'No man is an island entire of itself.'*
The hidden effect of peers on physical activity

* *John Donne, Meditation XVII*

Professor Paul Downward*
School of Sport, Exercise and Health Sciences
Loughborough University
Leicestershire
LE11 3TU
UK
p.downward@lboro.ac.uk
+44 1509 226365

Dr Simona Rasciute
School of Business and Economics
Loughborough University
Leicestershire
LE11 3TU
UK
s.rasciute@lboro.ac.uk
+44 1509 223620

***Corresponding Author**

Abstract

International public policy emphasises the need to increase current low levels of physical activity (WHO 2010). A large literature examines the reasons for the low levels of physical activity but tends to focus on the correlates of behaviour. This has prompted a call for more causal research to better support policy recommendations to change behaviour (Bauman et al. 2012). Using a large sample of individuals from the British Household Panel Survey (BHPS) between 1996/7 and 2006/7, a dynamic panel data analysis is employed to reveal a causal contemporaneous effect of a household peer's participation in physical activity on an individual's behaviour. The effect of a peer's physical activity on an individual's physical activity is found to be of a magnitude commensurate with the habits of the individual. An individual's participation in physical activity is also positively associated with their other leisure activity. The research suggests that an individual's physical activity takes place as part of a portfolio of household leisure, which health promotion needs to take account of.

Keywords

Physical activity, health, peer effects, panel data

1. Introduction

International policy advocates that individuals should engage in more physical activity (Department of Health, 2004; WHO 2010). The correlates of physical activity can be understood through an 'ecological' model, underpinned by a systematic review of the evidence. In this model up to 36 individual, interpersonal, environmental, regional/national and global factors are associated with differences in behaviours (Bauman et al., 2012). As far as the relationships between adult physical activity and individual characteristics are concerned, the literature suggests that ageing, being overweight, a greater perceived effort required for exercise, and occupational characteristics such as working are negatively associated with physical activity. In contrast being male, having a higher education level and a personal history of physical activity during adulthood are positively associated with physical activity. Bauman et al. (2012) also identify that the regional and national context can affect physical activity behaviour. This could be because of cultural differences between individuals in different regions or countries. It could also be because of policy variations in those localities with respect to transport planning, urban architecture and land use, such as the provision of sports facilities.

Of particular relevance for the current research is that Bauman et al. (2012) argue that there is no substantive evidence of the role of social support from friends and peers on an individual's physical activity and that there is an urgent need for more causal research to better underpin policy interventions. The current research contributes to meeting this need by investigating the causal role of peers on an individual's physical activity by providing a dynamic panel-data analysis of a large sample of individuals from the British Household Panel Survey (BHPS) between 1996/7 and 2006/7.

In Section 2 recent literature suggesting the importance of peers to physical activity is noted. Section 3 presents important methodological issues associated with the empirical identification of peer effects. The data and design of the research are presented in Section 4, with results presented and discussed in Section 5. Conclusions follow in Section 6.

2. Literature Review

Despite the claims by Bauman et al. (2012), other literature indicates the potential effect of peers on physical activity. This suggests that peer effects are lying hidden in current summaries of research findings and they are consequently not emphasised in policy recommendations.

One source of indirect evidence on the role of peers on physical activity lies in the literature investigating the impact of physical activity on either the subjective well-being (SWB) of individuals or the formation of social capital. Much of this research employs large-scale population data but is, however, correlational. Becchetti et al. (2008) identify that increases in SWB are positively associated with participation in sports as well as other social activities. This is supported by Downward and Rasciute (2011) who identify that SWB has a larger association with physical activity and sport that is undertaken with others rather than as an individual. With respect to social capital, Delaney and Kearny (2005) identify a positive association between sports participation and measures of individuals' socializing, participating in social organisations, and expressing greater trust in others. Gerlach and Brettschneider (2013) find that young people's membership of sports clubs is associated positively with greater feelings of social acceptance and respect. In a causal design, Felfe et al. (2011) identify that this is a determinant of increased levels of feeling good amongst friends. This literature highlights indirectly that physical activity is both mutually engaged in with, and enjoyed by, peers.

Another strand of literature that indirectly indicates the importance of peers to physical activity also draws upon large-scale population data, but focusses on identifying the social characteristics of an individual. Downward and Riordan (2007) identify an association between the shared characteristics of individuals and participation in sports. Farrell and Shields (2002), Downward (2007) and Humphreys and Ruseski (2015) find positive and negative associations between physical activity and marriage depending on the activity, with Wicker et al., (2009) and Ruseski et al., (2011) finding mixed effects according to the activity when there are children in the household. Finally, Rapp and Schneider (2013) identify that more formal cohabitation status, such as marriage, reduced physical activity more than cohabitation and dating relationships.

There is also some research that explores the role of peers on physical activity directly. One strand examines parental and child behaviours. Based on primary data, Seabra et al. (2007) identify associations between a child's participation and that of their parents, and particularly between female children and female parents. Boise at al. (2005) identify a positive association between a mother's and child's physical activity, whilst Bunke et al. (2013) a positive association between parents' and children's activity generally. In contrast, Downward et al. (2014) make use of a causal statistical design to identify that male parental participation in physical activity when growing up has a strong effect on their male child's, but not their female child's, participation. The effect from female parent's participation when growing up on male children is smaller, and smaller still on their female children.

