Universität St.Gallen

Sport Participation and Child Development in Less Developed Countries

Tim Pawlowski, Ute Schüttoff, Paul Downward, Michael Lechner

October 2014 Discussion Paper no. 2014-33

School of Economics and Political Science,
Department of Economics

University of St.Gallen

Editor: Martina Flockerzi
University of St.Gallen
School of Economics and Political Science
Department of Economics
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35
Email seps@unisg.ch

Publisher: School of Economics and Political Science
Department of Economics
University of St.Gallen
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25

Electronic Publication: Fax +41 71 224 31 35
http://www.seps.unisg.ch

Sport participation and Child Development in Less Developed Countries[1]

Tim Pawlowski, Ute Schüttoff, Paul Downward, Michael Lechner [2]

Authors' address:

Prof. Dr. Tim Pawlowski
University of Tübingen
Wilhelmstraße 12
D-72074 Tübungen
Email tim.pawlowski@uni-tuebingen.de

Ute Schüttoff
University of Tübingen
Wilhelmstraße 12
D-72074 Tübungen
Email ute.schuettoff@uni-tuebingen.de

Dr. Paul Downward
Loughborough University
Leicestershire, LE11 3TU, UK
Email p.downward@lboro.ac.uk

Prof. Dr. Michael Lechner
Swiss Institute for Empirical Economic Research (SEW)
University of St.Gallen
Varnbüelstrasse 14
CH-9000 St. Gallen
Phone +41 71 224 28 14
Fax +41 71 224 23 02
Email michael.lechner@unisg.ch
Website www.michael-lechner.eu

---

[1] This version: 6 October 2014. Comments are very welcome.

[2] Michael Lechner is also affiliated with CEPR and PSI, London, CESIfo, Munich, IAB, Nuremberg, and IZA, Bonn. A previous version of the paper was presented at the Economics Research Seminar of the Department of Economics at the University of St. Gallen (November 2013) as well as the meetings of the German Society of Sport Science, Konstanz (September 2013), the North American Association of Sports Economists, Denver (July 2014) and the European Sport Economics Association, Antwerp (September 2014). We thank participants for helpful comments and suggestions. The usual disclaimer applies.

Abstract

Previous research in *developed countries* suggests that sports participation can positively influence child development. We use panel data of a cohort of 658 children in Peru to test this relationship in a *less developed country* where conditions for child development are worse. To identify the causal effect of children's sport participation we exploit the panel structure of our data together with using propensity score matching. Our findings suggest that participation in a sports group has positive impacts on two of the key factors that are identified in the literature as central to child development, i.e. subjective health and social capital.

Keywords

Social capital, Human capital, Well-being, Health, Group participation, Sports.

JEL Classification

C14, D12, I21, J24.

I. Introduction

Policy makers have traditionally embraced the role of sport in providing health and social benefits for both, adults and children – becoming manifest for instance in the *European Sports Charter*. In line with this, numerous programs around the globe promote sports participation of (disadvantaged) children in less developed countries with the explicit objective to foster their skill formation and overall development.[3]

However and in contrast to the popular and widespread belief in the beneficial effects associated with sports participation, empirical evidence based on large-scale survey data is largely missing. Furthermore, the few existing studies exploring the outcomes of sports participation on young people have exclusively focused on *developed countries* (see Section II) where passive leisure activities such as watching TV are popular substitutes for physical activity, and employment as well as education possibilities to enhance human and social capital are widespread. This is different in poorer countries where access to education is more limited (Caballero, 2005) and the conditions to enhance social capital are worse (Bornstein *et al*., 2012). Surprisingly, studies analyzing the factors associated with child development in *less developed countries* have not considered testing for the (causal) relationships that might exist between sport and development (a recent example is Helmers and Patnam, 2011).

This paper contributes to the existing literature by formally testing for the presence of a causal relationship between sports participation and child development in Peru. Peru was chosen as a case-study because it is still developing and involved in programs to increase children's sports participation – most notably the *Volunteer Sports Programs*[4] and the *Sports Programs for (Disadvantaged) Children.*[5] Moreover, the sport (system) offers some similarities compared

[3] A prominent example is the UNICEF *Sport for Development* program (in cooperation with partners such as the Federation Internationale de Football Association, FIFA): http://www.unicef.org/sports/index_57597.html

[4] The aim of this program is to involve (young) people from abroad helping to teach children and adolescents aged 6 to 17 years a variety of sports in sports clubs/sports centers and at schools: http://www.volunteersports.co.uk/sports/school-sports/peru/

[5] The aim of this program is to provide a complement to the public education system for children and adolescents aged 6 to 15 years with physical and technical training for a sport as well as measures for educational and general development: http://proxied.changemakers.net/journal/02april/wolf.cfm

to those in developed countries, i.e. there are coaches in many sports groups (often organized in clubs) and the most popular sport is a team sport (football). Therefore, the country offers an interesting and highly relevant setting to test this relationship.

The data used covers a cohort of 658 children who had been surveyed in three periods (2002, 2006/07 and 2009) about all areas of their life.[6] While this data does not allow for identifying specific sports programs it enables us to analyze the effects associated with general participation in a sports group. By exploiting the panel structure of our data, and flexible estimation by propensity score matching, we argue that the identified associations have a causal interpretation. Overall, the findings suggest that programs in a less developed country, such as those provided in sport, can have positive developmental impacts for children. However, whilst we find positive impacts on subjective health and social capital we could not detect statistically significant effects on well-being and human capital.

The reminder of the paper is as follows: we start with a discussion of the related literature in Section II. Information on the characteristics of Peru and its sport systems is presented in Section III. Section IV outlines the identification and estimation strategies employed, before discussing the results in Section V and drawing conclusions in Section VI.

II. Literature review: The outcomes from sport for young people

Unlike the considerable literature based on large-scale survey data that examines the factors associated with why people participate in sport (for a recent overview, see Breuer *et al.*, 2011), and the growing literature examining the outcomes associated with sports participation for adults (see for example Downward *et al.*, 2012; Lechner and Downward, 2013, and Cabane and Lechner, 2014), there is relatively less research that formally tests the effects of sport on young peoples' lives. In the few studies available, the impact of sport on human capital – which suggests that exercise has beneficial impacts on cognitive functions e.g. to enhance learning and memory (Gomez-Pinilla, 2008) – is the most researched area. Several

[6] The data used in this publication come from Young Lives, a 15-year study of the changing nature of childhood poverty in Ethiopia, India (Andhra Pradesh), Peru and Vietnam (www.younglives.org.uk). Young Lives is core-funded by UK aid from the Department for International Development (DFID) and co-funded from 2010 to 2014 by the Netherlands Ministry of Foreign Affairs. The views expressed here are those of the author(s). They are not necessarily those of Young Lives, the University of Oxford, DFID or other funders.

studies have specifically focused on analyzing the direct and indirect effects of curricular and extra-curricular formal sport participation, either in education or in clubs, on cognitive and non-cognitive skill formation (A tabular overview is provided in Table A.1 in the Appendix).

Barron *et al.* (2000) make use of an instrumental variable approach to examine the effects of high school athletic participation on education and labor market outcomes in the US. Private school, school enrolment, measures of health, family income, and location (central city) as well as information on the individual's height and weight (for the National Longitudinal Survey of Youth) at the time the individual was a high school student are included as instruments. They find evidence that there is a direct impact of athletic participation on wages and educational attainment and argue that this reflects differences across an individual's ability or preference for leisure. However, some instruments are likely to have direct effects on their outcomes, for example the reputational impacts of the school, or the neighbourhood in which a child grows up could affect employment opportunities and are thus hardly credible. Further results are provided by Eide and Ronan (2001), who use an instrumental variable (IV) approach as well. They also use height as an instrument for participation. They show that for white male students (school-sponsored extracurricular) sports participation has a negative effect on their educational attainment, but a positive effect for white female students.

In a study using data from the Michigan Study of Adolescent Life Transitions (MSALT) Eccles *et al.* (2003) examine extracurricular activities and adolescent development. It is found that involvement in (school) team sports is associated with better educational outcomes (in grade 10 and 12). Furthermore, sports participants like school more and would rather attend college full time than is the case with none sports participants. However, only gender, intellectual aptitude (test for verbal and mathematical abilities), and the mother's educational level are entered as controls in the regression analysis. In contrast, a fixed effects strategy is used by Lipscomb (2007) to test whether sports activity provides an immediate return to student learning. Results show that, independent of individual ability, athletic and other club participation (in school) are both associated with an increase in math test scores and with Bachelor's degree attainment expectations. Athletic club participation is also associated positively with science test scores. Finally, Stevenson (2010) makes use of changes in US law providing for equal provision of sporting opportunities for both male and female students as a natural experiment to show that the post legislation increases in female sports participation increased female college attendance, employment and wages. 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. The interaction of the Title IX legislation with preexisting levels of boys' sports participation forms the instrument for the change in girls' athletic participation over the period of 1972–1978.

