1351.0.55.028

Australian Bureau of Statistics

Research Paper

Children's Participation in Organised Sporting Activity

www.abs.gov.au

New Issue

Research Paper

Children's Participation in Organised Sporting Activity

Anil Kumar, Peter Rossiter and Alexa Olczyk

Analytical Services Branch

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) THURS 29 OCT 2009

ABS Catalogue no. 1351.0.55.028

Views expressed in this paper are those of the author(s), and do not necessarily represent those of the Australian Bureau of Statistics.
Where quoted, they should be attributed clearly to the author(s).

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Mr Anil Kumar, Analytical Services Branch on Canberra (02) 6252 5344 or email <analytical.services@abs.gov.au>.

CONTENTS

CHILDREN'S PARTICIPATION IN ORGANISED SPORTING ACTIVITY

Anil Kumar, Peter Rossiter and Alexa Olczyk
Analytical Services Branch

ABSTRACT

In 2000, 2003 and 2006, the Australian Bureau of Statistics conducted surveys on *Children's Participation in Cultural and Leisure Activities*. We analyse the data from these surveys to identify factors which may have influenced children's participation in organised sporting activities.

Initially we apply a simple age-period-cohort accounting model to the full dataset and several subpopulations of interest. This reveals evidence of strong age effects which are consistently observed, even for groups of children that report significantly different rates of participation. Between 2000 and 2006, average participation rates rose overall by more than three percentage points, but this was not uniformly observed over all subpopulations. In particular, no increase in participation was reported among children from more disadvantaged areas. No evidence of cohort effects was detected.

We then fit a logistic regression model to the data, supplementing the age, period and cohort effects with a range of observed socio-demographic characteristics pertaining to the child, the family and the neighbourhood. Significant age and period effects are confirmed, and factors such as gender, parents' employment status, country of birth and the relative socioeconomic status of the neighbourhood are found to be strongly associated with children's participation rates. Children who spend more time watching television and/or using computers are also found to be less likely to participate in organised sporting activities.

1. INTRODUCTION

Participation in organised cultural and physical activities is an important element of a child's social development. In recent years, increasing awareness of the incidence of childhood obesity has particularly highlighted the desirability, on health grounds, for children to participate in regular physical activity. Participation in organised sport, as a subset of broader physical activity, is also important for the development of motor coordination skills, teamwork and physical fitness.

In this paper, we combine data from three surveys on *Children's Participation in Cultural and Leisure Activities* (CPCLA) – conducted by the Australian Bureau of Statistics in 2000, 2003 and 2006 – to specifically examine children's participation in organised sporting activities. We examine average rates of participation for children between the ages of six and fourteen years, and identify a number of socio-demographic factors which influence the propensity for particular groups of children to participate. Despite the relatively short period spanned by the three surveys, we also find evidence of positive trends in the participation rates.

As the data for this study have been collected in three independent cross-sectional surveys, we are unable to track the participation of individual children over time. However, by aligning the age, period and cohort characteristics of the data, we are able to observe the average participation of children from the same birth-year cohort, and compare their age-specific participation rates with those of other cohorts, while allowing for changing social and cultural influences.

Following an introduction to the CPCLA Survey in Section 2, we describe the construction of our 'pseudo-longitudinal' dataset in Section 3.

In Section 4, we employ a basic age-period-cohort accounting model to our dataset. This provides insights into the relative importance of these dimensions in explaining the variation in participation rates. We observe strong differences in the age, period and cohort effects between (a) boys and girls, (b) children born in Australia and those born overseas, (c) children who live in a capital city and children from rural and regional centres, and (d) children who live in areas of differing socioeconomic disadvantage.

To obtain further insights into the factors influencing participation in organised sporting activity, we develop a logistic regression model that includes a wider range of socio-demographic characteristics. Section 5 provides background and descriptive statistics on our additional explanatory variables, and Section 6 reports the results from our model. We summarise our findings in Section 7.