Finally, other research seeks to identify causal effects through statistical or experimental designs, but this time of non-familial peers. Carrell et al. (2011) examine individuals that have been randomly assigned to peer groups (squadrons) in the United States Air Force Academy and find statistically significant positive peer effects on fitness levels. Likewise, in a field experiment upon university students, Babcock and Hartman (2010) find greater increases in gym usage for individuals who also have more friends that have responded to an incentive to use the facilities.

An overall conclusion to draw from this discussion is that there is indirect and direct evidence of the role of peers in encouraging participation in physical activity. There is a gap in the current literature, however, exploring the general contemporaneous causal effects of adult peers on physical activity. This paper contributes to filling that gap by providing evidence from a large-scale population study to better inform policy.

3. Peer Effects and their identification

There are methodological difficulties in isolating peer effects in analysis. Manski (1993) identifies three main channels by which individuals belonging to the same group might behave similarly. Exogenous and correlated effects occur when an individual adjusts their behaviour in line with characteristics of the group or environment that faces it. Much of the literature above that analyses the indirect influences of peers, or the household structure of the individual focusses on these two effects. To directly assess peer effects, however, requires identification of endogenous effects. Here the individual varies their behaviour *directly* with the behaviour of a more aggregated grouping through joint determination. Figure 1, illustrates the possibilities for peer effects for an individual '1' from individuals i, i=1,2, with behaviours 'Y_i' and personal or environmental characteristics 'X_i'.

INSERT FIGURE 1 HERE

The arrows indicate pathways between the behaviours and characteristics focussing on individual 1. Pathway 1 represents the endogenous effect, which is the direct link between the behaviour of individual '1' and the behaviour of individual '2'. Pathway 2 represents the exogenous effect in which the characteristics of individual '2', or the environment that they face, also affects the behaviour of individual '1'. Finally, Pathway 3 indicates how the shared environmental or individual characteristics of individuals '1' and '2' exert influences on their respective behaviours without there being a direct link. Revealing causal effects through direct peer influences empirically requires identification of pathway 1.

This raises three related statistical challenges. The first is to control for pathways 2 and 3, as potential confounding influences that could occur through the observable and unobservable characteristics of individuals, or the environment that they share. The second is to estimate the effects of a peer on an individual's behaviour whilst accounting for the joint determination of their behaviour as indicated in pathway 1. As Dahl et al. (2014) argue this requires either the use of instrumental variables, or through the exogenous assignment of individuals to peer groups. Manski (1993, p.352) refers to this issue as a "reflection" problem. With specific reference to the definition of the peer group, a third related statistical issue arises. This is that using average group behaviour to measure peer effects in a regression would reveal a spurious effect only, which arises from a statistical identity (Angrist, 2014).

The current research explores the direct effects of the behaviour of peers through a regression analysis of an individual's physical activity in which another household member's physical activity is used as an independent variable. However, to address the empirical concerns noted above, first, observable factors that might influence behaviour are accounted for by being included in the regression analysis. These factors include individual characteristics such as age and gender as well as relationship status, the presence of children in the household and other observable factors, as noted in the literature above. Secondly, account is taken of unobservable characteristics that might influence physical activity through selection effects. This could be the case because common characteristics that lead an individual to belong to, or remain in, a peer group influence their individual physical activity. For example, this could be a 'taste for outdoor life'. Further, in the current analysis the household is chosen as the definition of a peer group because it is exogenously identified as an administrative unit in the data. In order to control for the unobservable selection effects, two-adult households are analysed in comparison to one-adult households. In this way the problematic use of the average behaviour of a reference group is avoided. Finally, a dynamic instrumental variable analysis is undertaken to isolate the causal effects connected with the contemporaneous choices of individuals in the same two-adult household. The next section of the paper outlines the components of this strategy in more detail after highlighting details of the data.

4. Data, Measurement and Methods

4.1. Data and Measurement of Variables

Longitudinal data from the BHPS are employed in the research. The survey commenced in 1991-1992 and ran in annual waves (from 1-18) when it was merged with Understanding Society, as part of a new longitudinal study from wave 19 in 2010-2011. For reasons that will be explained below, this research only makes use of the BHPS. The BHPS involved a face-to-face interview of each member of a nationally representative sample of more than 5,000 households, making each wave comprise of approximately 10,000 cases. All adult members of the household were interviewed, including new members as they arose, with children interviewed once they reached the age of 16. Some data appear in each wave as 'core questions' with others as 'rotating questions' in some waves but not others. The dependent variable measuring physical activity used in this analysis, and also the variable that is used to measure the peer physical activity behaviour of other household members, is a rotating question. It is one of a set of questions investigating leisure-time activity. These questions commenced with wave 6 in 1996-1997 but the rotation ended before the development of the Understanding Society survey in 2010-2011. Since then the questions on physical activity have varied. Specifically, for every two years in the period from 1996-1997 to 2008-2009 the question was asked: 'how frequently do you play sport or go walking or swimming?', with responses given as: 'at least once a week', 'at least once a month', 'several times a year', 'once a year or less' and 'never/almost never'. The responses to the question were originally coded as 5 'at least once a week' to 1 'Never/almost never but were recorded in the analysis to 4 'at least once a week' to 0 'Never/almost never' so that the lower category made more numerical sense. The other measures of leisure time behaviour, with the same format of responses and the preceding clause 'how frequently you do…' include:

- Go to watch live sport
- Go to the cinema
- Go to a concert, theatre or other live performance
- Have a meal in a restaurant, cafe or pub
- Go for a drink at a pub or club

- Work in the garden
- Do DIY, home maintenance or car repairs
- Attend leisure activity groups
- Attend meetings for local groups/voluntary organisations
- Do unpaid voluntary work.