Besides these papers three studies address the European context (Germany) and focus on the wider impacts of sport than on just human capital: Pfeifer and Cornelißen (2010) analyze the impact of participating in extra-curricular athletic activities on educational attainment (in the form of secondary school degrees and professional degrees) for young Germans. They use the 2000-2005 waves of the German Socio-Economic Panel with individuals having finished their education (aged 17-99) and where retrospective information on their adolescence is available. The sample includes 2,930 males and 3,071 females. Ordered probit models with generalized thresholds are applied to estimate the effect of participation in sport activities on the achievement of levels of secondary school degrees and professional degrees. Results show that participation in sport activities has a significant positive impact on educational attainment. The robustness of the results is tested by applying an IV-approach (using body height and city size as instruments).

Felfe *et al.* (2011) analyze the effect of sports club participation on skill development of children in Germany aged 3 to 10 years using data from the "German Health Interview and Examination Survey for Children and Adolescents" (KiGGS) as well as the "German Child Panel" (GCP). The effect of sports on children's cognitive (overall school grade) and non-cognitive skill measures (emotional problems, behavioral problems, hyperactivity, peer problems, and antisocial behavior) are estimated by employing matching methods. To avoid selection bias, a wide range of background characteristics are controlled for. The local availability of sports facilities is also included as an instrumental variable as a robustness check. Results show that there is a positive effect of participation in sports on children's cognitive and non-cognitive skills. In addition to these effects, Felfe *et al.* (2011) found that children participating in sports clubs score higher on different health and well-being measures.

Finally, Gerlach and Brettschneider (2013) analyze the impact of sports club participation by children and adolescents on social capital, i.e. social acceptance and respect by peers. Based on a longitudinal study of 1,637 individuals in Germany, they detect a significant positive

association using variance analyses. This study also indicates a positive relation on well-being (in line with Felfe *et al.*, 2011) while there is no significant positive effect on obesity.

III. Study program and institutional background

As the review of the existing literature suggests there is some evidence that sports participation in a variety of settings positively influences the development of young people in industrialized countries. In less developed countries, however, the conditions for the development of young people are generally different and certainly worse, at least when measured with resources available for fostering their development. For instance, children often grow up with malnutrition which is associated with poor developmental outcomes.[7] In addition, there is a dearth of social capital such as limited interpersonal cohesion, social networks and availability of civil organizations (Bornstein *et al*., 2012). Furthermore, manual labor and daily-survival activities demand high energy and physical *in*activity is not a major problem (Caballero, 2005). These differences make it interesting and highly relevant to test, whether sports participation affects child development also in less developed countries.

The data used is taken from the Young Lives (YL) study program which is an international study of childhood poverty and was established to pursue the changing lives of 12,000 children in Ethiopia, India (in the state of Andhra Pradesh), Peru and Vietnam over 15 years (2001 – 2016) (Young Lives, 2013). The countries in the YL study program were selected to reflect a wide range of cultural, political, geographical and social contexts. In each country 20 sentinel sites were selected non-randomly, with rich areas excluded and poor areas purposively over-sampled. Children in the right age group were sampled randomly in the selected sites (Young Lives Method Guide, 2011). To create two age cohorts 2,000 infants (aged between 6 and 18 months) and 700-1,000 older children (aged 7 to 8 years) were

[7.] For instance, Victora *et al.* (2008) show for cohort studies from Brazil, Guatemala, India, the Philippines, and South Africa that poor fetal growth or stunting in the first two years of life result in irreversible damage, e.g. shorter adult height, lower attained schooling, and reduced adult income. Further, malnutrition in early childhood can cause vocabulary deficits (e.g. Grantham-McGregor, 2002) and other deficiencies in school performance and intelligence (Grantham-McGregor, 1995). Alderman (2010) argues that early childhood nutrition is a long-term investment in health, nutritional and cognitive development. However, Crookston *et al.* (2011) point out that poor nutritional status alone does not account for children's cognitive deficits. Their study confirms that other factors such as wealth, maternal education, area of residence and number of siblings are also important determinants of verbal and quantitative ability.

selected. International experts developed the interview questionnaires (on child, household and community level) which were used in all four YL countries (Crookston *et al.*, 2011).[8]

Our study is focused on Peru since the country is involved in programs to increase children's sports participation and the sport (system) offers some similarities compared to those in developed countries, i.e. there are coaches in many sports groups (often organized in clubs) and the most popular sport is a team sport (football). This suggests that the research questions addressed in this paper that are ostensibly derived from a literature directed at more developed countries nonetheless have a direct policy relevance for a less developed country such as Peru. The country is located in the west of South America and is populated by 29.8 million people with 77% living in urban regions (in 2010) (USA in 2010: 82%). The poverty rate is 28% (USA in 2010: 15.1%) and there is a substantial inequality between rural and urban children (Escobal *et al.*, 2003). Life expectancy at birth is 71 years for men and 75 years for women (USA in 2013: 77 years for men; 81 years for women). The infant mortality rate is 21 deaths per 1,000 births (USA in 2014: 6.2 deaths per 1,000 live births) and health expenditure was 5.1% of GDP in 2010 (USA in 2011: 17.9%). Peru has, in comparison to many other developing countries, a quite high literacy rate. In 2004, the literacy rate for youth (15-24 years) was 97% and for adults almost 90%. Education in Peru through primary and secondary school is free for children from ages 7 to 16, although in practice it is inaccessible to many rural children because of the scarce supply of schools outside cities. Children enter primary school at the age of six (grades 1 to 6). Secondary school consists of 5 grades (grades 7 to12). There are vocational education, private and public technical colleges and universities for higher education (Classbase, 2013).[9]

In our study, the information was taken from the child questionnaires (answered by the child) and household questionnaires (answered by the primary caregiver) of the YL survey. Only children from the older cohort who participated in all three rounds (2002, 2006/07 and 2009) of the YL survey were analyzed (n=664). We decided to exclude children not enrolled in school in round 2 (n=6) from the sample as it seems that their living conditions differ much from the rest of the sample leaving n=658 children accessible for our analysis.

[8.] Some papers have already made use of the YL study program data (see Helmers and Patnam (2011) or Lordan and Frijters (2013) as an example).

[9.] All information on sports and sport systems in Peru were provided by interviews and discussions with Virginia Rey-Sánchez (Directora de comunicaciones; Estudio Niños del Milenio) as well as Mariluz Aparicio (Asistente de comunicaciones; Estudio Niños del Milenio). If not marked numbers are taken from the World Factbook (2013). A summary of information on Peru is provided in the Appendix (Tables A.2 and A.3)

IV. Empirical strategy

In line with only a few previous papers investigating the causal effects of sports (club) participation (e.g. Lechner, 2009; Lechner and Downward, 2013; Downward, Hallmann and Pawlowski, 2014) an average treatment effects analysis is conducted. In this study, the 'treatment' variable is taken from round 3 (r3) of the YL survey in 2009 and measures whether a child generally participates in a sports group or not (as a dummy variable with a score of 1 for yes). We are interested in the average effects of sports group participation on a randomly drawn child from the population (i.e. Average Treatment Effects: ATE), the average effects for those children who participated in sports groups (i.e. Average Treatment Effects on the Treated: ATET) and the average effects for those children who did not participate in sports groups (i.e. Average Treatment Effects on the Non-Treated: ATENT) (see Wooldridge, 2010, 905f.). The outcomes for which these effects are investigated are also measured in round 3.

Outcome variables

Human capital is measured by the Peabody Picture Vocabulary Test (PPVT). This test analyses receptive vocabulary and provides a quick estimate of verbal ability and scholastic aptitude (Pearson, 2013). In the test, a word is spoken and the child must choose the correct one of four pictures. Furthermore, in line with Becker (1962), who states that investments in schooling, on-the job training or medical care enhance individual productivity and therefore employment prospects, school enrolment is included as indicator for human capital (as a dummy variable) (An overview on all outcome variables is provided in Table A.4 in the Appendix.).

