.
- Angrist, J. D., and J. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton N.J..
- Barron, J. M., B. T. Ewing, and G. R. Waddell (2000). The effects of high school athletic participation on education and labor market outcomes, *Review of Economics and Statistics*, 82, 409-421.
- Becker, G.S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.
- Blundell, R., and M. Costa Dias (2002). Alternative Approaches to Evaluation in Empirical Microeconomics, *Portuguese Economic Journal*, 1, 91–115.
- Blundell, R., and M. Costa Dias (2009). Alternative Approaches to Evaluation in Empirical Microeconomics, *Journal of Human Resources*, 44, 565-640.
- Breuer, C., and P. Wicker (2008). Demographic and economic factors influencing inclusion in the German sport system. A microanalysis of the years 1985 to 2005, *European Journal for Sport and Society*, 5 (1), 33-42.
- Breuer, C., & P. Wicker (2009). Decreasing sports activity with increasing age? Findings from a 20-year longitudinal and cohort sequence analysis. *Research Quarterly for Exercise and Sport*, 80 (1), 22-31.
- Cabane, C. (2011). Do Sporty People Have Access to Higher Job Quality?, in P. Rodríguez, S. Késenne, & B.R. Humphreys (eds.), *The Economics of Sport, Health and Happiness*, Edward Elgar Publishing:UK.
- Carter, P. (2005). Review of National Sport Effort and Resources. *DCMS*, London.
- Coalter, F. (2007). *A wider social role for sport: Who's keeping the score?* London: Routledge.
- DCMS / Strategy Unit (2002). *Game Plan: A Strategy for Delivering Government's Sport and Physical Activity Objectives*. *DCMS*, London.
- DCMS (2010). *Plans for the legacy from the 2012 Olympic and Paralympic Games*, [http://www.culture.gov.uk/images/publications/201210\\_Legacy\\_Publication.pdf](http://www.culture.gov.uk/images/publications/201210_Legacy_Publication.pdf) (accessed 11th February, 2013).
- DCMS (2015). *Sporting Future – A New Strategy for an Active Nation* (accessed 18/5/16).
- Department of Health. (2004). *At least five a week. Evidence on the impact of physical activity and its relationship to health: A report from the chief medical officer*. [http://webarchive.nationalarchives.gov.uk/+/dh.gov.uk/en/publicationsandstatistics/publications/publications/policyandguidance/dh\\_4080994](http://webarchive.nationalarchives.gov.uk/+/dh.gov.uk/en/publicationsandstatistics/publications/publications/policyandguidance/dh_4080994) (accessed 18/5/16).
- Downward, P. M., A. Dawson and T. Dejonghe (2009). *Sports Economics: Theory, Evidence and Policy*, Elsevier: London.
- Downward, P. M., F. Lera-Lopez, and S. Rasciute (2012). The Economic Analysis of Sports Participation, in L. Robinson, P. Chelladurai, G. Bodet, and P. Downward (eds.), *International Handbook of Sports Management*, London: Routledge.
- Downward, P. M., and S. Rasciute (2011). Does Sport Make You Happy? An Analysis of the Well-being Derived from Sports Participation, *International Review of Applied Economics*, 25 (3), 331–348.
- Downward, P. M., and J. Riordan (2007). Social Interactions and the Demand for Sport: An Economic Analysis. *Contemporary Economic Policy*, 25, 518-537.

- Eide, E. R., and N. Ronan (2001). Is Participation in High School Athletics an Investment or a Consumption Good? Evidence from High School and Beyond, *Economics of Education Review*, 20, 431-442.
- Ewing, B. T. (1995). High school athletics and the wages of black males, *Review of Black Political Economy*, 24, 65-78.
- Ewing, B. T. (1998). Athletes and work. *Economics Letters*, 59, 113-117.
- Ewing, B.T. (2007). The Labor Market Effects of High School Athletic Participation Evidence From Wage and Fringe Benefit Differentials, *Journal of Sports Economics*, 8 (3), 255–265
- Felfe, C., M. Lechner, and A. Steinmayr (2016). Sports and Child Development, *PlosOne*, forthcoming.
- Frölich, M. (2007). Nonparametric IV estimation of local average treatment effects with covariates, *Journal of Econometrics*, 139, 35-75.
- Green, M. (2004). Changing Policy Priorities for Sport in England: The Emergence of Elite Sport Development as a key Policy Concern, *Leisure Studies*, 23, 365-385.
- Green, M., and B. Houlihan (2005). *Elite Sport Development: Policy Learning and Policy Priorities*. Routledge: London.
- Heckman, J. J. (1979). Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153-161.
- Houlihan, B. (1997). *Sport, Policy and Politics: A Comparative Analysis*, Routledge: London.
- Huber, M., M. Lechner, and A. Steinmayr, (2015). Radius matching on the propensity score with bias adjustment: tuning parameters and finite sample behaviour, *Empirical Economics*, 49, 1-31.
- Huber, M., M. Lechner, and C. Wunsch (2013). The performance of estimators based on the propensity score, *Journal of Econometrics*, 175, 1-21.
- Hyytinen, A., and J. Lahtonen (2013). The effect of physical activity on long-term income, *Social Science and Medicine*, 96, 129-137.
- Ichino, A., F. Mealli, and T. Nannicini (2008). From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity?, *Journal of Applied Econometrics*, 23, 305-32.
- Imbens, G. (2003). Sensitivity to exogeneity assumptions in program evaluation, *American Economic Review: Papers and Proceedings*, 93, 126-132.
- Imbens, G. W. (2004). Semiparametric Estimation of Average Treatment Effects under Exogeneity: A Review, *Review of Economic Studies*, 86, 4-29.
- Imbens, G. W., and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects, *Econometrica*, 62, 446-475.
- Imbens, G. W., and J. M. Wooldridge (2009). Recent Developments in the Econometrics of Program Evaluation, *Journal of Economic Literature*, 47 (1), 5-86.
- Imbens, G. W., and Th. Lemieux (2008). Regression discontinuity designs: A guide to practice, *Journal of Econometrics*, 142, 615–635
- Ipsos Mori (2007). The Active People Survey Technical Report.
- Jackson, M.O. (2011). An Overview of Social Networks and Economic Applications, in J. Benhabib, A. Bisin, and M.O. Jackson (eds.), *Handbook of Social Economics*, vol. 1, 511 - 585, Elsevier: Amsterdam.