A question addressing 'Visiting friends or relations or have them visit you' appeared in the first wave of data only so this was dropped from the analysis. The responses to these questions were recoded equivalently to physical activity. Table 1 gives a summary of the responses to the question regarding physical activity across the waves.

INSERT TABLE 1 HERE

With the exception of the last wave of the data in 2008-2009, most responses indicate that over half of the sample undertakes some physical activity once a week. There is some slight growth in the incidence of activity. However approximately 20% of the sample undertakes almost no physical activity, which illustrates the current public policy concerns, particularly as this measure of physical activity will capture a broad envelope of behaviours. Table 1 also reveals that the data for wave 2008-2009 is anomalous. In private communication with the UK Data Service Support team they recognise that they '…can't find any explanation for the 2008-2009 results being so different from previous years for some categories of the RLACTA variable. As you mention, there doesn't seem to be any significant changes in the question or response options that might explain it.' The issue is being addressed now by the survey team. Consequently, in all analysis this has been removed.

Table 2 summarises all of the variables used in the analysis. This includes the physical activity of an individual and the physical activity of their household peer, as well as variables that are used to control for observed confounding effects in the analysis which, as identified in Section 1, are likely to be correlated with physical activity. Descriptive statistics for the variables are presented and, for brevity, this includes means, standard deviations, and minimum and maximum values. Because the binary variables measure categories as either '1' or '0' the means of these variables give the sample proportions of the categories measured as '1'. The standard deviations, however, have no value. The sample size is based on the observations used in the estimation (n=34,624).

This reduction in sample size from Table 1 is due to two factors. The first, and major reason, is that the analysis is restricted to a sample of households of only two adults. This means that a clearly defined peer can be identified and, as discussed further in the next subsection, selection into the household controlled for. In order to calculate the value of the individual peer behaviour, the aggregate value of household physical activity was calculated and assigned to each individual in the household. From this total, the value of the individual's own physical activity was subtracted. This left a variable which measured the value of the individual's peer physical activity. If this method were used for households with more than two members the approach would not be able to distinguish between multiple and single peers. Another reason for the reduction in size is that the statistical control for selection also relies on a binary comparison of types of household. It follows that focussing on two-adult *versus* single-adult households provides a clear set of categories when analysing if an individual has a specific peer or not. The second and related reason for the reduction in sample size is because of missing values across the variables. In the analysis only dyadic households are examined in which observations on *all* variables are available for *each* pair of adult household members.

INSERT TABLE 2 HERE

Table 2 reveals that the physical activity of all individuals in the dyadic households has a mean of 2.92. On the scale of '0' to '4' noted earlier this equates to physical activity typically being at least once a month. Because of the design of the study the socio-economic data is not entirely representative of the population as a whole. Consequently 49% of the sample is male and the average age is approximately 43 years, which exceeds typical estimates for the UK (ONS, 2015). Total monthly incomes are approximately £1,297, with the standard deviation showing an expected skew.

The average number of children in the household is less than one, which suggests that most households do not have children, but those that do have children are more likely to have one in the particular age category. The highest mean value indicates that households are more likely to have children aged between 5 and 11 years old. Approximately 44% of the sample has higher education and 72% of the households include married individuals. Approximately 28% of the sample is not married and yet lives in a household of two adults. It is important, therefore, to control for marital status as the peer effect is not necessarily equivalent to the influence of a spouse *per se*. Households are also comprised of approximately 68% of individuals who are either self-employed, or full- or part-time employed, with approximately 14% retired, 9% looking after the family or home and 2% students. A long-term illness or disability is present for 4% of the sample.

Also included are variables that measure the other leisure activity of the individual and of their peer. As Downward and Rasciute (2010) show, individuals allocate their time across a range of leisure activities and, moreover, that physical activities are often substituted for other leisure activities. Consequently it is important to control for these potential effects. For each individual the value of their other leisure activity is the sum of the scores of the ten leisure activities noted above, which are each measured on the same scales as physical activity. The calculation of the peer values for other leisure is then undertaken in the same way as for physical activity. An average value of 14.62 across the ten activities suggests that engagement in these activities is less frequent than physical activity and is typically several times a year. Most of the activities across the waves have average values less than 2. For the variables: 'Have a meal in a restaurant, cafe or pub', 'Go for a drink at a pub or club' and 'Work in the garden', average values across the waves were between 2.3 and 2.7.

4.2 Controlling for unobservables

To control for unobservable factors that might promote physical activity, it is essential to focus on those households that actually have a direct peer, as discussed earlier. Conducting the analysis across both one and two-adult households would create problems arising from the presence of two types of zeros in the dataset. When there is no peer present, peer physical activity would be scored zero. Peer physical activity would also be scored zero, however, when the peer is present but did not undertake any physical activity. To isolate a genuine peer effect, only two-adult households should be analysed. The removal of one-adult households, however, introduces the problem of unobserved selection bias. For example, as noted earlier, a 'taste for outdoor life' that is shared by individuals might be a reason for co-habitation that affects the behaviour of a two-adult household, but simply cannot influence a one-adult household. To control for these unobserved effects, which would produce sample selection bias, the inverse Mills ratio was calculated and included in all regression analysis. The inverse Mills ratio is calculated from a probit regression on the probability of belonging to a two-adult *versus* a one-adult household. It arises from the properties of a truncated normal distribution (see Greene, 2011). The same correlates as those used in the analysis of peer effects plus the additional identifying variables of owner occupation of the house and the age cohort of the individual are included in the probit regressions. The implication is that these latter variables might also influence the possibility of the type of household to which the individual belongs. The inverse Mills ratio is calculated as indicated in equation 1, where the numerator is the probability of the outcome of belonging to a two-adult household, as opposed to a one-adult household, given a set of correlates 'Z' and estimated parameters λ. $\emptyset$ is the standard normal density function and Φ the standard normal cumulative distribution.