Social capital is defined by Putnam as 'those features of social organization, such as trust, norms and networks that can improve the efficiency of society by facilitating coordinated actions' (Putnam, 1993, 167) and thereby 'enable participants to act together more effectively to pursue shared objectives' (Putnam, 1995, 664ff.). These networks can be broad and wide ranging (Downward, Pawlowski and Rasciute, 2014). Consequently, the feeling of safety can be included in the concept of social capital. Feeling safe can be established through social relationships. Having friends and being helpful as well as the presence of older pro-social role models contribute to a sense of safety (Gilgun, 1996; Caprara *et al.*, 2000). Feeling safe is

also influenced by public (neighborhood) safety. People do not feel safe when they are regularly surrounded by, or see, violence. Hence, feeling safe is strongly connected with the effectiveness of public safety in keeping violence in a community low (Overstreet and Braun, 2000). In line with this, social capital is measured by neighborhood trust, friends support and respect by other children and feeling safe outside the house (safe environment) which are assessed on a scale 1 (strongly disagree) to 5 (strongly agree). In addition to the original (categorical) variables, dummy variables are deployed for these outcome variables with disagreement (1 to 3) as reference category. Finally, with respect to social capital, as indicated above it is interesting to investigate if sports group participation is a substitute or complement to *other group participation*. For example, it may act to bond similar individuals together, thus reducing links to others, or bridge differences between individuals. The latter is often cited as important for producing social cohesion, but bonding capital can be also important. For instance, Seefeldt and Erwing (1997) argue that sport participation is a substitute to gang behavior as adolescents join gangs because they provide an identity for its member and serve many functions that a family might do. Here other group participation is measured by dummy variables of memberships in after school, religious, school and youth groups.

Subjective well-being can be measured as the emotional response and global judgment of life satisfaction, happiness, eudemonia, or good psychological functioning (e.g. Diener *et al.*, 1999; Hornung, 2006). Although there are differences in the definitions of these terms, life-satisfaction and happiness are often used as synonyms. In this study, subjective well-being is measured as satisfaction with life where children should state their current position on a ladder (step 1 to 9). High steps denote high satisfaction with life. The ladder, as a representation of well-being, was first developed by Cantril (1965). It is known as the Cantril scale or Cantril ladder.

Health capital is measured as subjective health status and Body Mass Index (BMI). The question on the health status has to be answered by the children on a 5-point scale (1=very poor to 5=very good). The BMI is calculated by dividing weight by squared height.

Table 1 provides an overview of the outcome variables available together with the differences in the outcomes between the treatment and control groups based on simple t-tests of mean differences. Results indicate significant differences with regard to the average achievements in the vocabulary test (negative), the portion of children feeling that their friends will stand by

them during difficult times as well as the children's subjective well-being and health status (all positive). However, just comparing the outcomes of children participating with those not participating in a sports group might yield biased results if those two groups differ with respect to other characteristics that also influence the outcome variables (Imbens, 2004). To *identify* the effects of sports group participation, it is necessary that three central assumptions hold.

TABLE 1

Mean values for the outcome variables and t-test significances

	SG member in r3	*Not SG member in r3*	*t-test*
Human capital			
Vocabulary test (Peabody Picture Vocabulary Test score: 0-200)	94.03	98.36	**-2.4****
School enrolment (share of enrolled)	94%	93%	0.5
Social capital			
Neighbourhood trust (share of "agree" & "strongly agree")	32%	31%	0.2
Friends support (share of "agree" & "strongly agree")	69%	58%	**2.1****
Respect by children (share of "agree" & "strongly agree")	79%	82%	-0.7
Safe environment (share of "agree" & "strongly agree")	51%	44%	1.4
Subjective well-being			
Satisfaction with life (Cantril ladder: 1=dissatisfied to 9=satisfied)	6.50	6.04	**2.7*****
Health capital			
Health status (mean of 1-very poor...5-very good)	3.91	3.72	**2.8*****
Body Mass Index (weight / squared height)	21.05	21.10	-0.2
Other group participation			
After school club (share of members)	3%	6%	-1.6
Religious group (share of members)	5%	9%	-1.4
School club (share of members)	2%	2%	-0.2
Youth group (share of members)	4%	3%	-0.1
Observations	118	540	

Notes: Significance levels are indicated as *** ≡ p<0.01; ** ≡ p<0.05, * ≡ p<0.1; Significant effects are displayed in bold letters.

Identification

The first assumption is known as the no confounding or conditional independence assumption (CIA), the second is known as the stable unit treatment value assumption (STUVA)[10] and the third is known as common support or overlap assumption (OVERLAP)[11] (see Imbens, 2004).

[10.] STUVA demands that potential outcomes are unaffected by the particular assignment of the treatment. This condition requires that there are no unrepresented treatments (everybody is either 0 or 1) and that there are no relevant interactions between treatments (the fact that child *i* participates does not change the potential outcome of child *j*). We think that this assumption holds in our context at least with regard to an individual's perception of *their* experience. However, some spillovers could be latently present.

CIA implies that potential outcomes are conditionally independent of the treatment (*here*: sports group participation) for given values of any confounding variables X that might influence both. As noted in Section II, since previous studies confirm that selection into sports is not random, it is likely that a selection bias also exists in the context of our study. Therefore, given our research design we need to control for variables jointly determining sports group participation and the outcome variables. These confounding variables are used to estimate the conditional choice probabilities of being a sports group member, i.e. the propensity score (see Rosenbaum and Rubin, 1983), before matching similar individuals that just differ with respect to the treatment who can thus form the basis of a causal analysis.

In general, finding relevant and exogenous[12] confounders is not straightforward. Therefore, our choice of confounding variables is based on theoretical considerations and previous empirical findings related to sports (club) participation in developed countries since there is no study available, which focusses on sports participation in less developed countries. It is important to stress that to deal with the issue of possible endogeneity we use variables primarily from the period prior to the treatment and the outcomes, i.e. from round 1 (r1) or round 2 (r2) (see Table A.5 in the Appendix).

The first block of variables are socio-demographic and child environmental characteristics with gender and site (urban/rural). While the importance to control for gender is well documented in the literature (e.g. Lechner, 2009 for adults or McHale *et al,*. 2001 for children), the variable 'site' is included here since it appears plausible that urban areas in less developed countries might have a better provision of sports facilities than rural areas, which may influence sports activity in general (e.g. Limstrand, 2008; Pawlowski *et al.*, 2009; Huang and Humphreys, 2012). To describe the background and environment of the children, variables characterizing household and parents are taken into consideration. The household is represented by its size. Studies have found that adult sports participation is reduced by having children in a family especially for females (e.g. Farrell and Shields, 2002; Downward and Riordan, 2007). Further results show that there is a relationship between the sports

[11.] By estimating treatment effects the probability of the treatment (denoted by w) assignment given covariates, X, (propensity score $p(x) \coloneqq P(w = 1|X = x)$) is important. The overlap (or common support) assumption forecloses the possibility that $p(x)$ is ever zero or one which means that the perfect predictability of treatment given a set of confounding variables X is ruled out (Wooldridge, 2010). This ensures that for any given value of the confounding variables, a unit could potentially be observed with $w = 1$ or $w = 0$, $0 < p(x) < 1$. As indicated by our results we generally deal with a set of observations with appropriate matches.

[12.] For instance, while sports group participation might positively affect health status, children are in general more likely to be physically active (and organized in sports groups) if they do not suffer from health problems.

participation of children and their parents and siblings (e.g. Scheerder *et al.*, 2007; Downward, Hallmann and Pawlowski, 2014). In addition, the household situation in terms of wealth compared to other households (measured as a dummy variable, with 'richer or average' as the reference category) and a wealth index are considered. The latter measures the socioeconomic status of a household and incorporates housing quality (e.g. rooms per person), consumer durables and services (e.g. electricity or water) (Kumra, 2008). Previous studies showed that higher socio-economic status and higher incomes raise sports activity (e.g. Downward and Rasciute, 2010). Furthermore, the relationship of primary caregiver and child, as well as the educational level of the primary caregiver are included in the list of variables (completed secondary education). Finally, time-use variables (leisure, chore, study, sleep) are included measuring the hours spent for a certain activity per day.