- Kay, T. (2003). The Worklife Balance in Social Practice. *Social Policy and Society*, 2, 231-239.
- Kokolakakis, T., F. Lera-López, and T. Panagouleas (2012). Analysis of the determinants of sports participation in Spain and England, *Applied Economics*, 44, 2785-2798.
- Lechner, M. (2009). Long-run labour market and health effects of individual sports activities, *The Journal of Health Economics*, 28, 839–854.
- Lechner, M. (2010). Difference-in-Difference Estimation, *Foundations and Trends in Econometrics*, 4, 165 – 224.
- Lechner, M., & Sari, N. (2015). Labor market effects of sports and exercise: Evidence from Canadian panel data. *Labour Economics*, 35, 1-15.
- Lechner, M., R. Miquel, and C. Wunsch, (2011). Long-Run Effects of Public Sector Sponsored Training in West Germany, *Journal of the European Economic Association*, 9, 742-784.
- Lipscomb, S. (2007). Secondary school extracurricular involvement and academic achievement: a fixed effects approach, *Economics of Education Review*, 26 (4), 463–472.
- Long, J. E., and S. B. Caudill (1991). The Impact of Participation in Intercollegiate Athletics on Income and Graduation, *The Review of Economics and Statistics*, 73, 525-531.
- Maloney, M. T., and R. E. McCormick (1993). An Examination of the Role That Intercollegiate Athletic Participation Plays in Academic Achievement: Athletes' Feats in the Classroom, *Journal of Human Resources*, 28 (3), 555-570.
- McIntosh, P. (1980). Sport for All' Programmes Throughout the World, A Report for the International Council of Sport and Physical Education for UNESCO, <http://unesdoc.unesco.org/images/0005/000500/050052eb.pdf>, accessed 15<sup>th</sup> March, 2013.
- Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution, *The Journal of Political Economy*, 66 (4), 281–302.
- Morgan, S.L. and Winship, C. (2007). Counterfactuals and Causal Inference: Methods and Principles for Social Research, Cambridge University Press, New York.
- Nannicini, T. (2007). Simulation-based sensitivity analysis for matching estimators, *The Stata Journal*, 7 (3), 334-350.
- Pfeifer, C., and T. Cornelissen (2010). The impact of participation in sports on educational attainment – New evidence from Germany, *The Economics of Education Review*, 29, 94–103.
- Racine, J. S., and J. G. MacKinnon (2007). Inference via kernel smoothing of bootstrap P values, *Computational Statistics & Data Analysis*, 51 (12), 5949–5957.
- Rees, D., and J. J. Sabia, (2010). Sports participation and academic performance: Evidence from the National Longitudinal Study of Adolescent Health, *Economics of Education Review* 29, 751–759.
- Rodgers, B. (1977). *Rationalising sports policies. Sport in its social context: international comparisons (TCCC/DC 77/11E)*. Strasbourg: Council of Europe/Committee for the Development of Sport.
- Rodgers, B. (1978). *Rationalising sports policies. Sport in its social context: international comparisons (Technical Supplement; CDDS 78/19E)*. Strasbourg: Council of Europe/Committee for the Development of Sport.

