SOCIO-ECOLOGICAL PROFILES OF PARTICIPANTS IN SPORTS ACTIVITY IN POLAND: RANDOM TREE-BASED ANALYSIS

Elżbieta Biernat[1], Dawid Majcherek[2] , Łukasz Skrok[3]

[1] Department of Tourism, Collegium of World Economy, SGH Warsaw School of Economics, Warsaw, Poland
[2] Collegium of World Economy, SGH Warsaw School of Economics, Warsaw, Poland;
[3] Department of Business Economics, Collegium of World Economy, SGH Warsaw School of Economics, Warsaw, Poland

Abstract

While the benefits of sports activity are well-documented, people remain reluctant to engage in it in selected developed countries (including Middle and Eastern Europe). A comparison of empirical results indicates that the reasons behind such behaviour are heterogenous and complex. As implied by the socio-ecological framework, different spheres of life and the interactions between them matter. By using the results of a study on Polish society conducted in 2015, Social Diagnosis, and a random forest model, we show the key demographic, socio-economic characteristics and interactions between them in explaining the difference between engaging in sports activity and not doing so. In particular, we identify three crucial variables: age, participation of other adults in the same household in sports activity, and proclivity for restaurants, bars, pubs.

Key words: sport activity, socio-ecological model, prediction, analysis

Introduction

Sport is a common form of Leisure Time Physical Activity (LTPA), which has been shown (Eime et al., 2015; Thiel et al., 2018) to bring many health benefits and contribute to combatting the pandemic development of non-communicable diseases. In the last ten years, public health researchers have paid significant attention to physical activity (PA) determinants including sport activity (SA) (Kohl et al., 2012). Mostly, they have analysed individual factors such as gender, age, education, background, needs and motivations (Allender et al., 2006; Teixeira et al., 2012) and economic factors such as socio-economic status (Bauman et al., 2012; Beenackers et al., 2012) time budget and human capital (Downward, 2007; Hallmann et al., 2012). There is a general consensus that a high socio-economic position (including, occupational position, income and education level) positively correlates with higher LTPA. They argued that in most European countries, participation in sport decreases with age, that it is more often practised by men and people living in urban areas (Hoekman et al., 2017). They found that psychosocial factors such as knowledge, attitudes, beliefs and social support can influence people's behavioural changes (Conner & Sparks, 1996; Hawley-Hague et al., 2014). For example, the social environment (and especially social networking) can influence the SA of community members by replicating the behavioural patterns of influential people from networks of friends, families, peer and cultural groups (Christakis & Fowler, 2013; Datar & Nicosia, 2018; Yu et al., 2011).

There are also reports that indicate the opposite. For example, that it is the inhabitants of rural areas (Hoekman et al., 2017) and women (e.g. in Denmark and the Netherlands they are more active in sport, that the family can reduce men's participation in sport (Grima et al., 2017), and that sport takes up an increasing proportion of older people (Breedveld & Hoekman, 2011). This suggests that the set of determinants for SA, representing the behavioural determinants regarding its undertaking may be strongly dependent on the specificity (demographic, health, social, economic, institutional) of the surveyed community. The greater the value of empirical analyses conducted on different communities, both from the practical (for policy design) and theoretical (for understanding the processes leading to engaging or not engaging in SA) point of view. It therefore begs the question whether the existing evidence base to support the PA promotion policy provides an adequate understanding of the reasons for undertaking it or not. This uncertainty is confirmed by Ding et al. (2020), who, based on the 2012 and 2016 Lancet PA series, point to major knowledge gaps regarding PA Research, including mixed empirical evidence and dependence of results on selection of variables and their interactions. They emphasize the need for in-depth analyses– taking into account the multilevel relationships between individuals and their environment (as also indicated by other researchers (Ding et al., 2020; World Health Organization, 2018). After all, a number of factors concentrated in a holistic environment can influence a given behaviour, whether people are active in sport or not (Biernat et al., 2020). According to the socio-ecological approach

(Bronfenbrenner, 1979), the process that determines health behaviours is complex and includes intrapersonal, interpersonal and environmental factors. Every behavioural impact potentially interacts across these different factors (Sallis et al., 2008). Understanding such multilevel relationships will help to identify (from people's day-to-day environment) factors that may act as barriers or amenities for PA. It will also serve to identify population patterns of participation in sport (Stokols, 1996) which can be used as a basis for strategic policy planning or as leverage to change existing programmes and achieve desired results (Eime et al., 2015; Rowe et al., 2013).

The aim of our study, based on the socio-environmental framework recently applied in health and PA research (Hoekman et al., 2017) is to identify the key factors that coexist with the participation of Poles in sport. We want to look for interactions between these variables and build "socio-demographic" profiles that determine the likelihood of participation in sport.

Materials

We supplemented the dataset with measures of sports infrastructure availability based on the Statistics Poland dataset on sports facilities, by calculating the number of particular facilities per capita in the subregion (NUTS, Classification of Territorial Units for Statistics, 3 level) where each respondent lived during the period when the SD survey was conducted. (Statistics Poland, 2014).

We used a random subsample of 20% of all observations (n=4,449) from SD dataset wave 2015 as a training data set for modelling and we took the remaining 80% (n=17,671) as a test set for validation purposes and to compare the tested methods.

Our operational definition of SA in the analysis that follows stems directly from the question used in the SD survey. The (multiple choice) question was 'Do you practice any sport or PA?:' and the set of possible answers: 1. no, I do not practice any sport or PA; 2. aerobics; 3. running/jogging/Nordic walking; 4. gym; 5. cycling; 6. skiing or other winter sports; 7. swimming; 8. football or other team sports; 9. yoga; 10. martial arts; 11. another sport or type of PA. The total set of observations (n = 22,120) comprises the respondents that answered the above question and had no missing values for the variables constructed as described in Appendix 1.

Methods

In order to identify important variables for the participation of Poles in sport and to determine population profiles, SA probability modelling was carried out using two classification trees (cTree, pCART) and a random forest (rF). The article describes in detail the results of the Pruned Classification and Regression Tree (pCART), which

was selected as the best estimation method based on the largest area under the ROC (receiver operating characteristic) curve. Other model results are presented in appendices 2 (cTree) and 3 (rF).