2. DATA SOURCES

The data for this study were collected in three successive instances of the *Children's Participation in Cultural and Leisure Activities* (CPCLA) survey. These surveys were conducted by the Australian Bureau of Statistics in April 2000, April 2003 and April 2006 as part of the *Monthly Population Survey*.

The purpose of the CPCLA surveys is to collect data on the participation of children aged 5–14 years in selected cultural [1], sporting [2] and leisure [3] activities.

Specific data items collected in the survey include socio-demographic characteristics of the children and their parent(s), and information on the frequency and duration of the children's involvement in selected activities. For most cultural and sporting activities the collected information records participation within the preceding twelve months. For other activities (e.g. computer usage and television viewing), the survey reports the child's involvement during the two most recent school weeks. All information in the survey is provided by parents or guardians.

Sporting activities reported in the CPCLA survey are restricted to organised activities undertaken outside normal school hours. That is, the survey covers only one component of the time children may spend on physical recreation activities. Participation in informal or social sporting activities, or school-based sporting activities is not included.

The decision to focus on organised sporting activities has positive implications for the quality of the data collected. The range of activities within scope of the survey is likely to be easily understood and accurately reported by the respondent – the parent or guardian. As such activities often require personal or financial support from parents, as well as commitment by the child, it is likely that an accurate indication will be provided of the child's participation.

In the CPCLA survey, 'dancing' is classified as a cultural activity. Since dancing involves physical activity equivalent to other sporting pursuits and a high proportion of girls engage in this activity outside of school hours, we consider it appropriate to treat dancing as a substitute for sport. Therefore, in this study, we have included dancing within our definition of organised sporting activities. The effect of this decision is to raise the participation rate of girls by nine percentage points, thereby achieving a more comparable rate relative to boys.

[1] Cultural activities include music, singing, dancing and drama.

[2] The main organised sporting activities comprise: swimming, soccer, netball, tennis, basketball, Australian Rules, cricket, martial arts, athletics and gymnastics.

[3] The main leisure activities comprise: bicycle riding, skateboarding, rollerblading, playing computer games and watching television / videos / DVDs.

3. CONSTRUCTION OF THE 'PSEUDO-LONGITUDINAL' DATASET

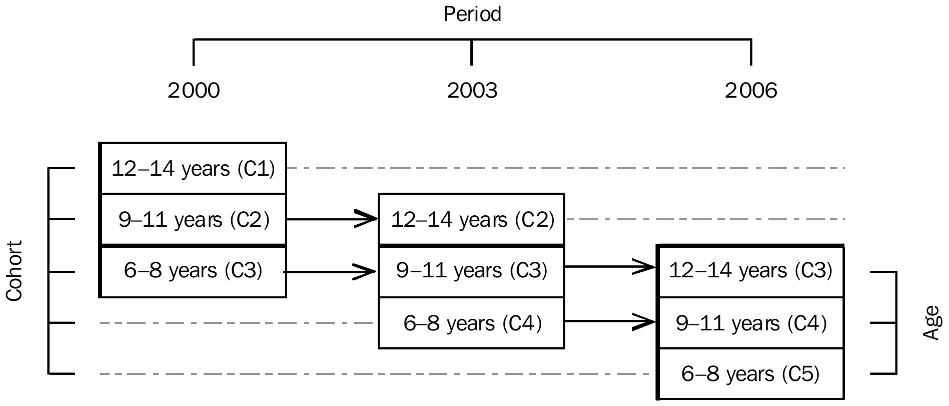
The data used in this study have been extracted from three repeated but independent cross-sectional survey datasets. This means that a consistent set of questions has been asked in all three surveys, but there is no overlap of individual respondents between surveys. Hence we cannot track changes in individual behaviour over time. We can, however, monitor changes in the average behaviour of several cohorts of children who share the same birth-year. For example, the cohort of children who were aged 6–8 years at the time of the April 2000 CPCLA survey were surveyed again in April 2003 (as 9–11 year olds) and again in April 2006 (as 12–14 year olds). It is this property of the pooled dataset that prompts us to describe the data as 'pseudo-longitudinal'.