- Rooth, D.-O. (2009). Obesity, attractiveness and differential treatment in hiring - a field experiment, *Journal of Human Resources*, 44, 710-735
- Rooth, D.-O. (2011). Work out or Out of Work: The Labour Market Return to Physical Fitness and leisure and Sport Activities, *Labour Economics*, 18, 399-409.
- Rosenbaum, P. (1997). Sensitivity analysis to certain permutation inferences in matched observational studies, *Biometrika*, 74, 13-26.
- Rosenbaum, P., and D. Rubin (1983): Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome, *Journal of the Royal Statistical Society Series B*, 45, 212-218.
- Spence, M. (1973). Job Market Signalling, *The Quarterly Journal of Economics*, 87(3), 355-374.
- Stevenson, B. (2010). Beyond the Classroom: Using Title IX to measure the return to High School sport, *The Review of Economics and Statistics*, 92 (2), 284–301
- Van der Klaauw, W. (2008). Regression–Discontinuity Analysis: A Survey of Recent Developments in Economics, *Labour*, 22, 219–245.
- Van Tuyckom, C., J. Scheerder, and P. Bracke (2010). Gender and age inequalities in regular sports participation: A cross-national study of 25 European countries, *Journal of Sports Sciences*, 28 (10), 1077-1084.
- Vytlacil, E. (2002). Independence, monotonicity and latent variable models: an equivalence result, *Econometrica*, 70, 331-341.
- WHO World Health Day (2002). <http://www.who.int/world-health-day/previous/2002/en/> accessed 24th June 2013.
- WHO (2007). Steps to health: A European framework to promote physical activity for health [http://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0020/101684/E90191.pdf](http://www.euro.who.int/__data/assets/pdf_file/0020/101684/E90191.pdf), accessed 24th June 2013.

**Appendix A: Further descriptive statistics**
*(online only)*

Table A.1 shows the descriptive statistics for the variables used in the estimation.

*Table A.1: Descriptive statistics for selected variables*

Variables	Subsample	Weighted sample means			
		Men		Women	
		26-45	46-64	26-45	46-64
Individual characteristics					
Number of children in household (x 100)		90	34	121	24
Ethnicity white in %		84	94	83	94
Age in years		36	54	36	54
Education (in %):Degree or equivalent		44	33	43	28
Higher education		10	11	9	10
A level		17	16	19	14
GCSE		21	22	23	29
Other education		5	15	4	15
Regions (in %)					
London		20	14	20	13
Northeast		3	3	3	4
Northwest		12	12	12	13
Yorkshire		10	11	10	11
West Midlands		10	10	10	10
East Midlands		9	10	9	10
East England		11	12	11	12
Southwest England		8	9	8	10
Southeast England		16	18	17	18
Regional characteristics of local authorities					
LA with largest city 100k-250k		18	16	18	17
LA with largest city below 100k		50	60	51	61
Average numbers of ... per head in %					
Children age 0 to 2		7.1	6.7	7.0	6.7
Long term unemployed		2.0	1.9	2.0	1.9
White ethnicity		86	89	87	90
University degree		32	30	32	30
Log of population in LA		5.2	5.1	5.2	4.9
Average numbers of ... per heads in %					
Illness: Muscular Skeletal		10	10	10	10
Mental anxiety		2.7	2.7	2.7	2.7
Sickness		0.4	0.3	0.3	0.3
# of facilities in LA:					
Athletic tracks		1.2	1.2	1.2	1.1
Golf courses		9.2	10	9.4	10
Grass pitches		208	211	209	209
Fitness suites		28	26	27	25
Indoor bowls		1.0	1.1	1.0	1.1
Indoor tennis centers		1.3	1.2	1.3	1.2
Sports halls		38	36	38	35
Swimming pools		18	17	18	17
Outcome variables					
Annual household income (in GBP)		32423	30432	27459	23830
Working (in %)		92	86	76	73
Unemployed (in %)		5.2	4.7	5.8	4.0
Student (in %)		1.9	0.2	3.5	0.4
Retired (in %)		0.1	8.3	0.0	16
Number of observations		15648	16658	24192	23063

Note: Weighted means using the sampling weights.

Table A.2 contains the full set of estimates for the marginal effects of the probit estimation of any sports (of moderate or vigorous) intensity vs. no such activity.

Table A.2: Average marginal effects for participation in any activity of higher intensity