Since SD is a complex study with broad number of variables, we based the preliminary selection of inputs to the models on theoretical framework comprising 4 groups of factors (figure 1):

1. Individual factors (age, BMI, dissatisfaction with health, joie de vivre, loss of libido, years of education completed, English language, occupational group, time spent working, time spent commuting, having as much energy as ever to work).
2. Intrapersonal factors (acquaintances met regularly, family members met regularly, number of friends, number of social meetings last month, number of entertainment facilities visited last month, number of restaurants etc. visited last month, marital status, household ratio of adults sport activity, household income last year, size of the living space, mobile phone, household income, number of children (age: 0-4) in the household, number of adults in the household).
3. Organisational factors (number of sports facilities of different types, related to the population in the subregion inhabited by the respondent).
4. Community factors (size class of living area of the household, work for local society, membership of organisations).
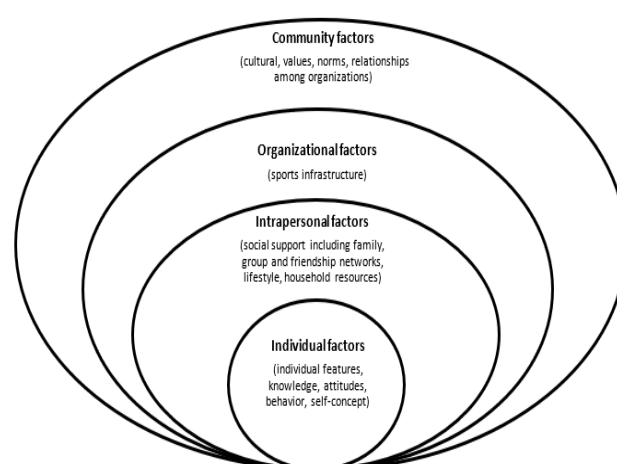

Figure 1. Adapted version of the socio-ecological model of sport activity determinants

All the figures presenting the results use abbreviated names of the above-mentioned variables. A specific description of the variables describing all the factors can be found in Appendix 1. Due to the large number of variables we only list variables shown in figure 3, supplementary figure 1 or supplementary figure 2 – the scope of variables used as inputs was broader. In particular, it also included variables describing: gender, happiness with life, satisfaction with income, volunteering, general trust, participation in elections, number of

older children in the household, involvement of children and adolescents in sports activity, father's education level, abuse of alcohol, recreational drug use, being vegetarian, smoking, having books in the household, not having a washing machine, being retired, being unemployed, working full time, being an entrepreneur, being self-employed, spending time caring for members of family, meeting friends, having physical problems, having headaches, sweating problems, breathing problems, palpitations, body aches, shivers, bladder problems, constipation, nosebleeds, blood pressure issues, fatigue unconnected with work, and number of sport facilities of various types, per capita, in a given subregion).

Figure 2 shows the ROC curves for the Conditional Inference Tree (cTree), pruned classification and regression tree (pCART) and Random Forest (rF) respectively.
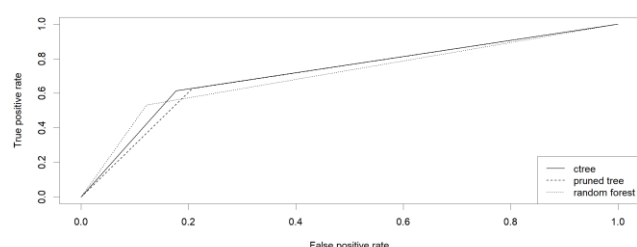
Figure 2. Comparison of sensitivity and specificity in different models

The area under the ROC curve presented on figure 2 is defined as AUC. The values of this measure for the three compared models are presented in Table 1, as well as the results of the PCC (percentage of observations correctly classified) estimation.

Table 1. Accuracy measure

Classification method	PCC	AUC
Condititonal Inference Tree	0.73	0.71
Pruned Tree	0.75	0.72
Random Forest	0.76	0.67

Note: PCC – percentage of observations correctly classified, AUC – area under curve

Table 1 shows that the random forest method (rF) has the highest accuracy, because about 76% of the observations were correctly classified, i.e. 3 p.p. more than for a single classification tree (ctree). Nevertheless, in terms of the AUC criterion, the pruned classification and regression tree (pCART) proved to be the best, with the field under the ROC curve being 0.72 and the rF having the lowest value.

The Pruned Classification and Regression Tree is based on a recursive node-splitting algorithm that minimises the measure of diversity in the nodes (Hastie et al., 2009). The rpart function from the party package was used for this purpose (Hothorn et al., 2020). Details are shown in appendix 4. Due to the risk of overfitting the model, the initial tree was truncated after the generation of the initial tree.

A pCART tree of optimal complexity is shown in Figure 3. The nodes (which are not leaves) contain the conditions for a sample split. The observations that meet the split condition set in the node move to the left descendant and the others to the right descendant. The leaves contain information on the most numerous category and the distribution of SA. By categorising information about the created split paths into groups with different SA probability levels, Table 2 was developed, which presents a detailed description of the profiles presented in Figure 3.

Results

The analysis conducted using a pruned Classification and Regression Tree (Figure 3) shows that three factors were fundamental when differentiating active and inactive people in Poland in 2015: use of restaurants, bars or pubs at least once in the previous month, percentage of other household members engaging in SA and age.

First of all, those frequenting restaurants, bars or pubs were more likely to be involved in SA (63% involvement in SA among those using restaurants etc. vs. 25% among non-users).

Secondly, SA was more often undertaken by respondents who declared a higher share of the remaining adult members of their household involved in SA. Therefore, as shown on the left side of Figure 3, 45% of respondents not using restaurants etc. participated in SA if at least 29% of other adults from their households also participated in SA. On the other hand, among respondents who did not use restaurants etc., but for whom less than 29% of other adults in their households participated in SA, only 17% participated in SA.