In addition to these variables, available lagged variables of all five outcome categories are used as they might also capture time constant confounders ('fixed effects') that are otherwise unobservable. Unfortunately, some of these variables are either measured differently compared to the outcome variables from round 3 or not available in round 2. *Human capital* is measured by school enrolment (r1), work (r2), reading and writing (r1), numeracy skills (r1) and 'scoremath' (r2). Reading, writing and numeracy are included as dummy variable whereas the reference categories are the ability to read/write/calculate. 'Scoremath' is measured by a test where children have to solve ten arithmetic problems (scores between 0 and 10 can be achieved). Furthermore, neighbourhood trust, respect by children and safe environment are included as lagged *social capital* measures. The variable 'friends support' is not available in round 2. Subjective health status was measured as 'compared to other children' and a dummy for long-term illness was included with the reference category no long-term illness to measure *health capital*. Further, body mass index from round 1 is included as well as height which is taken from round 3 (at the age of 15) as it is assumed, as with the literature in Section II, that height is not affected by sports activity of children, at least in the short run. However, height can have an effect on the physical activity of children. Finally, dummy variables for *group* memberships (after school-, religious-, and informal child group) are also used as confounding variables.

In summary, the aim of using this rich set of variables is to try and to control for some of the most important confounding effects between sports participation and outcomes. It is acknowledged, however, that there may well be further confounding factors that cannot be

controlled for but which might jointly influence sports participation and outcomes. These might include genetics, psychological, cognitive, and other emotional factors or behavioral attributes associated with factors such as diet that unfortunately cannot be taken into account. Therefore, the degree to which the estimated effects have a causal interpretation has limits. Nevertheless, in as such that these factors will have already influenced past outcomes, and have an effect that is stable over time, similar to a fixed effects, their effect will be accounted for by the inclusion of lagged outcome variables. As a consequence, we argue that the use of the confounding variables described above are sufficient to reveal the best evidence possible from this data to examine a causal link between sports group participation and the outcome measures of interest.

Estimation

Average treatment effects are estimated by radius matching based on the propensity score. This method enables comparison of the different outcomes of children being a member of a sports group (treated) with the outcomes of children who are not a member of a sports group. To avoid selection bias, the two groups are adjusted in their covariate composition (as described in the section before) based on the propensity to be member of a sports group, which is estimated using a probit model.[13]

Specifically, we apply an estimator that takes into account the methodological considerations of Lechner *et al.* (2011). Compared to standard nearest-neighbor matching this procedure is more precise because it incorporates the idea of radius matching (e.g. Dehejia and Wahba, 2002). Furthermore, the algorithm uses the initial matching weights in a second step of (weighted) regression adjustment, which has two advantages. Firstly, the estimator satisfies a so-called double robustness property: the estimator remains consistent if either the matching step is based on a correctly specified selection model, or the regression model is correctly specified (e.g. Rubin, 1979; Joffe *et al.,* 2004). Secondly, the regression adjustment should reduce small sample as well as asymptotic biases of matching (see Abadie and Imbens, 2008). Huber, Lechner, and Wunsch (2013) investigate the finite sample properties of this radius

[13.] For estimation, Bin_Match_4.0.1 is used, which is the Gauss version of the radius match command in STATA (see Huber *et al.,* 2014, for an investigation on the sensitivity of the estimator with respect to different values of various tuning parameters). It is used on Gauss Version 13. The Gauss program can be downloaded from www.michael-lechner.eu/software, while Stata and R versions are available from the respective repositories.

matching with bias adjustment algorithm along with many other matching type estimators and find it to be most highly ranked.

Concerning inferences, Abadie and Imbens (2008) show that for standard matching (i.e. based on a fixed number of comparison observations) bootstrap-based inference may be invalid. However, the matching algorithm used in our analysis is smoother than the one studied by Abadie and Imbens (2008) because it is based on a variable number of comparisons and uses the regression adjustment. For this reason, the bootstrap is likely to be a valid inference procedure in our context. To take account of the panel nature of the data, a block bootstrap is used in which all observations from the same individual over time form the block. To be more precise, inference is based on (i) bootstrapping programs 4,999 times; (ii) computing the bootstrap t-statistics of the respective average effects in each of the samples (normalized by the estimated effect); and (iii) estimating the p-value as the share of absolute bootstrap t-statistics that are larger than the absolute t-statistic in the original sample (see for instance MacKinnon (2006) for a discussion on bootstrapping symmetric statistics). This statistic is smoothed as suggested by Racine and MacKinnon (2007). Since the theoretical results by Abadie and Imbens (2008) and the simulation based results in Huber *et al.* (2013) suggest that the estimator is asymptotically normally distributed, bootstrapping the potentially pivotal t-statistic (computed under the assumption that the weights obtained to compute the control group are non-stochastic; see Lechner, 2002) has the advantage of potentially providing so-called asymptotic refinements and thus improving inference. In addition, we also checked the bootstrap distribution of the estimated effects directly (quantile method). The results are very similar (available on request).

V. Results

Since just comparing the outcomes of children participating with those not participating in a sports group might yield biased results if those two groups differ with respect to other characteristics that also influence the outcome variables, results such as simple comparisons presented in Table 1 will be misleading and must be treated with caution. In this section we present the estimation results based on the identification strategy as laid out above to identify, as much as possible, causal effects. We start with discussing the estimates of the probit model

for the variables of the selection process into sports group participation before providing the matching results, i.e. ATE, ATET and ATENT.

For ease of exposition, Table 2 provides an (abbreviated) overview of the mean values by treatment state and the selection-into-sports-group probit model. The complete version as well as the corresponding values for a second specification including interaction terms (Model 2) is displayed in Table A.6 in the Appendix. Model 2 serves as robustness check by taking possible interaction effects of sports group participation prior to the treatment and selected confounding variables into account. This more flexible specification may, consequently, lead to some near multicollinearity, and thus to a larger variability of the results.

There are slightly less males than females in the sports group sample (49%) compared to the non-sports group sample (54%). However, in line with the balance of gender, there is no association of being male with the probability to participate in a sports group. Interestingly, 42% of the sports group sample lives in urban sites whereas the figure for those who don't participate in a sports group is much higher (64%). According to the probit analysis there is a significant and negative association between living in urban areas and sports group participation. Furthermore, the percentage of households, which assess themselves as poorer compared to other households, is slightly higher in the sample of non-sports group members. Correspondingly a poorer household situation is significant and negatively associated with the probability of participating in a sports group. However, the objective measure of wealth (by the wealth index) shows no significant association.

In addition, being in a sports group in round 2 is highly significant and positively associated with being in a sports group in round 3. This is perhaps not surprising as 58% of those who are in a sports group in round 3 have already been in a sports group in round 2. With regard to time allocation, there are significant positive associations of leisure, study and chore time with the probability of participation in a sports group in round 3, as well as having done paid work in the last 12 months. Finally, a positive association of being in a religious group with participating in a sports group can be observed.

The results of Model 2 with the more flexible specification according to the interaction terms are broadly in line with the results of model 1 (Table A.6 in the Appendix). The main

difference is the highly significant negative (instead of positive) impact of sports group participation in round 2 which is ascribed to the incorporation of interaction terms.

TABLE 2

Descriptive statistics and selected average marginal effects of the probit model for the variables of the selection process into sports group participation

	Mean in subsamples		*Marginal effects*
	SG member in r3	*not SG member in r3*	*dependent variable: SG member in r3*
Socio-demographics and child environmental characteristics			
Male	0.49	0.54	-0.044 (0.23)
Household size	5.61	5.53	0.003 (0.737)
Household situation (r1)	0.24	0.27	**-0.063*** (0.054)
Wealth index (r1)	0.42	0.47	0.101 (0.478)
Wealth index	0.45	0.54	-0.172 (0.20)
Relationship caregiver (r1)	0.07	0.07	-0.023 (0.667)
Caregiver higher edu	0.47	0.55	0.013 (0.725)
Site	0.42	0.64	**-0.084*** (0.087)
Time allocation			
Sports group (r2)	0.58	0.38	**0.099***** (0.003)
Leisure time	2.42	2.57	**0.046*** (0.092)
Chore time	4.71	4.33	**0.051**** (0.041)
Study time	7.62	7.84	**0.056**** (0.016)
Sleeping time	9.35	9.33	-0.007 (0.618)
Human capital			
Work	0.34	0.26	**0.081**** (0.048)
Reading (r1)	0.09	0.09	-0.049 (0.359)
Writing (r1)	0.15	0.12	0.039 (0.472)
Numeracy	0.20	0.18	0.035 (0.396)
Score math	5.45	5.86	-0.007 (0.535)
Social capital			
Neighbourhood trust	0.56	0.56	-0.025 (0.431)
Respect by children	0.91	0.91	-0.002 (0.967)
Safe environment	0.65	0.57	0.047 (0.136)
Subjective well-being			
Satisfaction with life	5.99	6.00	-0.004 (0.566)
Health capital			
Health status	0.62	0.63	-0.006 (0.834)
Longterm illness	0.13	0.11	0.004 (0.923)
BMI (r1)	16.61	16.83	-0.004 (0.706)
Height (r3)	153.52	154.70	0.001 (0.681)
Other group participation			
After school club	0.13	0.08	0.008 (0.876)
Religious group	0.34	0.21	**0.081**** (0.035)
Informal child group	0.18	0.12	0.055 (0.270)
Observations / Efron's R^2	118	540	8.4

Notes: Significance levels are indicated as *** ≡ p<0.01; ** ≡ p<0.05; * ≡ p<0.1. Significant effects are displayed in bold letters; Average marginal effects are displayed in the last column derived from a probit analysis. Bootstrapped p-values are in brackets. The model includes a constant term and further covariates. The table with all variables is available in Appendix A.6. Number of bootstrap replications: 4.999. If not mentioned otherwise, all covariates are measured in r2.