Variables	Subsample	Average effects (in %-points)			
		Men		Women	
		26-45	46-64	26-45	46-64
Individual characteristics					
Number of children in household		1	2	-3	2
Ethnicity white		<b>14</b>	2	<b>13</b>	<b>12</b>
Age		<i>-8</i>	<i>-1</i>	<i>-.2</i>	<i>-.5</i>
Education: Degree or equivalent		<b>10</b>	<b>12</b>	<b>10</b>	<b>12</b>
Higher education		5	9	2	6
GCSE		-2	-5	-8	-3
Other education		<b>-14</b>	<b>-9</b>	<b>-18</b>	<b>-11</b>
Regions					
Northeast		10	-5	6	-4
Northwest		7	-7	3	-2
Yorkshire		3	-6	1	3
West Midlands		5	-9	5	-1
East Midlands		5	-7	4	1
East England		6	-5	5	1
Southwest England		5	-7	6	1
Southeast England		6	-4	4	0
Regional characteristics of local authorities <sup>0</sup>					
LA with largest city 100k-250k		-3	0	-3	-2
LA with largest city below 100k		-2	-2	-2	0
Average numbers of ... per head					
Children age 0 to 2		-18	-23	<b>-45</b>	<b>-49</b>
Long term unemployed		19	-60	23	-3
White ethnicity		107	11	<b>20</b>	5
University degree		<b>332</b>	<b>-18</b>	<b>36</b>	<b>16</b>
Log of population in LA		9	4	2	1
Average numbers of ... per heads					
Illness: Muscular Skeletal		-290	-34	-2	-23
Mental anxiety		54	<b>-97</b>	-31	-5
Sickness		-52	-165	66	-138
# of facilities in LA:					
Athletic tracks		<b>207</b>	14	54	17
Golf courses		191	4	12	10
Grass pitches		-20	1	1	-1
Fitness suites		6	-6	-7	2
Indoor bowls		-48	-66	-15	<b>-116</b>
Indoor tennis centers		81	<b>130</b>	25	-24
Sports halls		-5	-2	<b>-24</b>	9
Swimming pools		-11	-5	-2	20
Efron's pseudo-R <sup>2</sup> in %		4.0	5.5	5.6	4.6

Note: Average effects are computed as the weighted sample mean of the individual marginal effects. In case of dummy variables the average individual effects of changing the variable from 0 to 1 is used instead of the marginal effect. Inference is based on estimating the standard error by 1999 bootstrap replications taking clustering on the local authority level into account. Further control variables not mentioned in this table are constant term and sampling weight. Reference group for education is A-level, for the regional dummies it is London, for size of the city it is cities larger than 250'000 inhabitants. **Bold italics**: Coefficient is significant at 1% level; **Bold**: Coefficient is significant at 5% level; *Italics*: Coefficient is significant at 10% level.

## **Appendix B: Additional Results of Sensitivity Analysis (*online only*)**

### **B.1 Definition of the sample**

The data contains information on the long term health condition of the individuals. On the one hand, this variable may appear as a confounder as it influences current sports participation. On the other hand, one may also argue that if sports activity habits are only slowly changing, such a variable may be also considered as an outcome variable. The results shown so far are based on treating it as a confounder by removing individuals who indicate to be long term ill (about 12% of the younger cohort and 26% of the older cohort). Table B.1 shows the results based on the second interpretation. Comparing those results with those in Table 4 shows that there are no substantial or systematic differences that could be related to the definition of the samples. All main conclusions remain valid if the long term ill are included in the sample. Concerning the associations of sports with long-term illness, as expected, sports is negatively associated to the probability of being long-term ill. In (almost) all cases this effect is statistically significant at conventional levels. While it appears to be in the range of 2 to 7 %-points for the young cohort, for the older cohort it is about twice that magnitude. It appears to be exceptionally large for team sports. Table B.1: Conditionals associations of sport activities and labour market variables – the comparison to being non-active

Variables	Subsample	Average effects			
		Men		Women	
		26-45	46-64	26-45	46-64
Team or fitness or outdoor or leisure vs. no sports					
Long term illness		<b>-4.6</b>	<b>-8.2+</b>	<b>-3.1</b>	<b>-7.8</b>
Household income (annual, in GBP)		<b>5065</b>	<b>4355+</b>	<b>4247</b>	<b>3027</b>
Employment (in %)		<b>4.5</b>	<b>1.8+</b>	<b>5.5</b>	<b>1.4</b>
Unemployment (in %)		<b>-3.6</b>	<b>-3.1+</b>	<b>-1.8</b>	<b>-1.4</b>
Retirement (in %)		<b>-0.2</b>	<b>2.0+</b>	<b>-0.1</b>	<b>1.0</b>
Team vs. no sports					
Long term illness		<b>-7.1</b>	<b>-14.5+</b>	<b>-5.2</b>	<b>-14.5</b>
Household income (annual, in GBP)		<b>4592</b>	<b>3828+</b>	<b>4231</b>	1546
Employment (in %)		<b>5.3</b>	<b>5.7+</b>	<b>7.3</b>	<b>5.1</b>
Unemployment (in %)		<b>-5.1</b>	<b>-2.4+</b>	<b>-3.1</b>	<b>0.1</b>
Retirement (in %)		<b>-0.2</b>	<b>-2.9+</b>	<b>-0.1</b>	<b>-5.1</b>
Fitness or racquet sports vs. no sports					
Long term illness		<b>-5.2</b>	<b>-8.6</b>	<b>-3.8+</b>	<b>-9.1</b>
Household income (annual, in GBP)		<b>6563</b>	<b>4589</b>	<b>4244+</b>	<b>3291</b>
Employment (in %)		<b>4.9</b>	<b>1.4</b>	<b>5.5+</b>	<b>0.7</b>
Unemployment (in %)		<b>-3.9</b>	<b>-3.5</b>	<b>-1.9+</b>	<b>-1.3</b>
Retirement (in %)		<b>-0.2</b>	<b>2.9</b>	<b>-0.1+</b>	<b>1.2</b>
Outdoor vs. no sports					
Long term illness		<b>-3.5</b>	<b>-11.1</b>	<b>-4.4+</b>	<b>-9.1</b>
Household income (annual, in GBP)		<b>7637</b>	<b>5348</b>	<b>4194+</b>	<b>6053</b>
Employment (in %)		<b>5.9</b>	<b>2.4</b>	<b>-0.1+</b>	<b>-2.6</b>
Unemployment (in %)		<b>-6.0</b>	<b>-3.9</b>	<b>-1.4+</b>	<b>1.0</b>
Retirement (in %)		<b>0.1</b>	<b>2.2</b>	<b>0.1+</b>	<b>1.8</b>
Leisure vs. no sports					
Long term illness		<b>-2.5</b>	<b>-4.9</b>	<b>-1.3+</b>	<b>-3.6</b>
Household income (annual, in GBP)		<b>4092</b>	<b>4820</b>	<b>3703+</b>	<b>2529</b>
Employment (in %)		<b>4.8</b>	<b>0.8</b>	<b>4.3+</b>	<b>1.5</b>
Unemployment (in %)		<b>-4.4</b>	<b>-3.2</b>	<b>-1.0+</b>	<b>-0.5</b>
Retirement (in %)		<b>-0.2</b>	<b>3.2</b>	<b>0.0+</b>	<b>1.2</b>