On the other hand, as shown on the right-hand side of Figure 3, among people using restaurants etc. and over the age of 40, the percentage of involvement in SA was 69% – provided that 83% percentage other adults in the household participated in SA (profiles 12). On the other hand, if less than 83% of other adults in the household participated in SA, then only 43% of respondents aged 40 and over who used restaurants were engaged in SA.
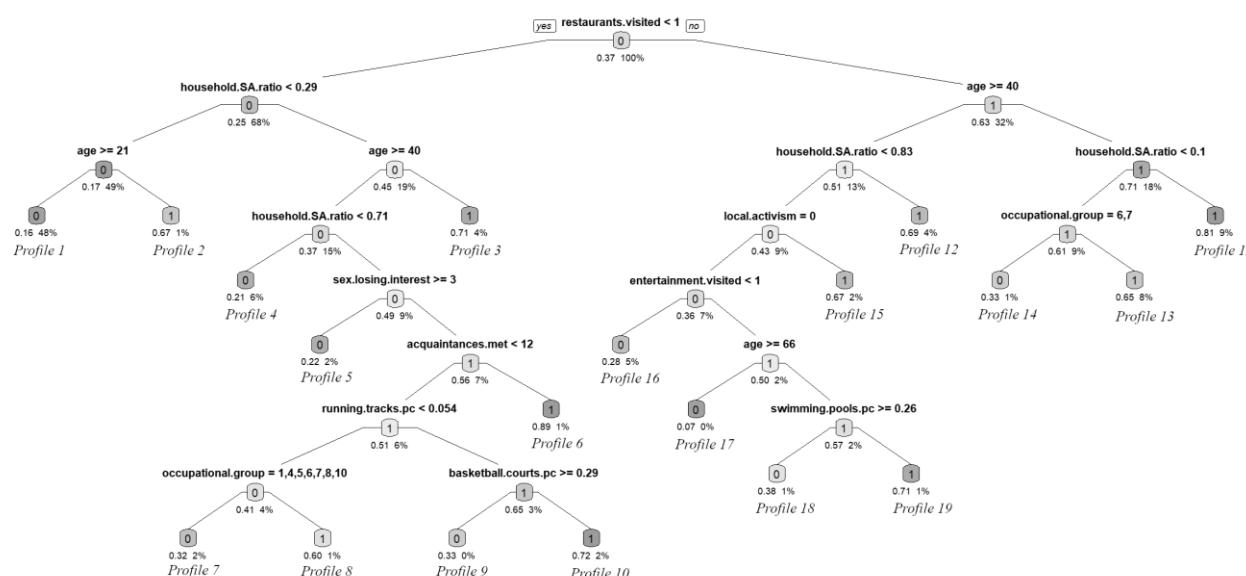
Figure 3. Classification tree from the pruned Classification and Regression Tree (pCART) model based on the training data set.

The right-hand side of Figure 3 shows that among people using restaurants etc. under 39 years old, as many as 81% were involved in SA if at least 10% of the remaining adults from their households were involved in SA (profiles 11). On the other hand, if the SA of the remaining adults in the household was low (below 10%), only 61% of respondents aged 39 and under who frequent restaurants, etc., were involved in SA.

Thirdly, as a rule, the higher the age, the fewer the people involved in SA. The left side of Figure 3 shows that this variable was important for people not going to restaurants etc., along with a low SA for other adults in the household (below 29%). In this group, as many as 67% of people under 21 years of age were involved in SA in (profile 2), whereas among older people (from 21 years of age) this figure fell to just 16% (profile 1).

The left side of Figure 3 also shows that as many as 71% of people under 40 years of age who do not go to restaurants etc. were involved in SA but the SA involvement of other adults in the household was high (from 29%; Profile 3). At the same time, only 37% of people aged 40 and over were involved in SA if they did not frequent restaurants but a significant proportion of other adults from their households were involved in SA (from 29%).

The right side of Figure 3 shows that among the people going to the restaurant etc., 71% of people under 40 years of age and only 51% of people over 40 years of age were involved in SA.
The conclusion about the key role of the three described variables is confirmed by the results of the analysis conducted using a random forest (appendix 3). The key role of the latter two variables is confirmed by the analysis carried out using a conditional inference tree (appendix 2). Due to the complexity of the relationships between the individual variables, the subsequent section of the description focuses on the final profiles (leaves of the tree) and not the subsequent variables (nodes). Based on pCART analysis, 19 respondent profiles were developed – ranked according to the probability of whether observations will be classified as being involved in SA (Table 2).

Profile 1 covers the most numerous group of Poles (almost 48% of all respondents). The probability of SA in this group is one of the lowest – just 16%. These are people aged >20 who have not been to a restaurant, bar or pub in the last month, for whom the percentage of people active in sports in the household is less than 29%.

The lowest (in the whole sample) probability of SA (7%) is characteristic of people from profile 17 (1% of all respondents). These are people aged 66 and over who have not undertaken any activity for the benefit of the local community in the last two years and the percentage of people active in sports in their household was less than 83%. However, they were in a restaurant, bar or pub and in a cinema, theatre or concert during the last month.
In contrast to the described groups are people from profiles 6 and 11. The highest probability of SA (89%) is among Poles from profile 6 (1% of all respondents). This includes people aged >40 years who still have a healthy libido. The percentage of people active in sports in their household is higher than 71%, and the number of friends met regularly during the year is higher than 12. These people have not been to a restaurant, bar or pub in the last month.
For Poles in profile 11 (9% of all respondents), the SA probability is 81%. These are people who are < 40 years old and who have been to a restaurant, bar or pub in the last month. The percentage of people active in sports in their household is higher than 10%.

Table 2. Profile description based on the pCART including the probability of SA and distribution of observations

Profiles	% of observations	Probability of SA (declining)	Characteristics of the persons examined
6	1%	89%	Persons aged ≥ 40 who have not been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sports in the household is > 71%, who still have a healthy libido and the number of friends met regularly during the year is > 12
11	9%	81%	Persons aged <40 years who have been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sport in the household is ≥10%
10	2%	72%	Persons aged ≥40 who have not been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sports in the household is > 71%, who still have a healthy libido, the number of friends met regularly during the year is < 12, the percentage of treadmills/ per capita in its region is ≥0.054 and the percentage of basketball pitches < 0.29
3	4%	71%	Persons aged <40 years. who have not been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sports in the household is ≥29%.
19	1%	71%	Persons aged 40-66 who have been to a restaurant, bar, pub, cinema, theatre or concert in the last month, for whom the percentage of people active in sports in the household is < 83%, who have not undertaken any activity for the benefit of the local community in the last 2 years, and the percentage of swimming pools in the region / per capita is <0.26
12	4%	69%	Persons aged ≥40 years who were in a restaurant, bar, pub in the last month, for whom the percentage of people active in sport in the household is ≥83%
2	1%	67%	Persons aged <21 years who have not been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sport in the household is < 29%
15	2%	67%	Persons aged ≥40 years who have been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sports in the household is < 83%, who have been active in the local community in the last 2 years
13	8%	65%	Persons aged <40 years, from a professional group other than 6 and 7 according to ISCO, who have been to a restaurant, bar, pub in the last month and for whom the percentage of people active in sport in the household is < 10%
8	1%	60%	Persons aged ≥40 years, from the professional group 2.3.9 according to ISCO, who have not been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sports in the household is > 71%, who still have a healthy libido, the number of friends regularly met during the year is < 12, the percentage of treadmills per capita in the region is < 0.054
18	1%	38%	Persons aged 40-66 who have been to a restaurant, bar, pub, cinema, theatre or concert in the last month, for whom the percentage of persons active in sports in the household < 83% who have not carried out any activity for the benefit of the local community in the last 2 years and the percentage of swimming pools in the region/ per capita is ≥0.26
9	0-1%	33%	People aged ≥40 who have not been to a restaurant, bar, pub in the last month, still have a healthy libido, for whom the percentage of people active in sports in the household is > 71%, the number of friends regularly met during the year is < 12, the percentage of treadmills per capita in the region is ≥0.054 and the percentage of basketball pitches ≥0.29
14	1%	33%	Persons aged <40 years, from professional groups 6 and 7 according to ISCO, who were restaurants, bars, pubs in the last month, for whom the percentage of people active in sport in the household is < 10%
7	2%	32%	Persons aged ≥40 years, from the professional group 1,4,5,6,7,8,10 according to ISCO, who were not restaurants, bars, pubs in the last month, who still have a healthy libido, for whom the percentage of people active in sports in the household is > 71%, the number of friends regularly met during the year is < 12, and the percentage of treadmills per one inhabitant in the region is < 0.054
16	5%	28%	Persons aged ≥40 years who have been to a restaurant, bar, pub in the last