Table 3 provides the matching results, i.e. ATE, ATET and ATENT. Since common support is given for more than 98% (respectively 94% in Model 2), the overlap assumption is not of concern. In line with the results of the simple t-tests of mean differences in the subsamples (Table 1), we could detect a significant *effect* of sports group participation on subjective health status and social capital (through the perceived support by friends in difficult times). Significantly, however, these results suggest that unlike in developed countries sports group participation may have much less relevance to human capital outcomes through education and consequently the labour market, and also subjective well-being.

TABLE 3

Radius matching results of sports group participation

	ATE	*ATET*	*ATENT*
Human capital			
Vocabulary test	-0.617 (0.727)	-3.419 (0.441)	-0.009 (0.989)
School enrolment	0.030 (0.114)	0.024 (0.564)	0.031 (0.122)
Social capital			
Neighborhood trust (d)	-0.008 (0.906)	0.023 (0.802)	-0.015 (0.825)
Friends support (d)	**0.121** (0.044)**	**0.136* (0.061)**	**0.118* (0.083)**
Respect by children (d)	-0.008 (0.858)	0.007 (0.907)	-0.011 (0.821)
Safe environment (d)	0.048 (0.451)	0.055 (0.441)	0.047 (0.516)
Subjective well-being			
Satisfaction with life	0.287 (0.168)	0.413 (0.128)	0.259 (0.248)
Health capital			
Health status	**0.303*** (0.001)**	**0.246** (0.018)**	**0.315*** (0.004)**
Body Mass Index	0.279 (0.381)	0.031 (0.939)	0.333 (0.384)
Other group participation			
After school club	-0.028 (0.457)	-0.040 (0.224)	-0.026 (0.565)
Religious group	-0.056 (0.100)	-0.108 (0.193)	-0.045 (0.266)
School club	0.017 (0.496)	-0.063 (0.545)	0.034 (0.357)
Youth group	-0.014 (0.293)	0.015 (0.568)	-0.020 (0.156)
No. of observations: 658			
No. of treated: 118 (17.93% of observations)			
Common support: 650 (Share in common support: 98.78%)			

Notes: Significance levels are indicated as *** ≡ p<0.01; ** ≡ p<0.05; * ≡ p<0.1. Significant effects are displayed in bold letters. p-values (in brackets) obtained from 4.999 bootstrap replications as the share of absolute bootstrap t-statistics that are larger than the absolute t-statistic in the original sample. The results are based on a version, which uses weighted logits instead of weighted OLS for the binary outcomes (weights used for inference take bias adjustment not into account). The full set of results is available in Appendix A.7.

VI. Conclusions

The few existing studies exploring the outcomes of sports participation on children have exclusively focused on developed countries. This paper contributes to the literature as it is the first to generally explore the possible outcomes of sports (group) participation in a less

developed country – Peru. Furthermore – and in contrast to most of the existing studies analyzing the factors associated with sports participation of adults and children – the identification strategy employed suggest that the estimated associations are (at least close to) having a causal interpretation.

The literature for the US or some European countries finds a broadly positive impact of sport on human capital, social capital, well-being and health. This is different in our study setting. Our results suggest that sports participation has positive impacts on children's subjective health and social capital. However, and in contrast to developed countries, we could not detect any statistically significant effects on well-being and human capital formation. Nonetheless, these results suggest that sport can influence two of the key factors that are identified as central to children's development in Bornstein *et al.* (2012) and, importantly, establish this in a population based study.

Overall, our results suggest that some of the aspirations of policy organizations that are noted in the introduction can be met through sport and consequently provides support for the use of sport to reach such policy targets. However, testing the dependence of the effects of sports group participation on the quality and format of the sports organization as well as the specificity of a country is an issue for future research.

References

Abadie, A. and Imbens, G. W. (2008). 'On the Failure of the Bootstrap for Matching Estimators', *Econometrica*, Vol. 76, pp. 1537–1557.

Alderman, H. (2010). 'The economic cost of a poor start to life', *Journal of Developmental Origins of Health and Disease*, Vol. 1, pp. 19–25.

Barron, J. M., Ewing, B. T., and Wadell, G. R. (2000). 'The effects of high school athletic participation on education and labor market outcomes', *Review of Economics and Statistics*, Vol. 82, pp. 409–421.

Becker, G. S. (1962). 'Investment in Human Capital: A Theoretical Analysis', *Journal of Political Economy*, Vol. 70, pp. 9–49.

Bornstein, M. C, Rebello Britto, P., Nonoyama-Tarumi, Y., Ota, Y., Petrovic, O., and Putnick, D. L. (2012). 'Child development in developing countries: Introduction and Methods', *Child Development*, Vol. 83, pp. 16–31.

Breuer, C., Hallmann, K., and Wicker, P. (2011). 'Determinants of sport participation in different sports, *Managing Leisure*', Vol. 16, pp. 269–286.

Caballero, B. (2005). 'A nutrition paradox - underweight and obesity in developing countries', *The New England journal of medicine*, Vol. 352, pp. 1514–1516.

Cabane, C. and Lechner, M. (2014). 'Physical activity of adults: a survey of correlates, determinants, and effects', *Discussion paper* No. 2014-28, University of St. Gallen, Department of Economics.

Cantril, H. (1965). '*The pattern of human concerns*', New Brunswick, Rutgers University Press.

Caprara, G., Barbanell, C., Pastorelli, C., Bandura, A., and Zimbardo, P. (2000). 'Prosocial foundations of children's academic achievement', *Psychological Science*, Vol. 11, pp. 302–306.

Classbase (2013). Education System in Peru, http://www.classbase.com/Countries/Peru/Education-System.

Crookston, B. T., Dearden, K. A., Alder, S. C., Porucznik, C. A., Stanford, J. B., Merrill, R. M., Dickerson, T. T., and Penny, M. E. (2011). 'Impact of early and concurrent stunting on cognition', *Maternal and Child Nutrition*, Vol. 7, pp. 397–409.

Dehejia, R. H. and Wahba, S. (2002). 'Propensity-score-matching methods for nonexperimental causal studies', *The Review of Economics and Statistics*, Vol. 84, pp. 151–161.

Diener, E., Suh, E. M., Lucas, R. E. and Smith H. L. (1999). 'Subjective Well-Being: Three Decades of Progress', *Psychological Bulletin*, Vol. 125, pp. 276–302.

Downward, P. and Rasciute, S. (2010). 'The Relative Demands for Sports and Leisure in England', *European Sport Management Quarterly*, Vol. 10, pp. 189–214.

Downward, P. and Riordan, J. (2007). 'Social Interactions and the Demand for Sport: an Economic Analysis', *Contemporary Economic Policy*, Vol. 25, pp. 518–537.

Downward, P., Lera-Lopez, F., and Rasciute, S. (2012). 'The economic analysis of sport participation', in L. Robinson, P. Chelladurai, G. Bodet and P. Downward (eds.), *Routledge Handbook of Sport Management*, London, Routledge.

Downward, P., Hallmann, K., and Pawlowski, T. (2014). 'Assessing parental impact on the sports participation of children: A socio-economic analysis of the UK', *European Journal of Sport Science*, Vol. 14, pp. 84–90.

Downward, P., Pawlowski, T., and Rasciute, S. (2014). 'Does associational behaviour raise social capital? A cross-country analysis of trust', *Eastern Economic Journal*, Vol. 40, pp. 150– 165.