$$InvMills = \frac{\emptyset(-Z\lambda)}{1 - \Phi(-Z\lambda)} \qquad (1)$$

The inverse Mills ratio rescales the standard normal density function to sum to one allowing for the truncation of the distribution, in this case focussing on cases of two-person households rather than also one-person households, as indicated by the denominator.

4.3 Controlling for joint determination

Given that the data type to be analysed is panel data, the following general static regression model, in equation 2, can be used as a basis for estimating the physical activity for any given individual.

$$PA_{it} = \sum \alpha_j Z_{it} + \beta_k PA_{mt} + \sum \gamma_l W_i + \delta_n Inverse\ Mills_{it} + \mu_i + \varepsilon_{it} \qquad (2)$$

Here, PA refers to participation in physical activity, and Z the observable characteristics that might moderate the participation for individual 'i' and which can vary over time. PA_{mt} is the physical activity of the peer 'm'. W, in contrast, are characteristics such as gender which do not vary over time. μ_i represents unobserved person-specific and time-invariant effects and ε_{it} is an idiosyncratic disturbance term. The inverse Mills ratio is included for the reasons indicated above.

Traditionally equation 2 can be estimated by either the pooled Ordinary Least Squares estimator (OLS) or random or fixed effects panel estimators. These estimators would, however, have some limitations. The first is that, as identified in Section 3, peer effects are intrinsically manifestations of simultaneous behaviour and consequently instrumental variable analysis needs to be undertaken in order to remove the bias that this causes in estimating the effect of the peer effects variable. The second, and related, issue is that the above models are likely to nest implicit dynamic behaviour caused by the 'habit persistence' of participation in physical activities like sport (see for example, Downward and Riordan, 2007; Downward et al., 2015). This would be manifested in the presence of serial correlation of the residuals.

This would suggest that a dynamic instrumental variable panel-data estimator is a better option for analysis. As well as controlling for simultaneous behaviour, the dynamic analysis would also isolate the purely contemporaneous impact of peer effects on behaviour as estimates are conditional on the history of the model and thus likely past influences (See, also, for example, Greene, 2011; Piper, 2015). A suitable model is presented in equation 3, which shows that lags of PA are included as regressors compared to equation 2.

$$PA_{it} = \sum \varphi_p PA_{it-p} + \sum \alpha_j Z_{it} + \beta_k PA_{mt} + \sum \gamma_l W_i + \delta_n Inverse\ Mills_{it} + \mu_i + \varepsilon_{it} \qquad (3)$$

Establishing an adequate dynamic panel data instrumental variable model is not straightforward because of the problems of identifying sets of relevant and suitable instruments. Two very useful options are 'difference' and 'system' Generalised Methods of Moments (GMM) estimators. These have been developed, as versions of fixed effects estimators, specifically for panel data analysis (Roodman, 2009). Both of these estimators can rely on instruments based on lags of variables already contained within the dataset, whose validity can be tested. The estimation of a dynamic panel model, such as equation 3, by OLS would create bias as the lagged dependent variable would be correlated with the unobserved fixed effects. Difference GMM, as with the fixed effects estimator, uses first-differences to eliminate the unobserved effects and employs lagged values of the endogenous variables as instruments. System GMM adds to this model an equation in levels. This means that the unobserved person specific characteristics remain in the data and would require an additional instrumental variable unless there is no correlation between the differences of the explanatory variables and the unobserved individual fixed effects. If this is the case the levels of the variables can be instrumented with their own first-differences.

5. Results and Discussion

Table 3 provide benchmark estimates from the fixed effects, random effects and OLS estimators. Because the dependent variable is not a strictly scaled variable, with likely impacts on the heteroscedasticity of the error term, robust standard errors are used to construct inferences, and also to control for other potential heterogeneities in the idiosyncratic disturbances. For empirical reasons the fixed effects estimator is most appropriate in the current context as indicated by the estimated Hausman test statistic [$\chi 2(27) = 381.07$ (0.000).

Overall, all three estimators indicate the importance of peer effects through the significance of the variable 'PAPeer', with a one unit change in the scale of participation of the peer increasing participation of an individual of between 0.155 and 0.224 units of the scale. The magnitude of the coefficient declines in moving towards the favoured fixed-effects estimator. The inverse Mills ratio is also insignificant in the fixed effects estimator, but it is significant for the other cases. This indicates that the latter estimators do not fully control for unobserved selection effects. The inverse Mills ratio is also negative in both the random-effects and OLS cases. This suggests that the unobservable factors selecting individuals into two-adult compared to one-adult-households reduces participation in physical activity, which suggests some mitigation against peer effects and could account for the higher peer effects coefficients in these estimators.