			month, for whom the percentage of people active in sports in the household is < 83%, who have not carried out any activity for the benefit of the local community in the last 2 years and have not been to a cinema, theatre or concert in the last month
5	2%	22%	Persons aged ≥40 years who have not been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sports in the household is > 71% and who have completely lost interest in sex
4	6%	21%	Persons aged ≥40 years who have not been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sport in the household is 29%-71%
1	48%	16%	Persons aged ≥21 who have not been to a restaurant, bar, pub in the last month, for whom the percentage of people active in sport in the household is < 29%
17	1%	7%	Persons aged ≥66 who have been to a restaurant, bar, pub, cinema, theatre or concert in the last month, for whom the percentage of people active in sport in the household is < 83% and who have not carried out any community activities in the last 2 years and last month

Note: ISCO (*ISCO – International Standard Classification of Occupations*, 2008); 1 Public authorities, senior officials and managers; 2 Specialists; 3 Technicians and other associate professionals; 4 Clerical workers; 5 Service and sales workers; 6 Farmers, horticulturists, foresters and fishermen; 7 Industrial and artisan workers; 8 Plant and machine operators and assemblers; 9 Simple workers; 10 Armed Forces

Discussion

The aim of this study was to identify the key factors that determine the participation of Poles in sport and to define the profiles of the respondents – in terms of their likelihood of participating in sport. In order to include a wide range of variables (from different spheres of human life) and recognise the importance of dynamic, multi-directional influences between people and their living environment, we applied a social and environmental approach. The results obtained may be of great importance for shaping pro-health policy (McLeroy et al., 1988; Stokols, 1996).

The first thing that came to our attention was that either Poles are active (SA probability was 60% and more) or not (SA probability – 38% and less). These percentages can be interpreted as SA probabilities for a given profile. There were no cases between these ranges. From this perspective, the lack of profiles with probabilities close to 50% is a positive result. It means that no profile has been constructed for which engagement or non-engagement in SA are equally probable events. What is more, it allows to distinguish – with an (unweighted) average for the whole SA sample of 37% – profiles with exceptionally frequent SA (6 and 11 – covering 10% of observations in total), profiles with a high share in SA (10, 3, 19, 12, 2, 15, 13 and 8 – covering 23% of observations in total), profiles with a close to average share in SA (18, 9, 14, 7 and 16 – covering 10% of observations in total) and profiles with a low share in SA (5, 4, 1 and 17 – covering 57% of observations in total). This gives some indication to decision-makers to pay attention to radically different behaviours in different social groups.

On the one hand, we have a 10% "sports elite" (profiles 11 and 6) – younger people, probably from the modern upper-middle-class (which shall be discussed in more detail later) and older people who are very active socially and in life. On the other hand, there are people with very low participation in SA (profile 5, 4, 1 and 17), including: (1) after 66 years of age, (2) at the age of 40 – using restaurant services to a lesser extent, with less life activity, (3) from households with low participation in SA. Between these extremes, profiles with an average SA level can be identified (18, 9, 14, 7 and 16). These include mainly people over 40, including those working in traditional sectors of the economy, with a lower frequency of interpersonal contact, who are less often involved in social activities and from households with low participation in SA.

Our analysis suggests that the SA of Poles may be subject to a set of sociodemographic variables. One ought to bear in mind that, especially for nodes other than the root note or the two top internal nodes (i.e. except for restaurants, household engagement in SA and age of the respondent), the selection of variables relies not only on the choice of specific method of estimation (i.e. type of tree), but also on the (random) selection of observations for the estimation sample. In particular, some variables (especially those describing different types of sports facilities subregionally) tend to be substitutes for each other for slightly different subsamples or trees. They do, however, represent underlying factors that coincide with SA. Therefore, indirect interpretation might be more relevant than attempts to relate them to SA directly.

At the first level of the classification tree appears a key factor for explaining the SA of Poles, i.e. visiting a restaurant/bar/pub during the last month. This may be connected, first of all, with changes taking place within society, i.e. moving away from the traditional family day model and eating out in the so-called "city" – especially for a specific group of people, i.e. : (1) with a fairly high socio-economic status, with disposable income and enjoying life; (2) reserving time, for example for SA. This may indirectly indicate the relationship of SA with a

metropolitan, modernist lifestyle. Namely, it is common for households to substitute household production with market services. This includes cooking and catering services (Been et al., 2015). The inclination to choose various ways to satisfy a basic need (food) depends, inter alia, on remuneration and the hourly rate of pay, which affects the valuation of leisure time and the relative prices of market services (Becker, 1965). In other words, the development of the range of market services has enabled sufficiently wealthy households to substitute home production, while allowing them to devote more time to professional work or leisure activities, including sport. In recent years, there have been reports concerning the inclusion of SA in the lifestyle of people previously unrelated to it, associated with modern sectors of the economy (Muniowski, 2018).