Eccles, J. S., Barber, B. L., Stone, M., and Hunt, J. (2003). 'Extracurricular Activities and Adolescent Development', *Journal of Social Issues*, Vol. 59, pp. 865–889.

Eide, E. R. and Ronan, N. (2001). 'Is participation in high school athletics an investment or a consumption good? Evidence from high school and beyond', *Economics of Education Review*, Vol. 20, 431–442.

Escobal, J., Lanata, C., Madrid, S., Penny, M., Saavedra, J., Suarez, P., Verastegui, H., Villar, E., and Huttly, S. (2003). '*Young Lives Preliminary Country Report: Peru*', http://r4d.dfid.gov.uk/PDF/Outputs/YoungLives/PeruPreliminaryreport.pdf.

Farrell, L. and Shields, 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)*, Vol. 165, pp. 335–348.

Felfe, C., Lechner, M., and Steinmayr, A. (2011). 'Sports and child development', *Discussion Paper* No. 6105, Forschungsinstitut zur Zukunft der Arbeit.

Gerlach, E. and Brettschneider, W.-D. (2013). '*Aufwachsen im Sport. Befunde einer 10-jährigen Längsschnittstudie zwischen Kindheit und Adoleszenz*' (Sportentwicklungen in Deutschland Band 23), Aachen, Meyer & Meyer.

Gilgun, J. (1996). 'Human development and adversity in ecological perspective, part 2: Three patterns', *Families in society*, Vol. 77, pp. 459–476.

Gomez-Pinilla, F. (2008). 'The influences of diet and exercise on mental health through hormensis', *Aging Research Review, Vol.* 7, pp. 49–62.

Grantham-McGregor, S. (1995). 'A review of studies of the effect of severe malnutrition on mental development', *The Journal of Nutrition*, Vol. 125, pp. 2233S–2238S.

Grantham-McGregor, S. (2002). 'Linear growth retardation and cognition', *Lancet*, Vol. 359, p. 542.

Joffe, M. M., Ten Have, T. R., Feldman, H., and Kimmel, S. E. (2004). 'Model Selection, Confounder Control, and Marginal Structural Models: Review and New Applications', *The American Statistician*, Vol. 58, pp. 272–279.

Helmers, C. and Patnam, M. (2011). 'The formation and evolution of childhood skill acquisition: Evidence from India', *Journal of Development Economics,* Vol. 95, pp. 252–266.

Hornung, B. R. (2006). 'Happiness and the pursuit of happiness: A sociocybernetic approach', *Kybernetes*, Vol. 35, pp. 323–346.

Huang, H. and Humphreys, B. (2012). 'Sports Participation and Happiness: Evidence from U.S. Micro Data', *Journal of Economic Psychology*, Vol. 33, pp. 776–793.

Huber, M., Lechner, M., and Steinmayr, A. (2014). 'Radius matching on the propensity score with bias adjustment: finite sample behaviour, tuning parameters and software implementation', Forthcoming in *Empirical Economics*.

Huber, M., Lechner, M., and Wunsch, C. (2013). 'The performance of estimators based on the propensity score', *Journal of Econometrics*, Vol. 175, pp. 1–21.

Imbens, G. W. (2004). 'Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review', *The Review of Economics and Statistics*, Vol. 86, pp. 4–29.

Kumra, N. (2008). 'An Assessment of the Young Lives Sampling Approach in Andhra Pradesh, India', *Technical Note No. 2*.

Lechner, M. (2002). 'Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programmes by Matching Methods', *Journal of the Royal Statistical Society, Series A*, Vol. 165, pp. 59–82.

Lechner, M. (2009). 'Long-run labour market and health effects of individual sports activities', *The Journal of Health Economics,* Vol. 28, pp. 839–854.

Lechner, M. and Downward, P. (2013). 'Heterogeneous Sports Participation and Labour Market Outcomes in England', *IZA Discussion Paper* No. 7690.

Lechner, M., Miquel, R., and Wunsch, C. (2011). 'Long-Run Effects of Public Sector Sponsored Training in West Germany', *Journal of the European Economic Association*, Vol. 9, pp. 742–784.

Limstrand, T. (2008). 'Environmental Characteristics Relevant to Young People's use of Sports Facilities: a Review', *Scandinavian Journal of Medicine & Science*, Vol. 18, pp. 275–287.
Lipscomb, S. (2007). 'Secondary school extracurricular involvement and academic achievement: a fixed effects approach', *Economics of Education Review*, Vol. 26, pp. 463–472.
Lordan, G. and Frijters, P. (2013). 'Unplanned pregnancy and the impact on sibling health outcomes', *Health Economics*, Vol. 22, pp. 903–914.
MacKinnon, J. G. (2006). 'Bootstrap Methods in Econometrics', *The Economic Record*, Vol. 82, pp. S2–18.
McHale, S., Crouter, A. C., and Tucker, C. J. (2001). 'Free-time activities in middle childhood: links with adjustment in early adolescence', *Child Development*, Vol. 72, pp. 1764–1778.
Overstreet, S. and Braun, S. (2000). 'Exposure to community violence and post-traumatic stress symptoms: Mediating factors', *American Journal of Orthopsychiatry*, Vol. 70, pp. 263–271.
Pawlowski, T., Breuer, C., Wicker, P., and Poupaux, S. (2009). 'Travel Time Spending Behaviour in Recreational Sports: An Econometric Approach with Management Implication', *European Sport Management Quarterly*, Vol. 9, pp. 215–242.
Pearson (2013). Peabody Picture Vocabulary Test, Fourth Edition, http://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm? Pid=PAa30700.
Pfeifer, C. and Cornelißen, T. (2010). 'The impact of participation in sports on educational attainment – New evidence from Germany', *Economics of Education Review*, Vol. 29, pp. 94–103.
Putnam, R. (1993). *Making democracy work: Civic traditions in modern Italy*, Princeton University Press, Princeton, NJ.
Putnam, R. (1995). 'Tuning in, tuning out: The strange disappearance of social capital in America', *Political Science and Politics*, Vol. 28, pp. 664–683.
Racine, J. S. and MacKinnon, J. G. (2007). 'Inference via kernel smoothing of bootstrap P values', *Computational Statistics & Data Analysis*, Vol. 51, pp. 5949–5957.
Rosenbaum, P. and Rubin, D. (1983). 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, Vol. 70, pp. 41–55.
Rubin, D. B. (1979). 'Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies', *Journal of the American Statistical Association*, Vol. 74, pp. 318–328.
Scheerder, J., Taks, M., and Lagae, W. (2007). 'Teenage girls' participation in sports. An integrational analysis of socio-cultural predictor variables', *European Journal for Sport and Society*, Vol. 4, pp. 133–150.
Seefeldt, V. D. and Ewing, M. E. (1997). 'Youth Sports in America: An Overview', *President's Council on Fitness and Sport: Research Digest*, Vol. 2, pp. 1–14.
Stevenson, B. (2010). 'Beyond the Classroom: Using Title IX to measure the return to High School sport', *The Review of Economics and Statistics*, Vol. 92, pp. 284–301.
Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Ritcher, L. *et al*. (2008). 'Maternal and child undernutrition: consequences for adult health and human capital', *Lancet*, Vol. 371, pp. 340–357.
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd Ed.), MIT Press, Cambridge, MA.
World Factbook (2013). https://www.cia.gov/library/publications/the-world-factbook/
Young Lives (2013). What we do, http://www.younglives.org.uk/what-we-do.
Young Lives Method Guide (2011). Sampling, http://www.younglives.org.uk/files/methods-guide/methods-guide-sampling.