Note: Inference: 999 bootstrap replications; bootstrapping p-values, smoothed version, linear bias adjustment, symmetric p-values used. **Bold italics**: Variable is significant at 1% level; **Bold**: Variable is significant at 5% level; *Italics*: Variable is significant at 10% level. +: Inference based on asymptotic standard errors.

## B.2 Formal sensitivity analysis of matching estimation

Since the sensitivity of the matching estimates with respect to potentially missing confounders is an issue that might seriously bias results, there are several suggestions in the literature on how to assess such sensitivities. It appears to us that a sensitivity analysis should try to understand the sensitivity due to unobserved confounders given already existing control variables (thus ruling out the approach by Altonji, Elbers, and Taber, 2005), and at the same time not introduce new restrictions, for example in the form of a requirement to model the relationship between outcomes, treatment, and confounders. This second condition rules out the early approaches by Rosenbaum and Rubin (1993), Rubin (1997), as well as by Imbens (2003).



However, the simulation based approach suggested in Ichino, Mealli, and Nannicini (2008) and Nannicini (2007) fulfils those criteria and can be used together with any matching method. The idea is simulate how the inclusion of a confounder that is related to the potential outcomes and the treatment affects the results. This simulated confounding variable is included in the propensity score and the estimation proceeds in the standard way. Comparing the estimates with and without this new covariate indicates the sensitivity of the chosen matching specification to missing confounding variables. Although, they propose this approach specifically for the estimation of the average treatment effect on the treated, the general principle can be applied to all mean treatment effects usually considered in the literature.

In order to simplify the simulations, in this approach the unobserved confounder as well as the outcomes are treated as binary and independent of (included) covariates. In this case, the key design parameters for simulating the new dummy variables,  $U$ , are its probability for taking a value of 1 that varies for the four strata defined by outcome and treatment, i.e. ( $Y$  denotes the outcome variable and  $D$  denotes the treatment). If the values vary with treatment and outcome, then including  $U$  among the confounders has the potential to affect the results even after controlling for all covariates in the model. Since for each observation 'i' the realisation of  $U$  will be random, we follow the suggestion of Ichino, Mealli, and Nannicini (2008) to repeat the simulations many times and then take the mean of the effects over the simulations. Furthermore, we choose the design probabilities such that they reflect an artificial unbalanced design expected to lead to potentially substantial biases as well as reflecting the correlation patterns of two important observed confounders, namely the size of the local authority and whether individuals have obtained an academic degree. The latter two scenarios attempt to answer the question of how much bias would be there if there are other confounding variables omitted that share the same 'confounding structure/potential' as those two important control variables, but are at the same independent of those and all other already included controls.

Using the flexible approach of Ichino, Mealli, and Nannicini (2008) and Nannicini (2007) requires further decisions to become implementable in our setting. Many of those decisions have also the purpose of keeping the complexity, and thus the computational costs, at bay. First, as outcome variable we chose income and split it at the median (for the purpose of simulation the confounders only). Income is more relevant here than other labour market states, because it is directly related to potential missing variables, like ability, self-discipline, or beauty, for example. The second decision is on how to choose the values of  $p^{ij}$ . We consider four scenarios: The first is a ‘neutral’ scenario with  $p^{ij} = 0.5, \forall i, j$ , in which the simulated confounder should not lead to any bias providing a benchmark for checking whether the randomness induced by the simulation has any impact.<sup>15</sup> It provides the benchmark to compare the more selective scenarios with. The second scenario is based on an unbalanced design with respect to income, i.e.  $P(U = 1 | Highincome, D = 1) = P(U = 1 | Highincome, D = 0) = 0.4$ . The corresponding values for low income are 0.2 (to be denoted *S1* in the respective table). Two further scenarios mimic the relation of two important confounders to outcome and treatment (population size of local authority, *S2*, education, *S3*). Thus, in those scenarios, the  $p^{ij}$ ’s vary over the different samples and comparisons.