Secondly, it may be a result of the fashionable trend for watching sports broadcasts in public places (e.g. in cafés) and sharing sports emotions with friends, which may affect the social integration of people with similar interests, building relationships and mutual influence of active leisure patterns (Strzemińska et al., 2012). It is also consistent with the results of research – e.g. in the UK – showing that people who consume alcohol are significantly more often involved in sports/physical activity (Downward & Riordan, 2007). Finally, this may be a result of such premises being transformed into so-called "third places" (Strzemińska et al., 2012), which – alongside "first" places (private homes) and "second" (i.e. a public area) – build an interactive space, providing an opportunity for people to join forces in different areas of life and "full of life" throughout the day. In the morning they are offices for professionals (who set up their mobile offices in cafés), while in the evenings they turn into music clubs, workshops or game rooms. There are more and more such places in Poland. They include sports cafés (Dołżycka, 2017). They go beyond the role of a catering establishment and create an environment that is a source of authentic energy, in which further actions are planned and move from "being" to "co-creating" subsequent events (Strzemińska et al., 2012). Cafes/restaurants/pubs, which are located in shopping malls, have an important place within the locality. There, sports attractions are not only limited to a fitness club or a climbing wall. The best shopping malls in the world have bowling alleys, ice rinks, extreme sports centres, ski slopes and aqua parks with waveforms for surfers (Colliers International, 2017). This can foster a specific, sporting active lifestyle (especially in larger cities). Of course, the most numerous group of regulars and activists are young adults (Strzemińska et al., 2012). The energy and creativity of this group is most strongly influenced by the activity of metropolitan cafés. Furthermore, cafés – with their game corners – are also an attractive space for parents with children. The multi-generational landscape of the guests of these premises is completed by seniors, who have started to discover them as a space for various activities, e.g. connected with Third Age Universities. The windows of many Warsaw cafes feature "seniors-friendly" stickers, awarded as part of the Society of Creative Initiatives campaign.

The identification of the decision tree clusters also shows that participation in SA depends (apart from visiting restaurants/bars/pubs) on several other factors. These include both individual factors (i.e. age, occupation, libido) and intrapersonal factors (i.e. percentage of people active in sports in the household, number of regular friends) and organisational factors (i.e. number of sports facilities per capita in the sub-region) and community factors (i.e. activity for the local community and visiting cinemas/theatres/concerts). More specifically, the relevant determinants of SA vary depending on the context under consideration (from the associated variable) and offer different behavioural outcomes, i.e. the probability of undertaking SA. The analysis of interaction factors revealed 19 possible profiles for Poles in terms of this issue.

The highest probability of undertaking SA was in profile 6 (89%) and 11 (81%). In the former, these were people aged ≥40 years who still have a healthy libido, for whom the percentage of people active in sports in the household is >71% and who have not been to a restaurant/bar/pub in the last month, but the number of friends regularly met during the year was >12. This indicates that different environments can jointly influence the formation of people's behaviour – as confirmed by other studies (Franzini et al., 2010; Sallis et al., 2006). Downward and Rasciute (2016) show that the SA of other household members contributes to activation, and this effect is particularly strong for women. (Downward & Rasciute, 2016). By influencing each other and to some extent SA, they can become moderators of this activity. Thus, it may be that Poles aged ≥40 will only practice sport if they are sexually active (the cause-effect relationship is not clear) and if they have enough time and proper organisation of family life. However, the key relationship of personal SA with that of household members may also result from the need to maintain an appropriate work-life balance in terms of the household. What we mean here is to provide members of the household with adequate resources for engaging in activities other than those directly related to work or household duties (including material resources and time).

The available literature does not support the conclusion that people aged ≥40 years will only play sports if they are sexually active and have enough time and proper organisation of family life. The evidence base in this regard is small and inconsistent. However, there is evidence of the links between these factors and SA. Salvador et al. (Salvador et al., 2009), for example, point to the role of family and friend support in this issue, and Prins et al prove that this support can go beyond motivation and reflect actual action (Prins et al., 2010). The physical activity of those in the immediate environment can be conducive to

imitation (Gallagher et al., 2010), especially if the people involved are strongly connected (McMurdo et al., 2012).

Profile 11, with the second highest probability of undertaking SA, features Poles aged <40 years, who visited a restaurant/bar/pub in the last month and for whom the percentage of people active in sports in the household was ≥10%. This confirms the previously discussed phenomenon that so-called "third places" may constitute an important interactive space – especially for young adults (twenty-something age group) (Strzemińska et al., 2012) who do not have family support. If it were not for this type of relationship, it is not known whether these people would take up sport. This conclusion is also supported by the fact that profile 1 (with one of the lowest probabilities of undertaking SA – 16%) includes people aged ≥21 who have not been to a restaurant/bar/pub in the last month.

The least active in sports were Poles from the 17th profile (SA probability – 7%). They were ≥66 years old, had not visited a restaurant/bar/pub or cinema/theatre/concert in the last month, and the percentage of people active in sports in their household was <83%. Furthermore, these people have not carried out any activities for the benefit of the local community in the last 2 years. This indicates that the behaviour of older people depends not only on individual predecessors or physical environment (Fishbein, 2000; Kremers et al., 2006), but also on community factors. According to research, such factors play an important role in shaping people's awareness and influencing their SA (Sallis et al., 2006). The social environment determines values, norms, behaviour, and thus influences action (Tyszka, 1993). Social networks, which form as a result of social life, build social capital, which is proven to be conducive to PA (Lechner & Downward, 2017). In our opinion, community factors such as visiting a cinema/theatre/concert or a restaurant/bar/pub, and acting for the benefit of the local community, can act as mediators or SA effects (as indicated by (Skrok et al., 2019). The mere provision of recreational and sports infrastructure – despite the fact that it is an important SA condition (Nathan et al., 2014)– may not be enough to keep seniors active (Giles-Corti & Donovan, 2003). In other words, outdoor gyms, bicycle paths, etc. will be used more often if the social environment supports this project – as the literature confirms (Franzini et al., 2010). According to Boulton et al. (2018), the social element of engaging in activities was found to be an important motivator to adhere to PA (Boulton et al., 2018). In the case of the elderly, this is greatly facilitated by living in districts with a wide range of cultural and entertainment attractions on offer (Nathan et al., 2014).

The aforementioned sports and recreation infrastructure is an SA determinant, which is only revealed in some cases (in groups of people over 40 years of age) and only at level 6 of the decision tree. Thus, its potential importance is relevant for only 8% of the sample and is conditioned by other factors. It is not significant for Poles, for whom the probability of SA is <32%. On the other hand, it is important for the active ones, i.e. from the following profiles: 10, 19, 8, 18, 9 and 7 (where the probability is 72%; 71%; 60%; 38%; 33% and 32% respectively). This confirms previous reports by Dishman et al (1985), according to which active people are twice as likely as inactive people to claim that more availability of sports facilities would increase their participation in PA (Dishman et al., 1985). Nevertheless, studies so far provide evidence that there is a link between taking up sport and the perceived and objectively measured availability of sports and leisure facilities (Davison & Lawson, 2006) and that infrastructure is not essential to the growth of PA (Ferreira et al., 2007; Van Der Horst et al., 2007).

The fact that sports infrastructure reveals itself as an important variable only among people >40 years of age may result from the fact that, as Biernat and Piątkowska (2018) write, current sports facilities seem to be aimed at young, athletic people (Biernat & Piątkowska, 2018). Adults do not feel like they are being addressed. Besides, for young people, sports infrastructure is not so important (the opportunities to participate in SA are created organically, according to their needs and fashions for specific sports).