INSERT TABLE 3 HERE

Other variables that are included in the analysis to control for observable confounding effects have a more varied influence on physical activity. Variables that are significant in the random effects and OLS estimates, but not the fixed effects estimates, include children aged between 12 and 15 years, the age of the individual, and being married. All of these are negatively signed and could be suggestive of implicit family and peer effects constraining participation. The implication could be that greater time constraints through responsibilities can reduce participation in physical activity. There is also a negative sign between physical activity and being employed and being a student, which could also suggest greater time or income constraints. However, there is contrary evidence that the presence of children aged between 5 to 11 years is positively associated with physical activity in the fixed effects estimates compared to the random effects and OLS estimates. This could suggest a family influence that encourages participation. As indicated in Section 2, it might be that a combination of changing specific activities as children age account for the changes in influence of behaviour on adults. The fixed effects estimator also identifies an insignificant effect of having a higher education on physical activity compared to a positive significant effect in the random effects and OLS estimates. This could be because of better control for unobserved heterogeneity, for example, capturing the influence of education through employment, or because the effect of education is being masked with changes over time. The latter is measured by the wave variable. Both this and education are not significant in the fixed effects estimates, but they are both significant in the random effects and OLS estimates. In this respect, one potential difference between the fixed effect and the other estimators is that the former focusses on person-specific variations over time. The wave variable could therefore more closely account for time varying behaviour. Investment in human capital through education could thus be manifest in employment. Stronger effects of variables that vary more over time are also likely. For example, variations in the individual's age and Ch5to11 being significant as compared to other variables could be because the former naturally increases over time with the progression of waves. It is also the case that the proportion of households with children aged 5 to 11 years falls from approximately 44% to 37% over the waves, whereas all of the other variables are much more stable.

Regardless of these variations, all models, however, also identify that the other leisure activities of an individual are positively associated with their participation in physical activity, but the other leisure activity of their peers is negatively associated with their activity. This suggests that there could be household trade-offs of leisure time.

Table 4 presents the results of the dynamic instrumental variable panel-data estimator. The relevance of the dynamic analysis is indicated by a test of the serial correlation of the residuals of the preferred static fixed effects estimation. Rejection of the null hypothesis of no first order serial correlation ($F(1, 5799) = 54.255$; Prob > F= 0.000) reveals the presence of dynamic behaviour (Wooldridge, 2002; Drukker, 2003). Results are presented for the whole sample and also disaggregated for males and females as it is well-known, for example, that participation in activities like sports are gendered (Downward et al., 2015).

In the current analysis difference GMM is employed because of supportive diagnostic test results. This was not the case with the system GMM estimator. These diagnostic tests include, first, examining for the absence of first order serial correlation and, second, examining the validity of the instruments.

The difference GMM estimator directly addresses the serial correlation that will arise from estimating fixed effects in a dynamic context, as discussed earlier. However, if the idiosyncratic error term ε_{it} is itself serially correlated of order 1 then this will be correlated with the lagged differences of the dependent variable. The potential presence of such serial correlation is explored by testing the null hypothesis of no second order serial correlation in differences in ε_{it}, to check for first-order serial correlation in the levels of ε_{it}. This is because of the overlapping time periods for the levels of the residuals that are implied (Roodman, 2009). The diagnostic statistic AR(2) is used to address the null hypothesis. As can be seen from the bottom of the table, the reported p-value of a test of the null hypothesis of no second order serial correlation is accepted for all cases as the AR (2) value is not statistically significant. In the case of females the value of the statistic is less than 0.2, which is seen to be desirable. As suggested by Roodman (2009), experimentation with the specification, by including a further lag of non-physical activity leisure behaviour in the female equation, strengthens the serial correlation diagnostics but the coefficients hardly change, so the original results are presented in the table.

The Hansen J test is used to assess the overall validity of the set of instruments employed in the model. Difference-in-Hansen tests for subsets of the instruments for specific endogenous variables are also used. In the current context one of the endogenous variables is the lagged physical activity of the individual, as the physical activity behaviour of individuals overtime is likely to be related to common omitted factors such as a 'taste for outdoor life' as noted earlier. The physical activity of peers is also clearly treated as endogenous. All of the other variables in the model were treated as exogenous including one-period lags of Othleisure and OthPeer. As can be seen from the bottom of Table 4, all of the reported p-values suggest acceptance of the null hypothesis of the validity of the instruments. This stands, in contrast, to the System GMM estimator for which the null hypothesis of no second order serial correlation in differences can be rejected at a 10% level of significance and the Hansen test of suitable instruments rejected at less than 1% significance.

Table 4 reveals that the significant variables correspond to the fixed effects estimator, with the exceptions of being employed or not, being a student or not, and having a long-term illness or not, becoming insignificant. This reduction in significance of variables might be expected given the dynamic specification of the model in which lagged behaviour is used to model current behaviour. Nonetheless, there is clear evidence of a peer effect of a comparable size to the fixed effects estimator and similar patterns of the other leisure activity of household members are observed. The dynamic model also reveals that the coefficients on the lagged physical activity behaviour, capturing past habit, is much larger for males than females. Significantly however, the peer effects are generally larger than the habit effects and, in turn, both of these effects are larger for males than for females. The difference in the effects is, however, greater for females. The results also show that an individual's non-physical activity leisure can reinforce their physical activity.