Appendix

TABLE A.1

Large-scale empirical studies exploring possible effects of child and youth sports participation on their human capital, social capital, subjective well-being and health

Study	*Country and Data*	*Methods*	*Results*
Barron, Ewing and Wadell (2000)	USA, National Longitudinal Survey of Youth (NLSY; n=1,047) and of High School Class (NLS-72; n=3,014)	Two-period model of time allocation (probit model)	*Human capital:* Athletic participation directly affects wages and educational attainment (EA). But effects appear to reflect differences across individuals in ability or value of leisure.
Eide and Ronan (2001)	USA, High School and Beyond data set (HSB) (n=11,577)	Ordinary least squares and instrumental variable	*Human capital:* Sports participation has a neg. effect on the EA of white male student athletes, a pos. effect on the EA and earnings of black male student athletes, and a pos. effect on the EA of white female student athletes.
Eccles, Barber, Stone and Hunt (2003)	Michigan, USA MSALT (n=1,259)	Longitudinal regression analysis	*Human capital:* Participation in most extracurricular activities leads to better educational outcomes by controlling for social class, gender and intellectual aptitude. Participation on school sports teams increased educational outcomes, but higher drinking rates, too.
Lipscomb (2007)	USA, National Education Longitudinal Study (NELS) (n=16,305)	Fixed effects strategy	*Human capital:* Athletic and club participation are associated with an increase in math test scores and an increase in Bachelor's degree attainment expectations.
Stevenson (2010)	National Center for Education Statistics (NCES) and Public Use Micro Sample (PUMS) (n varies per data set)	Instrumental variables	*Human capital:* Analysis across states in changes between pre- and post-legislation reveals that a rise in state sports participation for females generates an increase in female college attendance and in labour force.
Pfeifer and Cornelißen (2010)	GSOEP (n=6,050)	Generalized ordered probit models	*Human capital:* After controlling for important variables, a significant and positive effect of sport on educational attainment could be found.
Felfe, Lechner and Steinmayr (2011)	KiGGS, GCP and information on the available sports facilities (n=5,632)	Matching estimation strategy and instrumental variable (IV)	*Human capital, subjective well-being and health:* Sports club participation of children in Germany aged 3 to 10 years has a positive effect on skill development (cognitive and non-cognitive), general well-being and health.
Gerlach and Brettschneider (2013)	Germany, Longitudinal Study of children in a region in Germany (Paderborn) (n=1,637)	Variance analysis	*Social capital, subjective well-being and health:* Sports club participation of children and adolescents has a significant positive effect on social acceptance and respect by peers and well-being. However, there is no significant positive effect on obesity.

TABLE A.2

General information on Peru

Location (population)	West of South America (29.8 million)
Population (urban/rural)	77% in urban areas (USA: 82%)
Poverty (% of pop. living below poverty line)	27.8% (USA: 15.1%)
Health	
Life expectancy (in years)	men: 71, women: 75 (USA: men: 77, women 81)
Infant mortality rate (per 1,000 live births)	21 deaths (USA: 6.2)
Health expenditure	5.1% of GDP (GDP: $ 292.2 billion in 2010) (USA: 17.9%)
Education	
Age of school enrolment	a) 6 years
Grades (1st and 2nd education)	b) 12 grades
Literacy rate	c) 90% and 97% for youth (15-24 years) (USA: 99%)

Notes: Available US figures are provided for a comparison.
Source: Numbers are taken from the World Factbook (2013).

TABLE A.3

Information on sports and sport systems in Peru

Popular sports	football
General organization of sports groups	mainly football clubs great demand for sports clubs from children; clubs have to be paid to be member; public parks with sports facilities (pay for access)
Aim of sports groups	practicing, preparing for competitions, playing sports in general
Degree of organization in sports groups	There are coaches in the sports groups organized in clubs.

Source (contact persons) : Virginia Rey-Sánchez (Directora de comunicaciones) and Mariluz Aparicio (Asistente de comunicaciones), Estudio Niños del Milenio

TABLE A.4

Description of outcome variables (taken from round 3)

Variable	*Label*	*Scale*
Human capital		
Vocabulary test	Peabody Picture Vocabulary Test (PPVT)	Metric: achieved points in the test
School enrolment	Is child currently in school?	*dummy:* 0 – no, 1 – yes
Social capital		
Neighbourhood trust	Most people in my neighbourhood can be trusted.	1 – strongly disagree; 2 – disagree; 3 – more or less; 4 – agree; 5 – strongly agree
Neighbourhood trust (d)	Most people in my neighbourhood can be trusted.	*dummy:* 0 – strongly disagree – more or less, 1 – agree & strongly agree
Respect by children	Other children in my class treat me with respect.	1 – strongly disagree; 2 – disagree; 3 – more or less; 4 – agree; 5 – strongly agree
Respect by children (d)	Other children in my class treat me with respect.	*dummy:* 0 – strongly disagree – more or less, 1 – agree & strongly agree
Safe environment	I feel save when I go out of the house on my own.	1 – strongly disagree; 2 – disagree; 3 – more or less; 4 – agree; 5 – strongly agree
Safe environment (d)	I feel save when I go out of the house on my own.	*dummy:* 0 – strongly disagree – more or less, 1 – agree & strongly agree
Friends support	My friends will stand by me during difficult times.	1 – strongly disagree; 2 – disagree; 3 – more or less; 4 – agree; 5 – strongly agree
Friends support (d)	My friends will stand by me during difficult times.	*dummy:* 0 – strongly disagree – more or less, 1 – agree & strongly agree
Subjective well-being		
Satisfaction with life	Current position on ladder	Cantril ladder: 1=dissatisfied to 9=satisfied
Health capital		
Health status	In general is your health good or poor?	1=very poor, 2=poor, 3=average, 4=good, 5=very good
Body mass index (BMI)	Calculated BMI=weight / squared(height)	Value of the BMI
Other group participation		
After school club	Member of an after school club	*dummy:* 0 – no, 1 – yes
Religious group	Member of a religious group	*dummy:* 0 – no, 1 – yes
School club	Member of a school club	*dummy:* 0 – no, 1 – yes
Youth group	Member of a youth group	*dummy:* 0 – no, 1 – yes

TABLE A.5

Description of confounding variables (taken from round 1 and 2; only "height" taken from round 3)

Variables	*Label*	*Scale*
Socio-demographics and child environmental characteristics		
Male*	Gender of the child.	*dummy:* 0 – female, 1 – male
Site*	Site where the child lives. (r2)	*dummy:* 0 – rural, 1 – urban
Household size*	Number of household members. (r2)	*metric*
Household situation	Household situation (wealth) compared to others. (r1)	*dummy:* 0 – richer or average, 1 – poorer
Wealth index (wi)*	Measures the socioeconomic status of households. (r1, r2)	*metric* (value between 0 and 1)
Relationship caregiver	Relationship of caregiver and child. (r1)	*dummy:* 0 – mother is primary caregiver, 1 – other
Caregiver higher edu*	Caregiver completed secondary or post-secondary school or university. (r2)	*dummy:* 0 – did not complete, 1 – completed
Time allocation		
Sports group	Sports group membership in r2.	*dummy:* 0 – no, 1 – yes
Leisure time*	Time spent for general leisure on a typical day. (r2)	*metric*
Chore time	Time spent for caring for others, domestic tasks, and family farm or business on a typical day. (r2)	*metric*
Study time*	Time spent at school and for studying outside school on a typical day. (r2)	*metric*
Sleeping time*	Time spent for sleeping on a typical night. (r2)	*metric*
Human capital		
Work	Has child done paid work in last 12 months? (r2)	*dummy:* 0 – no, 1 – yes
Reading	Child can't read letters, words or sentences. (r1)	*dummy:* 0 – can read, 1 – can't read
Writing	Child can't write sentences. (r1)	*dummy:* 0 – can write, 1 – can't write
Numeracy	Answer to the calculation 2 times 4. (r1)	*dummy:* 0 – can calculate; 1 – can't calculate
Scoremath*	Mathematic achievement test. (r2)	*metric* (scores between 0 and 10)
Social capital		
Neighbourhood trust*	Most people in my neighbourhood can be trusted. (r2)	*dummy:* 0 – strongly disagree – more or less, 1 – agree & strongly agree
Respect by children	Other children in my class treat me with respect. (r2)	
Safe environment	I feel safe when I go out of the house on my own. (r2)	
Subjective well-being		
Satisfaction with life*	Current position on ladder. (r2)	Cantril ladder: 1 - dissatisfied to 9 - satisfied
Health capital		
Health status*	Health compared to other children. (r2)	*dummy:* 0 – similar or better, 1 – worse
Long-term illness*	Child has long time health problems. (r2)	*dummy:* 0 – no, 1 – yes
BMI (Body Mass Index)	Weight / squared height (r1)	*metric*
Height	Height of the child (r3)	
Other group participation		
After school club*	Member of an after school club (r2)	*dummy:* 0 – no, 1 – yes
Religious group	Member of a religious group (r2)	*dummy:* 0 – no, 1 – yes
Informal child group	Member of an informal child group (r2)	*dummy:* 0 – no, 1 – yes

Notes: * An interaction term with the variable "sports group" is included in model 2.