Our results show that the presence of infrastructure is not a direct factor that affects the SA of people aged >40 years. Other variables also play an important role in this process, for example the percentage of people active in sports in the household. This fact was previously pointed out by other researchers (Giles-Corti & Donovan, 2003). Among others, they demonstrated that mediators between sport facilities and SA may include attitude and intentions that perception of accessibility can be the result of cognitive interpretation and processing (van der Horst et al., 2010).

The last significant variable in the model was membership of a professional group. We cannot fully confirm a known correlation (Biernat & Piątkowska, 2018; Grima et al., 2018) – that people with a higher professional status have a higher probability of SA. On the other hand, we can state that Poles from occupational group 6 and 7 (according to ISCO – farmers, horticulturists, foresters and fishermen, as well as industrial workers and craftsmen), are among those with a low probability of participating in sport (profile 14 and 7 – 33 and 32% respectively), which has been proven previously (Biernat et al., 2017). Analysis of the decision-making tree has shown that community factors are important in this respect, including, among others, visiting a restaurant/bar/pub in the last month and intrapersonal factors, including the number of friends regularly met during the year and the percentage of people in the household who are active in sport. This indicates that these factors

have the possibility to inspire changes in the behaviour of people in these groups.

Strengths and limitations

The strength of our study is that it is based on Social Diagnosis (SD), which has the advantages of a relatively large, representative sample and an extensive set of socio-economic indicators. The multilevel structure of SD gave the possibility to calculate variables in the household, including, as it turns out, the crucial percentage of other adults active in sports. The analysis using random trees allowed for the selection of the most important factors and identification of interactions between them. This, in turn, made it possible to build a (simultaneously complex and concise) picture of SA in Poland. The estimation of alternative models (cTree and random forest) and fine-tuning of parameters, on the one hand, made it possible to choose the best possible representation of the study results and, on the other hand, to carry out a robustness check regarding the weight of individual variables in explaining activity (i.e. their hierarchy) and assessing model stability. The lack of an a priori approach to defining interactions between the variables also made it possible to obtain unexpected results, for example, indicating the key role of visiting restaurants, bars, pubs and the like, or the absence of a gender-specific variable in the final results.

The limitation of the test is the instability of its set of variable parts. It is not possible to present in one random tree. On the other hand, the use of a random forest to solve this problem does not facilitate the construction of a similar illustration. In this sense, some of the results should be interpreted indirectly. This is also influenced by the lack of measures in the source survey (SD) describing the availability of sports infrastructure and the limited accuracy of information about the respondents' place of residence. For this reason, it was only possible to sub-regionally assign a measure of the availability of sports facilities (NUTS 3) to the respondents. Another limitation of the survey is the lack of information about the volume, intensity and frequency of the respondents' SA, as well as the failure to include forms such as dance in the questionnaire. The SD survey was fully self-assessed.

When interpreting the results, it should also be remembered that due to the purpose of the study, the analysis did not use the panel nature of the data and methods allowing for the assessment of the cause and effect relationship between the analysed variables.

Conclusion

Our results have shown that the identification of key factors for taking up sport should be based on a social and environmental framework as well as the interactions between the different levels, since the multi-level relationship between an individual and his or her environment is important in this respect. The variable that proved crucial in describing the variability of Poles' SA is whether or not they visited a restaurant/bar/pub in the last month. Of course, this variable does not determine sports activity per se. Participation in SA also depends on individual factors (age, occupation, libido), intrapersonal (percentage of people active in sports in the household, number of regular friends or visits to the cinema/theatre/concert), organisational (percentage of sports infrastructure per capita in the region) and community (activity for the local community). SA determinants vary depending on the context under consideration (from the related variable) and give different behavioural results, i.e. the probability of undertaking SA. By influencing each other and to some extent SA, they become moderators of this activity. Nineteen profiles of Poles were identified according to SA probability, demonstrating that they are either active or not. This gives some indication to the decision-makers that such strongly differentiated behaviours are noticed. The results also show that increasing the physical activity of Poles requires a wider scope of activities than just promoting and increasing sports and recreation infrastructure. In the case of people aged 20-40, cultural factors may be of particular importance, both within the household and in terms of organising life outside it.

References

Allender, S., Cowburn, G., & Foster, C. (2006). Understanding participation in sport and physical activity among children and adults: A review of qualitative studies. *Health Education Research*, *21*(6), 826–835. https://doi.org/10.1093/her/cyl063

Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J., & Martin, B. W. (2012). Correlates of physical activity: Why are some people physically active and others not? *The Lancet*, *380*(9838), 258–271. https://doi.org/10.1016/S0140-6736(12)60735-1

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal*, *75*(299), 493–517. JSTOR. https://doi.org/10.2307/2228949

Been, J., Rohwedder, S., & Hurd, M. D. (2015). Home Production as a Substitute to Market Consumption? Estimating the Elasticity Using Houseprice Shocks from the Great Recession. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2745152

Beenackers, M. A., Kamphuis, C. B. M., Giskes, K., Brug, J., Kunst, A. E., Burdorf, A., & van Lenthe, F. J. (2012). Socioeconomic inequalities in occupational, leisure-time, and transport related physical activity among European adults: A systematic review. *The International Journal of Behavioral Nutrition and Physical Activity*, *9*, 116. https://doi.org/10.1186/1479-5868-9-116

Biernat, E., Bartkiewicz, P., & Buchholtz, S. (2017). Are Structural Changes in Polish Rural Areas Fostering Leisure-Time Physical Activity? *International Journal of Environmental Research and Public Health*, *14*(4). https://doi.org/10.3390/ijerph14040372

Biernat, E., & Piątkowska, M. (2018, July 31). *Stay active for life: Physical activity across life stages*. Clinical Interventions in Aging. https://doi.org/10.2147/CIA.S167131

Biernat, E., Skrok, Ł., Majcherek, D., & Nałęcz, H. (2020). Socioecological Profile of Active Adults. Sport as a Whole-life Choice. *Physical Culture and Sport. Studies and Research*, *85*(1), 59–76. https://doi.org/10.2478/pcssr-2020-0007

Boulton, E. R., Horne, M., & Todd, C. (2018). Multiple influences on participating in physical activity in older age: Developing a social ecological approach. *Health Expectations : An International Journal of Public Participation in Health Care and Health Policy*, *21*(1), 239–248. https://doi.org/10.1111/hex.12608

Breedveld, K., & Hoekman, R. (2011). Measuring sports participation in the Netherlands – the need to go beyond guidelines. *European Journal for Sport and Society*, *8*(1–2), 117–132. https://doi.org/10.1080/16138171.2011.11687873

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: Examining dynamic social networks and human behavior. *Statistics in Medicine*, *32*(4), 556–577. https://doi.org/10.1002/sim.5408

Colliers International. (2017, October 2). *Nadchodzi zmiana? Nowy wymiar rozrywki i rekreacji w centrach handlowych [Is change coming? A new dimension of entertainment and recreation in shopping centers]*. Outsourcing Portal. http://www.outsourcingportal.eu/pl/nadchodzi-zmiana-nowy-wymiar-rozrywki-i-rekreacji-w-centrach-handlowych

Conner, M., & Sparks, P. (1996). The theory of planned behaviour and health behaviours. In *Predicting health behaviour: Research and practice with social cognition models* (pp. 121–162). Open University Press.