Overall, the analysis confirms the presence of peer effects on physical activity and that these effects are reinforced by the individual's other leisure activity and not offset by the peer's other leisure activity. Moreover, the role of the peer is potentially of greater relevance than an individual's own habits and past behaviour, and this is particularly so for females. The portfolio of leisure activity undertaken by households is thus revealed to be an important domain in which to consider physical activity promotion, particularly with respect to females. The implication is that policy should focus on either or both the collective promotion of activity to peers and more 'selfish' behaviour of females in engaging in physical activity. This might focus on encouraging individuals to engage in activities together more than, perhaps, focussing on individual performance 'targets' and competition such as, for example, in promotions that target individuals to lose weight or to focus on improving personal times in running etc.

It should be recognised that these results may have most relevance to the UK as this is from where the data are drawn. It is clear that the results rely on individuals having access to leisure time and choice over this, which is not always the case internationally with variations in socio-economic circumstances and cultural practice. The measure of physical activity is also relatively broad and does not capture the intensity of effort. Moreover, to generate robust statistical insights two-adult based households provided the focus. Clearly, this work needs to be extended to consider additional peers and different social and cultural contexts. Nonetheless, the results do provide a unique insight into behaviour in the general population and also indicates how the study of other longitudinal data might take place.

6. Conclusions

There is now widespread public policy concern about the need to increase physical activity in society. Current research tends to draw upon the correlates of physical activity in cross-sectional samples. There is consequently a call for more innovative research designs and statistical methods to contribute to causal research. This paper makes a contribution to the literature by focussing on the generally neglected impact of peer effects on activity in large-scale population data. Based on a large sample of individuals from the British Household Panel Survey (BHPS) between 1996/7 and 2006/7, static and dynamic panel data analysis is conducted to isolate peer effects on an individual's physical activity, in an innovative design, by controlling for observable confounding influences and also unobservable factors that could lead to selection into a two-adult versus one-adult household. The preferred model is a dynamic panel data model that accounts for the endogeneity between peers. Analysis reveals a causal contemporaneous effect of a household peer's participation in physical activity on an individual's behaviour of a size commensurate with the habits of the individual for males, but greater than this for females. An individual's participation in physical activity is also positively associated with their other leisure activity. The research suggests that health promotion needs to recognise the role of physical activity as part of a portfolio of household leisure to target complementarities in behaviour, particularly for females who appear to have less habit in their behaviour and are more likely to accommodate the interests of their peers. This suggests either a need to encourage them to be more ‘selfish’ in their behaviours or to direct focus on peers to facilitate greater group-based activity.

These results are important because recent commentaries of the scope and impact of health promotion have suggested that interventions have tended to focus primarily on the individual, more than the interpersonal setting, and also on nutritional and physical activity drawing upon motivational and educational mechanisms (Sallis et al., 2012 within the school and community setting (Golden and Earp, 2012; Mozzafarian et al., 2012). The current research suggests that the household peer is of profound importance to an individual’s physical activity. Future research should investigate this further as part of an ecological analysis linking the household and its access to leisure time, with opportunities in the broader community, cultural and built environment.

References

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics, 30*, 98-108.

Babcock, P. S., & Hartman, J. L. (2010). *Networks and Workouts: Treatment Size and Status Specific Peer Effects in a Randomized Field Experiment,*

Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J., Martin, B. W., & Lancet Physical Activity Series Working Group. (2012). Correlates of physical activity: Why are some people physically active and others not? *The Lancet, 380*(9838), 258-271.

Becchetti, L., Pelloni, A., & Rossetti, F. (2008). Relational goods, sociability, and happiness. *Kyklos, 61*(3), 343-363.

Bois, J. E., Sarrazin, P. G., Brustad, R. J., Trouilloud, D. O., & Cury, F. (2005). Elementary schoolchildren's perceived competence and physical activity involvement: The influence of parents' role modelling behaviours and perceptions of their child's competence. *Psychology of Sport and Exercise, 6*(4), 381-397.

Bunke, S., Apitzsch, E., & Bäckström, M. (2013). The impact of social influence on physical activity among adolescents–a longitudinal study. *European Journal of Sport Science, 13*(1), 86-95.

Carrell, S. E., Hoekstra, M., & West, J. E. (2011). Is poor fitness contagious?: Evidence from randomly assigned friends. *Journal of Public Economics, 95*(7), 657-663.

Dahl, G. B., Løken, K. V., & Mogstad, M. (2014). Peer effects in program participation. The American Economic Review, 104(7), 2049-2074.

Delaney, L., & Kearney, E. (2005). *Sport and Social Capital in the United Kingdom: Statistical Evidence from National and International Survey Data, DCMS,*

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

Downward, P., & Rasciute, S. (2011). Does sport make you happy? an analysis of the well-being derived from sports participation. *International Review of Applied Economics, 25*(3), 331-348.

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

Downward, P., Dawson, P., & Mills, T. C. (2015). Sports participation as an investment in (subjective) health: A time series analysis of the life course. *Journal of Public Health (Oxford, England),* doi:fdv164 [pii]

Farrell, L., & 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), 165*(2), 335-348.

Felfe, A. C., Lechner, M., & Steinmayr, A. (2011). Sports and child development.

Gerlach, E., & Brettschneider, W. (2013). Aufwachsen mit sport: Befunde einer 10-jährigen längsschnittstudie zwischen kindheit und adoleszenz Meyer & Meyer Verlag.