Finally, there is the issue of how to obtain inference in this case and of how many times to draw from the distribution of  $U$ . With respect to inference, it appears natural to use bootstrap inference as the true asymptotics are not known for the matching estimator used. Since this procedure is computational very intensive in our rather large samples, we set the number of draws of  $U$  (within each bootstrap replication) to 19 and use only 99 bootstrap replications. The test statistics are based on the difference of the baseline scenario to the various confounding

---

<sup>15</sup> The average effects obtained under the baseline are almost identical to the main results given in Table 4.

scenarios using the quantile method.<sup>16</sup> In Table B.3, we report those deviations from the baseline results as well as their significance. As before, we focus on the comparisons to non-participation. They are in the form of deviations from the baseline scenario (i.e. the situation when the additional covariate  $U$  is not confounding). Overall, it is fair to conclude that sensitivity is limited, as only 6.4% (17 out of a total of 264 effects) of the deviations from the baseline scenario are significant at the 5% level. This is despite the fact that the confounders are constructed in a way to be fairly ‘harmful’.

Only four income differences, all related to scenario 2, are large enough to change the conclusions from the main estimation. Two of them are related to racquet and outdoor sports for young men. The other two differences are observed for racquet and leisure sports for older women. In all four cases, these large negative earnings effects go together with reduced employment, although the latter difference is only significant for older women.

For the other sports done by young men or older women as well as for the results of the two other populations no such severe sensitivities appear. In detail, we find the following: First, the employment, unemployment, and retirement results are very robust against these types of omitted confounders. The deviations from the baseline scenario are small and never significant. Second, since the way the simulation of  $U$  is geared towards particularly ‘attacking’ the earnings results, it is not surprising that in some populations and some sports types the deviation from the baseline earnings effects are statistically significant and sometimes of a non-negligible quantity. In particular, such significant effects appear for sports of young men, team sports of young women, for fitness sports, for racquet sports of women and young men, and leisure sports

---

<sup>16</sup> Even with those not too large numbers, the sensitivity analysis took 5 months to compute using our off-the-shelf multi-core hardware and reasonably efficiently programmed Gauss codes which include explicit multithreading. Obviously, using the quantile method with such a small number of observations makes inference at low significance levels unreliable. However, we expect to have reasonable inference at the 5 and 10% level nevertheless.

of women. When such effects appear, their magnitude indicates that the estimated effects may overestimated by 10 to 50%.

In conclusion, our sensitivity check suggest in general that the main findings are fairly robust with respect to confounders that are in the range of what might be expected for missing confounding variables. However, the true earnings effects might be somewhat lower than the ones presented in Table 4. Nevertheless, the main conclusion of positive earnings effects for all sports activities is not substantially challenged by the findings of the sensitivity analysis, nor are the results for the other labour market outcomes considered. The exceptions appear to be racquet and outdoor sports for which large sensitivities appear among the population of young men.

*Table B.3: Sensitivity check – Difference of effects under different confounding scenarios relative to the baseline scenario*

Variables	Subsample	Average effects			
		Men		Women	
		26-45	46-64	26-45	46-64
Team or fitness or outdoor or leisure vs. no sports					
Household income (annual, in GBP)	S1	<b>-986</b>	--	--	4.8
	S2	-967	--	--	-44
	S3	<b>-1042</b>	--	--	-752
Employment (in %)	S1	-0.2	--	--	-0.2
	S2	-0.2	--	--	-0.3
	S3	-0.2	--	--	-0.6
Unemployment (in %)	S1	0.1	--	--	0.1
	S2	0.1	--	--	0.1
	S3	0.1	--	--	0.1
Retirement (in %)	S1	-0.0	--	--	0.1
	S2	-0.0	--	--	0.2
	S3	-0.0	--	--	0.5
Team vs. no sports					
Household income (annual, in GBP)	S1	-792	-360	-1394	-1263
	S2	-850	-107	-1477	-690
	S3	-728	-379	<b>-1456</b>	-1657
Employment (in %)	S1	0.3	-0.4	-1.0	0.4
	S2	0.3	-0.2	-1.0	-0.1
	S3	0.2	-0.4	-1.1	0.6
Unemployment (in %)	S1	0.1	0.1	0.6	0.2
	S2	0.1	0.1	0.7	0.2
	S3	0.2	0.1	0.6	0.4
Retirement (in %)	S1	0.0	0.3	0.0	-0.5
	S2	0.0	0.1	0.0	-0.1
	S3	0.0	0.2	0.0	-0.9
Fitness sports vs. no sports					
Household income (annual, in GBP)	S1	<b>-1320</b>	-534	<b>-723</b>	35
	S2	<b>-1096</b>	-296	-642	23
	S3	<b>-1422</b>	<b>-1141</b>	<b>-1386</b>	<b>-728</b>
Employment (in %)	S1	-0.2	0.1	-0.3	-0.2
	S2	-0.2	0.0	-0.2	-0.2
	S3	-0.0	-0.1	-0.3	-0.0
Unemployment (in %)	S1	0.3	0.0	0.0	0.1
	S2	0.3	0.0	-0.1	0.1
	S3	0.3	0.2	-0.0	-0.1
Retirement (in %)	S1	-0.0	-0.1	0.0	0.1
	S2	-0.0	-0.0	0.0	0.1
	S3	-0.0	0.3	0.0	0.1

Note: Table B.3 to be continued.