Czapiński, J., & Panek, T. (2015). *Social diagnosis—Objective and Subjective Quality of Life in Poland*. http://www.diagnoza.com/

Datar, A., & Nicosia, N. (2018). Assessing Social Contagion in Body Mass Index, Overweight, and Obesity Using a Natural Experiment. *JAMA Pediatrics*, *172*(3), 239–246. https://doi.org/10.1001/jamapediatrics.2017.4882

Davison, K. K., & Lawson, C. T. (2006). Do attributes in the physical environment influence children's physical activity? A review of the literature. *The International Journal of Behavioral Nutrition and Physical Activity*, *3*, 19. https://doi.org/10.1186/1479-5868-3-19

Ding, D., Ramirez Varela, A., Bauman, A. E., Ekelund, U., Lee, I.-M., Heath, G., Katzmarzyk, P. T., Reis, R., & Pratt, M. (2020). Towards better evidence-informed global action: Lessons learnt from the Lancet series and recent developments in physical activity and public health. *British Journal of Sports Medicine*, *54*(8), 462–468. https://doi.org/10.1136/bjsports-2019-101001

Dishman, R. K., Sallis, J. F., & Orenstein, D. R. (1985). The determinants of physical activity and exercise. *Public Health Reports*, *100*(2), 158–171.

Dołżycka, N. (2017, February 18). 40+ klimatycznych polskich pubów, kawiarni i restauracji, które powinien znać każdy geek [40+ cheerful Polish pubs, cafes and restaurants that every geek should know]. *Life Geek*. https://lifegeek.pl/geek-knajpy-w-polsce/

Downward, P. (2007). Exploring the Economic Choice to Participate in Sport: Results from the 2002 General Household Survey. *International Review of Applied Economics*, *21*(5), 633–653. https://doi.org/10.1080/02692170701474710

Downward, P., & Rasciute, S. (2016). 'No man is an island entire of itself.' The hidden effect of peers on physical activity: John Donne, Meditation XVII. *Social Science & Medicine*, *169*, 149–156. https://doi.org/10.1016/j.socscimed.2016.09.038

Downward, P., & Riordan, J. (2007). Social Interactions and the Demand for Sport: An Economic Analysis. *Contemporary Economic Policy*, *25*(4), 518–537. https://doi.org/10.1111/j.1465-7287.2007.00071.x

Eime, R. M., Sawyer, N., Harvey, J. T., Casey, M. M., Westerbeek, H., & Payne, W. R. (2015). Integrating public health and sport management: Sport participation trends 2001–2010. *Sport Management Review*, *18*(2), 207–217.

Ferreira, I., van der Horst, K., Wendel-Vos, W., Kremers, S., van Lenthe, F. J., & Brug, J. (2007). Environmental correlates of physical activity in youth—A review and update. *Obesity Reviews: An Official Journal of the International Association for the Study of Obesity*, *8*(2), 129–154. https://doi.org/10.1111/j.1467-789X.2006.00264.x

Fishbein, M. (2000). The role of theory in HIV prevention. *AIDS Care*, *12*(3), 273–278. https://doi.org/10.1080/09540120050042918

Franzini, L., Taylor, W., Elliott, M. N., Cuccaro, P., Tortolero, S. R., Janice Gilliland, M., Grunbaum, J., & Schuster, M. A. (2010). Neighborhood characteristics favorable to outdoor physical activity: Disparities by socioeconomic and racial/ethnic composition. *Health & Place*, *16*(2), 267–274. https://doi.org/10.1016/j.healthplace.2009.10.009

Gallagher, N. A., Gretebeck, K. A., Robinson, J. C., Torres, E. R., Murphy, S. L., & Martyn, K. K. (2010). Neighborhood factors relevant for walking in older, urban, African American adults. *Journal of Aging and Physical Activity*, *18*(1), 99–115. https://doi.org/10.1123/japa.18.1.99

Giles-Corti, B., & Donovan, R. J. (2003). Relative Influences of Individual, Social Environmental, and Physical Environmental Correlates of Walking. *American Journal of Public Health*, *93*(9), 1583–1589.

Grima, S., Grima, A., Thalassinos, E., Seychell, S., & Jonathan. (2017). Theoretical Models for Sport Participation: Literature Review. *International Journal of Economics and Business Administration*, *V*(Issue 3), 94–116. https://doi.org/10.35808/ijeba/138

Grima, S., Grima, A., Thalassinos, E., Seychell, S., & Spiteri, J. V. (2018). A Study of the Determinants of Sports Participation by Maltese Nationals. *European Research Studies Journal*, *XXI*(2), 110–133.

Hallmann, K., Wicker, P., Breuer, C., & Schönherr, L. (2012). Understanding the importance of sport infrastructure for participation in different sports – findings from multi-level modeling. *European Sport Management Quarterly*, *12*(5), 525–544. https://doi.org/10.1080/16184742.2012.687756

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* (2nd ed.). Springer-Verlag. https://doi.org/10.1007/978-0-387-84858-7

Hawley-Hague, H., Horne, M., Campbell, M., Demack, S., Skelton, D. A., & Todd, C. (2014). Multiple Levels of Influence on Older Adults' Attendance and Adherence to Community Exercise Classes. *The Gerontologist*, *54*(4), 599–610. https://doi.org/10.1093/geront/gnt075

Hoekman, R., Breedveld, K., & Kraaykamp, G. (2017). Sport participation and the social and physical environment: Explaining differences between urban and rural areas in the Netherlands. *Leisure Studies*, *36*(3), 357–370. https://doi.org/10.1080/02614367.2016.1182201

Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2020). *party: A Laboratory for Recursive Partytioning* (1.3-4) [Computer software]. https://CRAN.R-project.org/package=party