TABLE A.6

Descriptive statistics and average marginal effects of the probit models for the variables of the selection process into sports group participation (model 1 and 2)

Variables	*Mean in subsamples*		*Marginal effects Model 1*	*Marginal effects Model 2*
	SG member in r3	*not SG member in r3*	*dependent variable: SG member*	*dependent variable: SG member*
Socio-demographics and child environmental characteristics				
Male	0.49	0.54	-0.0442 (0.23)	-0.037 (0.463)
Male x SG	0.32	0.23	-	0.002 (0.981)
Household size	5.61	5.53	0.003 (0.737)	-0.002 (0.842)
Household size x SG	3.33	2.14	-	0.009 (0.532)
Household situation (r1)	0.24	0.27	**-0.063*** (0.054)	**-0.07*** (0.064)
Wealth index (r1)	0.42	0.47	0.101 (0.478)	0.111 (0.445)
Wealth index	0.45	0.54	-0.172 (0.20)	-0.296 (0.125)
Wealth index x SG	0.27	0.19	-	0.214 (0.342)
Relationship caregiver (r1)	0.07	0.07	-0.023 (0.667)	-0.018 (0.744)
Caregiver higher edu	0.47	0.55	0.013 (0.725)	0.053 (0.286)
Caregiver higher edu x SG	0.26	0.20	-	-0.079 (0.221)
Site	0.42	0.64	**-0.084*** (0.087)	-0.105 (0.179)
Site x SG	0.25	0.21	-	0.061 (0.546)
Time allocation				
Sports group	0.58	0.38	**0.099***** (0.003)	**-0.365**** (0.033)
Leisure time	2.42	2.57	**0.046*** (0.092)	0.040 (0.202)
Leisure time x SG	1.45	0.98	-	0.003 (0.895)
Chore time	4.71	4.33	**0.051**** (0.041)	**0.050*** (0.081)
Study time	7.62	7.84	**0.056**** (0.016)	**0.055*** (0.060)
Study time x SG	4.36	2.88	-	-0.006 (0.761)
Sleeping time	9.35	9.33	-0.007 (0.618)	-0.026 (0.229)
Sleeping time x SG	5.53	3.58	-	0.04 (0.184)
Human capital				
Work	0.34	0.26	**0.081**** (0.048)	0.072 (0.100)
Reading (r1)	0.09	0.09	-0.049 (0.359)	-0.046 (0.379)
Writing (r1)	0.15	0.12	0.039 (0.472)	0.042 (0.433)
Numeracy	0.20	0.18	0.035 (0.396)	0.033 (0.414)
Scoremath	5.45	5.86	-0.007 (0.535)	-0.006 (0.726)
Scoremath x SG	3.21	2.20	-	0.001 (0.967)
Social capital				
Neighbourhood trust	0.56	0.56	-0.025 (0.431)	-0.049 (0.269)
Neighbourhood trust x SG	0.36	0.21	-	0.067 (0.326)
Respect by children	0.91	0.91	-0.002 (0.967)	0.001 (0.984)
Safe environment	0.65	0.57	0.047 (0.136)	0.048 (0.145)
Subjective well-being				
Satisfaction with life	5.99	6.00	-0.004 (0.566)	-0.012 (0.262)
Satisfaction with life x SG	3.69	2.35	-	0.018 (0.243)
Health capital				
Health status	0.62	0.63	-0.006 (0.834)	-0.015 (0.731)
Health status x SG	0.34	0.22	-	0.006 (0.925)
Long-term illness	0.13	0.11	0.004 (0.923)	-0.047 (0.454)
Long-term illness x SG	0.10	0.04	-	0.124 (0.334)
BMI	16.61	16.83	-0.004 (0.706)	-0.001 (0.903)
Height	153.52	154.70	0.001 (0.681)	0.001 (0.651)
Other group participation				
After school club	0.13	0.08	0.008 (0.876)	-0.046 (0.541)
After school club x SG	0.26	0.04	-	0.079 (0.531)
Religious group	0.34	0.21	**0.081**** (0.035)	**0.087*** (0.059)
Informal child group	0.18	0.12	0.055 (0.270)	0.055 (0.289)
Observations / Efron's R^2	118	540	8.4	10.2

Notes: Significance levels are indicated as *** ≡ p<0.01; ** ≡ p<0.05; * ≡ p<0.1. Significant effects are displayed in bold letters; Average marginal effects are displayed in the last column derived from a probit analysis. Bootstrapped p-values are in brackets. The model includes a constant term. Number of bootstrap replications: 4.999. If not mentioned otherwise, all covariates are measured in r2.

TABLE A.7

Radius matching results of sports group participation (model 1 and 2)

Outcomes	*Model 1*			*Model 2*		
	ATE	*ATET*	*ATENT*	*ATE*	*ATET*	*ATENT*
Human capital						
Vocabulary test	-0.617	-3.419	-0.009	-0.67	2.310	-1.347
	(0.727)	(0.441)	(0.989)	(0.706)	(0.561)	(0.510)
School enrolment	0.030	0.024	0.031	0.009	0.024	0.006
	(0.114)	(0.564)	(0.122)	(0.737)	(0.575)	(0.851)
Social capital						
Neighbourhood trust	-0.027	-0.061	-0.019	-0.078	-0.14	-0.064
	(0.810)	(0.648)	(0.880)	(0.522)	(0.340)	(0.648)
Neighbourhood trust (d)	-0.008	0.023	-0.015	-0.014	-0.066	-0.002
	(0.906)	(0.802)	(0.825)	(0.819)	(0.354)	(0.970)
Friends support	0.182	0.177	0.183	0.138	0.22	0.12
	(0.164)	(0.206)	(0.221)	(0.401)	(0.125)	(0.539)
Friends support (d)	**0.121****	**0.136***	**0.118***	0.097	**0.132***	0.089
	(0.044)	**(0.061)**	**(0.083)**	(0.175)	**(0.062)**	(0.278)
Respect by children	0.122	0.172	0.111	**0.137***	0.135	0.138
	(0.110)	(0.129)	(0.182)	**(0.097)**	(0.144)	(0.134)
Respect by children (d)	-0.008	0.007	-0.011	0.025	-0.032	0.038
	(0.858)	(0.907)	(0.821)	(0.629)	(0.566)	(0.529)
Safe environment	0.021	0.048	0.015	0.095	0.004	0.116
	(0.883)	(0.758)	(0.926)	(0.518)	(0.975)	(0.503)
Safe environment (d)	0.048	0.055	0.047	0.038	-0.026	0.053
	(0.451)	(0.441)	(0.516)	(0.585)	(0.821)	(0.509)
Subjective well-being						
Satisfaction with life	0.287	0.413	0.259	0.323	**0.505****	0.281
	(0.168)	(0.128)	(0.248)	(0.227)	**(0.040)**	(0.351)
Health capital						
Health status	**0.303*****	**0.246****	**0.315*****	0.152	**0.260*****	0.127
	(0.001)	**(0.018)**	**(0.004)**	(0.308)	**(0.008)**	(0.518)
Body Mass Index	0.279	0.031	0.333	0.259	0.361	0.235
	(0.381)	(0.939)	(0.384)	(0.546)	(0.303)	(0.635)
Other group participation						
After school club	-0.028	-0.040	-0.026	-0.041	-0.075	-0.034
	(0.457)	(0.224)	(0.565)	(0.609)	(0.197)	(0.440)
Religious group	-0.056	-0.108	-0.045	-0.042	-0.038	-0.043
	(0.100)	(0.193)	(0.266)	(0.364)	(0.407)	(0.389)
School club	0.017	-0.063	0.034	0.032	0.002	0.039
	(0.496)	(0.545)	(0.357)	(0.609)	(.)	(0.612)
Youth group	-0.014	0.015	-0.020	0.019	0.0004	0.023
	(0.293)	(0.568)	(0.156)	(0.687)	(0.983)	(0.687)
	No. of observations: 658 No. of treated: 118 (17.93% of observations) Common support: 650 (Share in common support: 98.78%)			No. of observations: 658 No. of treated: 118 (17.93% of observations) Common support: 623 (Share in common support: 94.68%)		

Notes: Significance levels are indicated as *** ≡ p<0.01; ** ≡ p<0.05; * ≡ p<0.1. Significant effects are displayed in bold letters. p-values (in brackets) obtained from 4.999 bootstrap replications as the share of absolute bootstrap t-statistics that are larger than the absolute t-statistic in the original sample. The results are based on a version, which uses weighted logits instead of weighted OLS for the binary outcomes (weights used for inference take bias adjustment not into account).