Golden, S. D., & Earp, J. A. (2012). Social ecological approaches to individuals and their contexts: Twenty years of health education & behavior health promotion interventions. Health Education & Behavior : The Official Publication of the Society for Public Health Education, 39(3), 364-372. doi:10.1177/1090198111418634 [doi]

Greene, W. H. (2011). Econometric analysis, 7th. London, Prentice Hall

Humphreys, B. R., & Ruseski, J. E. (2015). The economic choice of participation and time spent in physical activity and sport in Canada. International Journal of Sport Finance, 10(2), 138.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. The Review of Economic Studies, 60(3), 531-542.

Mozaffarian, D., Afshin, A., Benowitz, N. L., Bittner, V., Daniels, S. R., Franch, H. A., . . . American Heart Association Council on Epidemiology and Prevention, Council on Nutrition, Physical Activity and Metabolism, Council on Clinical Cardiology, Council on Cardiovascular Disease in the Young, Council on the Kidney in Cardiovasc. (2012). Population approaches to improve diet, physical activity, and smoking habits: A scientific statement from the american heart association. *Circulation, 126*(12), 1514-1563. doi:CIR.0b013e318260a20b [pii]

ONS (2015) http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk--england-and-wales--scotland-and-northern-ireland/mid-2014/sty-ageing-of-the-uk-population.html (last retrieved 13th April, 2016)

Piper, A. T. (2015). Sliding down the U-shape? A dynamic panel investigation of the age-well-being relationship, focusing on young adults. *Social Science & Medicine, 143*, 54-61.

Rapp, I., & Schneider, B. (2013). The impacts of marriage, cohabitation and dating relationships on weekly self-reported physical activity in germany: A 19-year longitudinal study. *Social Science & Medicine, 98*, 197-203.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system gmm in stata. stata journal, 9 (1), 86-136. *For Online Publication,*

Sallis, J. F., Floyd, M. F., Rodriguez, D. A., & Saelens, B. E. (2012). Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation, 125*(5), 729-737. doi:10.1161/CIRCULATIONAHA.110.969022 [doi]

Seabra, A. F., Mendonca, D. M., Thomis, M. A., Peters, T. J., & Maia, J. A. (2008). Associations between sport participation, demographic and socio-cultural factors in portuguese children and adolescents. *European Journal of Public Health, 18*(1), 25-30. doi:ckm049 [pii]

Wicker, P., Breuer, C., & Pawlowski, T. (2009). Promoting sport for all to age-specific target groups: The impact of sport infrastructure. *European Sport Management Quarterly, 9*(2), 103-118.

Wooldridge, J. M. 2002. econometric analysis of cross section and panel data. *The MIT Press*.

WHO (2010). Global recommendations on physical activity for health, Retrieved from http://www.who.int/dietphysicalactivity/publications/9789241599979/en/index.html. (last retrieved 29th September 2016).

Figure 1: Endogenous, Exogenous and Correlational Effects

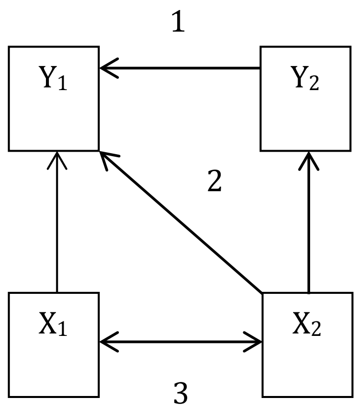

Table 1: Participation in Physical Activity (%)

Walk, swim, play sport	1996-1997	1998-1999	2000-2001	2002-2003	2004-2005	2006-2007	2008-2009	Category Total (n)
Never/almost never	19.68	23.15	23.14	19.15	19.65	18.81	82.02	27,375
Once a year or less	4.20	3.69	3.23	4.66	3.84	2.76	3.34	3,404
Several times a year	9.97	9.23	8.29	7.71	8.48	6.88	4.29	7,163
At least once a month	12.89	12.49	11.63	11.89	11.86	11.17	3.88	9,993
At least once a week	53.27	51.45	53.71	56.60	56.16	60.38	6.47	45,122
Year total (n)	9,126	10,546	15,060	15,683	14,775	14,413	1,3454	93,057

n.b. Column percentages for each year

Table 2: Variables for Analysis

Variable	Description	Mean	Std.	Minimum	Maximum
PA	Physical activity (0 = never to 4 = At least once a week)	2.92	1.49	0	4
PAPeer	Physical activity of peer (0 = never to 4 = At least once a week)	2.92	1.49	0	4
Sex	Sex of individual (1=male; 0=female)	0.49	0.50	0	1
Age	Age of Individual in years	43.11	14.57	16	78
Othleisure	Other non-physical activity leisure (Sum of 10 other activities)+	14.62	5.62	0	40
OthPeer	Other non-physical activity leisure of peer (Sum of 10 other activities)+	14.62	5.62	0	40
Income	Total income last month (£)	1,297.13	1,222.34	0	36,027.71
Ch0to2	Number of children 0 to 2 years old	0.11	0.33	0	3
Ch3to4	Number of children 3 to 4 years old	0.11	0.33	0	2
Ch5to11	Number of children 5 to 11 years old	0.37	0.71	0	6
Ch12to15	Number of children 12 to 15 years old	0.16	0.44	0	3
Ch16to18	Number of children 16 to 18 years old	0.02	0.14	0	2
HE	Higher Education (1=yes; 0 = no)	0.44	0.50	0	1
Married	Married (1=yes; 0=no)	0.72	0.45	0	1
Employed	Employed (1=yes; 0=no)	0.68	0.47	0	1
Retired	Retired (1=yes; 0=no)	0.14	0.34	0	1
Familycare	Look after the family (1=yes; 0=no)	0.09	0.28	0	1
Student	Student (1=yes; 0=no)	0.02	0.15	0	1
Longill	Have long-term illness (1=yes; 0=no)	0.04	0.20	0	1