Table B.3 continued

Variables	Subsample	Average effects			
		Men		Women	
		26-45	46-64	26-45	46-64
Racquet sports vs. no sports					
Household income (annual, in GBP)	S1	-1190	161	-1684	1291
	S2	<b>-19626</b>	131	-1779	<b>-17932</b>
	S3	<i>-1275</i>	<i>-758</i>	<b>-2628</b>	<i>-376</i>
Employment (in %)	S1	0.1	0.3	-1.2	0.7
	S2	-2.4	0.1	-0.9	-7.8
	S3	0.1	-0.5	-0.7	0.2
Unemployment (in %)	S1	-0.1	0.0	0.2	0.7
	S2	2.4	0.0	0.1	-0.3
	S3	-0.1	-0.1	-0.2	0.3
Retirement (in %)	S1	0.0	-0.4	0.0	-0.3
	S2	0.0	-0.1	0.0	5.6
	S3	0.0	-0.5	0.0	-0.2
Outdoor vs. no sports					
Household income (annual, in GBP)	S1	-1245	155	-437	-119
	S2	<b>-20556</b>	118	-934	-436
	S3	<i>-945</i>	<i>-26</i>	<i>-1344</i>	<i>-1226</i>
Employment (in %)	S1	-3.1	0.1	-0.1	0.5
	S2	0.5	0.1	0.0	0.6
	S3	0.5	0.3	0.0	0.0
Unemployment (in %)	S1	-0.3	0.0	0.2	-0.1
	S2	0.8	-0.0	0.1	-0.1
	S3	-4.1	-0.0	0.2	0.2
Retirement (in %)	S1	0.1	-0.1	0.0	-0.2
	S2	0.2	-0.1	0.0	-0.3
	S3	0.1	-0.2	0.0	0.2
Leisure vs. no sports					
Household income (annual, in GBP)	S1	-1011	-167	-737	-255
	S2	-799	-63	-780	<b>-13550</b>
	S3	<i>-1101</i>	<i>-717</i>	<b>-1284</b>	<i>-877</i>
Employment (in %)	S1	0.3	0.2	-0.1	-0.1
	S2	0.1	0.2	-0.1	<b>-7.1</b>
	S3	0.2	-0.1	-0.1	-0.3
Unemployment (in %)	S1	-0.3	-0.0	0.0	0.0
	S2	-0.3	-0.0	0.0	0.5
	S3	-0.1	-0.1	0.0	0.5
Retirement (in %)	S1	-0.0	-0.2	-0.0	0.0
	S2	-0.0	-0.2	-0.0	<b>5.0</b>
	S3	-0.0	0.2	-0.0	-0.1

Note: Difference of effects in confounding scenario to baseline scenario (S1:  $P(U = 1 | \text{high income}) = 0.4$ ;  $P(U = 1 | \text{low income}) = 0.2$ ; S2: 'population in LA', S3: 'education degree'. Inference: 99 bootstrap replications and 19 draws of simulated binary confounder; quantile method, smoothed version, linear bias adjustment, symmetric p-values used. **Bold:** Variable is significant at 5% level; *Italics:* Variable is significant at 10% level. 1% level is not indicated because of low number of bootstrap replications. In the comparisons of sports vs. no sports for older men and younger women, some of the probit estimates of the propensity scores did not converge for the simulated confounders. Thus, these results are omitted from the table.

### **B.3 Instrumental variable estimation**

One may argue that the availability of fitness facilities has a direct effect on individual labour market outcomes, at least once we control for the main variables that capture regional economic differences. If true, the availability of such facilities may serve as instruments and provide an alternative way to identify the effects. The results of conventional two stage least square estimations with this instrument are contained in Tables B.4 and B.5 using different transformations of the instrument. The conclusions from all these estimations: Since the first stages are not strong enough for the sample size available (sometimes they are not even significant) is that it is impossible to determine the effects precisely enough. Thus, although these results do not contradict the matching results, this robustness check has very limited power.