Kohl, H. W., Craig, C. L., Lambert, E. V., Inoue, S., Alkandari, J. R., Leetongin, G., Kahlmeier, S., & Lancet Physical Activity Series Working Group. (2012). The pandemic of physical inactivity: Global action for public health. *Lancet (London, England)*, *380*(9838), 294–305. https://doi.org/10.1016/S0140-6736(12)60898-8

Kremers, S. P., de Bruijn, G.-J., Visscher, T. L., van Mechelen, W., de Vries, N. K., & Brug, J. (2006). Environmental influences on energy balance-related behaviors: A dual-process view. *International Journal of Behavioral Nutrition and Physical Activity*, *3*(1), 9. https://doi.org/10.1186/1479-5868-3-9

Lechner, M., & Downward, P. (2017). Heterogeneous sports participation and labour market outcomes in England. *Applied Economics*, *49*(4), 335–348. https://doi.org/10.1080/00036846.2016.1197369

McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An ecological perspective on health promotion programs. *Health Education Quarterly*, *15*(4), 351–377. https://doi.org/10.1177/109019818801500401

McMurdo, M. E. T., Argo, I., Crombie, I. K., Feng, Z., Sniehotta, F. F., Vadiveloo, T., Witham, M. D., & Donnan, P. T. (2012). Social, environmental and psychological factors associated with objective physical activity levels in the over 65s. *PloS One*, *7*(2), e31878. https://doi.org/10.1371/journal.pone.0031878

Muniowski, Ł. (2018). Geek Is the New Jock: The Relationship Between Geek Culture and Sports. In K. E. Lane (Ed.), *Age of the Geek: Depictions of Nerds and Geeks in Popular Media* (pp. 129–147). Springer International Publishing. https://doi.org/10.1007/978-3-319-65744-8_7

Nathan, A., Wood, L., & Giles-Corti, B. (2014). Exploring Socioecological Correlates of Active Living in Retirement Village Residents. *Journal of Aging and Physical Activity*, *22*(1), 1–15. https://doi.org/10.1123/japa.2012-0189

Prins, R. G., van Empelen, P., Te Velde, S. J., Timperio, A., van Lenthe, F. J., Tak, N. I., Crawford, D., Brug, J., & Oenema, A. (2010). Availability of sports facilities as moderator of the intention-sports participation relationship among adolescents. *Health Education Research*, *25*(3), 489–497. https://doi.org/10.1093/her/cyq024

Rowe, K., Shilbury, D., Ferkins, D. L., & Hinckson, E. (2013). *Sport development and physical activity promotion: An integrated model to enhance collaboration and understanding.* http://dx.doi.org/10.1016/j.smr.2012.12.003

Sallis, J. F., Cervero, R. B., Ascher, W., Henderson, K. A., Kraft, M. K., & Kerr, J. (2006). An ecological approach to creating active living communities. *Annual Review of Public Health*, *27*, 297–322. https://doi.org/10.1146/annurev.publhealth.27.021405.102100

Sallis, J. F., Owen, N., & Fisher, E. B. (2008). Ecological models of health behavior. In *Health behavior and health education: Theory, research, and practice, 4th ed* (pp. 465–485). Jossey-Bass.

Salvador, E. P., Florindo, A. A., Reis, R. S., & Costa, E. F. (2009). Perception of the environment and leisure-time physical activity in the elderly. *Revista de Saude Publica*, *43*(6), 972–980. https://doi.org/10.1590/s0034-89102009005000082

Skrok, Ł., Majcherek, D., Nałęcz, H., & Biernat, E. (2019). Impact of sports activity on Polish adults: Self-reported health, social capital & attitudes. *PLOS ONE*, *14*(12), e0226812. https://doi.org/10.1371/journal.pone.0226812

Statistics Poland. (2014). *[Data set] Surveys on sports facilities—Results of KFT-OB/a, KFT-OB/b, KFT-1 surveys. Data Made Available by Statistics Poland under a Data-Sharing Agreement.*

Stokols, D. (1996). Translating social ecological theory into guidelines for community health promotion. *American Journal of Health Promotion: AJHP*, *10*(4), 282–298. https://doi.org/10.4278/0890-1171-10.4.282

Strzemińska, A., Mencwel, J., & Olesińska, K. J. (2012). *„Kawiarnie Kultury. Diagnoza." Raport z badania ["Culture Cafes. Diagnosis." Research report]* (p. 28). https://stocznia.org.pl/app/uploads/2016/01/Kawiarnie_Kultury._Raport1.pdf

Teixeira, P. J., Carraça, E. V., Markland, D., Silva, M. N., & Ryan, R. M. (2012). Exercise, physical activity, and self-determination theory: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 78. https://doi.org/10.1186/1479-5868-9-78

Thiel, A., Thedinga, H. K., Barkhoff, H., Giel, K., Schweizer, O., Thiel, S., & Zipfel, S. (2018). Why are some groups physically active and others not? A contrast group analysis in leisure settings. *BMC Public Health*, *18*(1), 377. https://doi.org/10.1186/s12889-018-5283-2

Tyszka, A. (1993). *Kultura jest kultem wartości. Aksjologia spoleczna—Studia i szkice [Culture is a cult of values. Social axiology—Studies and sketches (I).]* (I). NORBERTINUM.

Van der Horst, K., Oenema, A., te Velde, S. J., & Brug, J. (2010). Do individual cognitions mediate the association of socio-cultural and physical environmental factors with adolescent sports participation? *Public Health Nutrition*, *13*(10A), 1746–1754. https://doi.org/10.1017/S1368980010002314

Van Der Horst, K., Paw, M. J. C. A., Twisk, J. W. R., & Van Mechelen, W. (2007). A brief review on correlates of physical activity and sedentariness in youth. *Medicine and Science in Sports and Exercise*, *39*(8), 1241–1250. https://doi.org/10.1249/mss.0b013e318059bf35

World Health Organization. (2018). *Global action plan on physical activity 2018-2030: More active people for a healthier world /cWorld Health Organization.*

Yu, G., Renton, A., Schmidt, E., Tobi, P., Bertotti, M., Watts, P., & Lais, S. (2011). A multilevel analysis of the association between social networks and support on leisure time physical activity: Evidence from 40 disadvantaged areas in London. *Health & Place*, *17*(5), 1023–1029. https://doi.org/10.1016/j.healthplace.2011.07.002

Corresponding information:

Received:05.08.2020.
Accepted:13.01.2021.
Correspondence to: Elżbieta Biernat
University: SGH Warsaw School of Economics, al. Niepodległości 162, 02-554 Warsaw, Poland
Faculty: Department of Tourism, Collegium of World Economy
Phone : (22) 564 68 55, 564 92 39
E-mail: elzbieta.biernat@sgh.waw.pl
University: University School of Physical Education in Wroclaw