n=34,624

+ Each activity is scored 0 to 4 as with physical activity

Table 3: Static Estimates

	Fixed Effects	Random Effects	OLS
PA_{t-1}	n/a	n/a	n/a
PAPeer	0.155***	0.192***	0.224***
	(20.01)	(31.35)	(37.84)
Invmills	-0.331	-0.545***	-0.467***
	(-1.87)	(-5.81)	(-5.57)
Sex	n/a	-0.0583*	-0.0844***
		(-2.52)	(-4.55)
Age	-0.00775	-0.00859***	-0.00881***
	(-0.27)	(-7.23)	(-8.90)
Othleisure	0.0449***	0.0687***	0.0791***
	(18.05)	(40.59)	(52.06)
OthPeer	-0.00551*	-0.00674***	-0.0120***
	(-2.35)	(-3.98)	(-7.57)
Income	-0.0000107	-0.000000151	-0.00000111
	(-1.23)	(-0.02)	(-0.16)
Ch0to2	-0.0245	-0.0322	-0.0309
	(-0.85)	(-1.45)	(-1.33)
Ch3to4	-0.0346	-0.0494*	-0.0496*
	(-1.34)	(-2.32)	(-2.18)
Ch5to11	0.0642**	0.00491	-0.000317
	(3.24)	(0.43)	(-0.03)
Ch12to15	0.0118	-0.0527**	-0.0575***
	(0.40)	(-3.05)	(-3.42)
Ch16to18	0.0548	-0.0733	-0.0688
	(0.56)	(-1.17)	(-1.09)
Wave	0.0261	0.0307***	0.0299***
	(0.47)	(6.51)	(6.34)
HE	-0.0487	0.0945***	0.0763***
	(-1.08)	(4.82)	(4.84)
Married	-0.188	-0.295***	-0.240***
	(-1.74)	(-4.84)	(-4.36)
Employed	-0.162**	-0.187***	-0.203***
	(-2.94)	(-4.49)	(-4.73)
Retired	0.0114	0.0229	0.0649
	(0.17)	(0.43)	(1.26)
Familycare	-0.0229	-0.00733	0.00811
	(-0.36)	(-0.15)	(0.16)
Student	-0.170*	0.0493	0.0646
	(-2.00)	(0.86)	(1.09)
Longill	-0.250**	-0.616***	-0.754***
	(-2.62)	(-9.20)	(-12.14)
Region dummies	Yes	Yes	Yes
R^2	0.17***	0.190***	0.191***
n	34,624	34,624	34,624
number of groups	12,784	12,784	n/a

z statistics in parentheses for Random Effects, t statistics for Fixed Effects and OLS
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Dynamic Estimates

	All	Males	Females
PA_{t-1}	0.0857***	0.114***	0.0667*
	(4.11)	(3.60)	(2.52)
PAPeer	0.153***	0.197***	0.126***
	(6.61)	(7.16)	(4.03)
Invmills	15.48	7.358	14.47
	(1.60)	(0.78)	(1.41)
Sex	n/a	n/a	n/a
Age	-0.260*	-0.245	-0.217
	(-1.99)	(-1.90)	(-1.22)
Othleisure	0.0476***	0.0518***	0.0473***
	(11.56)	(8.84)	(8.26)
OthPeer	-0.00311	-0.00841	-0.00392
	(-0.70)	(-1.61)	(-0.65)
Income	-0.000240	-0.000118	-0.000274
	(-1.74)	(-0.90)	(-1.65)
Ch0to2	0.959	0.288	1.104
	(1.57)	(0.55)	(1.52)
Ch3to4	0.308	0.0508	0.378
	(1.36)	(0.25)	(1.41)
Ch5to11	0.190*	0.0978	0.233*
	(2.48)	(1.36)	(2.30)
Ch12to15	0.666	0.273	0.735
	(1.74)	(0.88)	(1.57)
Ch16to18	3.826	1.542	4.006
	(1.67)	(0.82)	(1.49)
Wave	0.306	0.392	0.223
	(1.74)	(1.73)	(0.92)
HE	-0.438*	-0.352	-0.367
	(-2.27)	(-1.59)	(-1.80)
Married	8.874	3.856	9.055
	(1.61)	(0.79)	(1.41)
Employed	0.0203	0.0493	-0.0705
	(0.16)	(0.27)	(-0.51)
Retired	-0.236	-0.0611	-0.303
	(-1.31)	(-0.32)	(-1.22)
Familycare	-0.450	-0.587	-0.567
	(-1.70)	(-1.12)	(-1.67)
Student	-1.082	0.0412	-1.377
	(-1.50)	(0.08)	(-1.55)
Longill	-0.575*	-0.0345	-0.950*
	(-2.11)	(-0.12)	(-2.57)
Region dummies	Yes	Yes	Yes
AR(2)	0.315	0.832	0.118
Hansen's J	0.582	0.514	0.347
Difference -in-Hansen: PA_{t-1}	0.482	0.388	0.323
PAPeer	0.444	0.420	0.470
n	12,104	6,016	6,088
number of groups	5,570	2,753	2,817
number of instruments	36	36	36

z statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$