*Table B.4: Conditionals associations of sport activities and labour market variables – the comparison to being non-active*

Variables	Subsample	Average effects			
		Men		Women	
		26-45	46-64	26-45	46-64
Team or fitness or outdoor or leisure vs. no sports					
First stage (x 1000)		0.16	0.01	0.06	<b>0.13</b>
Household income (annual, in GBP)		86138	--	--	55509
Employment (in %)		0.8	--	--	63.7
Unemployment (in %)		-23.8	--	-12.4	7.7
Retirement (in %)		3.4	--	-32.6	-53.3
Team vs. no sports					
First stage (x 1000)		0.29	0.15	-0.01	0.05
Household income (annual, in GBP)		30339	6354	--	--
Employment (in %)		-69.5	--	--	--
Unemployment (in %)		13.4	-48.1	--	--
Retirement (in %)		50.4	--	--	--
Fitness sports vs. no sports					
First stage (x 1000)		<b>0.38</b>	<b>0.46</b>	0.24	0.13
Household income (annual, in GBP)		57902	18362	59751	--
Employment (in %)		-17.6	45.5	-0.8	--
Unemployment (in %)		15.8	-20.2	-17.6	--
Retirement (in %)		1.3	-24.3	3.6	--
Racquet sports vs. no sports					
First stage (x 1000)		-0.09	<b>0.51</b>	<b>0.38</b>	<b>0.32</b>
Household income (annual, in GBP)		--	1307	29815	37968
Employment (in %)		-14.5	--	-41.8	-49.1
Unemployment (in %)		--	--	49.3	20.0
Retirement (in %)		-12.3	-5.2	0.3	-49.2
Outdoor vs. no sports					
First stage (x 1000)		<b>0.65</b>	<b>0.49</b>	<b>0.41</b>	0.20
Household income (annual, in GBP)		51368	30504	58761	62868
Employment (in %)		-9.4	--	3.8	--
Unemployment (in %)		-13.8	-60.7	-40.8	18.9
Retirement (in %)		1.4	-45.5	1.8	--
Leisure vs. no sports					
First stage (x 1000)		-0.07	0.10	0.25	0.02
Household income (annual, in GBP)		--	--	86805	--
Employment (in %)		30.7	--	--	--
Unemployment (in %)		--	--	4.4	--
Retirement (in %)		12.1	41.1	--	--

Note: Inference: 199 bootstrap replications; quantile method. **Bold italics**: Variable is significant at 1% level; **Bold**: Variable is significant at 5% level; *Italics*: Variable is significant at 10% level. Instrument measures available facilities related to specific sports at LA level per 10.000 inhabitants. 2SLS results. --: Estimated effects not reported because they violate support of the respective outcome variable.



*Table B.5: Conditionals associations of sport activities and labour market variables – the comparison to being non-active (standardized instruments)*

Variables	Subsample	Average effects			
		Men		Women	
		26-45	46-64	26-45	46-64
Team or fitness or outdoor or leisure vs. no sports					
First stage (x 1000)		-0.5	-2.4	-7.5	1.0
Household income (annual, in GBP)		--	--	--	67385
Employment (in %)		--	--	-32.9	-10.9
Unemployment (in %)		--	--	87.6	-15.6
Retirement (in %)		0.3	14.0	0.3	2.1
Team vs. no sports					
First stage (x 1000)		-1.8	1.9	-0.8	0.5
Household income (annual, in GBP)		43773	50647	--	--
Employment (in %)		64.8	--	--	--
Unemployment (in %)		-22.3	--	37.8	-22.2
Retirement (in %)		-10.7	16.2	-4.9	--
Fitness sports vs. no sports					
First stage (x 1000)		<b>0.63</b>	-1.7	-2.5	-0.9
Household income (annual, in GBP)		31035	--	--	55962
Employment (in %)		10.2	--	--	--
Unemployment (in %)		-0.3	--	--	--
Retirement (in %)		3.1	--	-1.3	--
Racquet sports vs. no sports					
First stage (x 1000)		-0.2	-0.8	-2.3	-0.2
Household income (annual, in GBP)		--	--	--	-6909
Employment (in %)		-57.3	--	66.6	--
Unemployment (in %)		31.8	--	56.6	13.9
Retirement (in %)		-12.6	--	4.3	--
Outdoor vs. no sports					
First stage (x 1000)		0.3	2.8	-0.5	<b>2.9</b>
Household income (annual, in GBP)		88360	53100	--	64021
Employment (in %)		59.2	27.1	-21.1	-11.4
Unemployment (in %)		--	-17.4	--	-27.1
Retirement (in %)		14.8	-3.1	4.3	52.2
Leisure vs. no sports					
First stage (x 1000)		-0.9	0.7	-0.1	<b>3.0</b>
Household income (annual, in GBP)		--	--	--	36258
Employment (in %)		--	--	--	-11.9
Unemployment (in %)		--	--	--	3.0
Retirement (in %)		--	18.9	--	-24.5

Note: Inference: 199 bootstrap replications; quantile method. ***Bold italics***: Variable is significant at 1% level; **Bold**: Variable is significant at 5% level; *Italics*: Variable is significant at 10% level. Instrument measures available facilities related to specific sports at LA level per 10.000 inhabitants normalized to mean zero and standard deviation one. 2SLS results. --: Estimated effects not reported because they violate support of the respective outcome variable.