Inherent in our definition of 'pseudo-longitudinal' data is the notion of three interrelated dimensions in the data:

- the *age* dimension reflects the age of the child at the time of their inclusion in the survey;
- the *period* dimension indicates the date on which the survey was conducted (e.g. April 2003); and
- the *cohort* dimension groups together children who share common birth-years.

Given the consistent three-year intervals between CPCLA surveys, we shall find it analytically convenient to define a conformable structure of three-year age-groups and three-year birth cohorts. This structure is illustrated in figure 3.1.

3.1 Constructing a pseudo-longitudinal dataset from cross-sectional surveys



The three columns (*periods*) of figure 3.1 correspond to the three CPCLA surveys, conducted in 2000, 2003 and 2006. Within each survey, the data are subdivided by the *age* of the child into three age-groups (6–8 years, 9–11 years, 12–14 years). The rows in figure 3.1 have been aligned to illustrate the five resulting birth-year *cohorts*.

As the CPCLA survey does not collect the actual dates of birth of the children in the sample, we can only compute approximate birth-years – by subtracting the child's age from the survey year. For example, children who were aged nine years in April 2000 have will an imputed birth-year of 1991 (which may or may not correspond with the child's actual calendar year of birth). The specification of the five cohorts, in terms of age and birth-year, is clearly shown in table 3.2, which also reports the numbers of children included in our 'pseudo-longitudinal' dataset.

3.2 Number of children surveyed, by cohort and survey year

			Survey year			
Cohort	*Age (years)*	*Birth-years*	*2000*	*2003*	*2006*	*Total*
Cohort 1	12–14 (in 2000)	1986–1988	2,772	.	.	2,772
Cohort 2	9–11 (in 2000)	1989–1991	2,977	2,635	.	5,612
Cohort 3	6–8 (in 2000)	1992–1994	2,973	2,709	2,673	8,355
Cohort 4	6–8 (in 2003)	1995–1997	.	2,640	2,689	5,329
Cohort 5	6–8 (in 2006)	1998–2000	.	.	2,493	2,493
Total			8,722	7,984	7,855	24,561

Note that our choice to align the width of the age-group definitions and cohort specifications with the three-year interval between surveys is not strictly essential. Indeed in instances where the interval between surveys is larger (say, ten years) it would be impractical to do so. However, retaining 15 cohorts (corresponding to the single birth-years 1986–2000) would not provide us with any additional insight into cohort differences, as the three single-year cohorts corresponding to each three-year cohort are always observed simultaneously (in the same survey and with the same frequency). It might also be argued, on *a priori* grounds, that any differences detected between cohorts which are separated by less than three years would necessarily be spurious.

We could perhaps have retained single years of *age* within a framework of three-year birth *cohorts*, as the age effect is expected to be strong, and significant differences in behaviour might plausibly be observed between children whose ages differ by one or two years. Primarily for the sake of exposition, however, we choose to employ three-year age-groups in our analysis.

We note that our 'pseudo-longitudinal' dataset is 'unbalanced' – that is, not all cohorts have the same number of observations over time. Only one of the birth-year cohorts is observed over all three age-groups; two cohorts are only observed in two consecutive surveys and the remaining two cohorts are only observed in a single survey. This will prove important in our later analysis when we opt to employ cohort-based weights in our logistic regression model.

4. AN AGE-PERIOD-COHORT ANALYSIS OF SPORTS PARTICIPATION

In this section we specifically examine age, period and cohort influences on children's participation in organised sporting activity. We initially analyse the full sample, and then apply the same model to selected subsets of the pseudo-longitudinal dataset.

We observe that age and period effects are very consistent, even for groups of children whose average rates of participation are very different. However we find no discernible evidence of consistent cohort effects. These observations will influence the specification of the logistic model which we develop in subsequent sections to identify socio-demographic factors that explain the differences in average participation rates.

Table 4.1(a) reports the average rates of participation in organised sporting activity, cross-tabulated by age and period. The averages in the margins (labelled 'Total') represent crude measures of the age and period effects. Table 4.1(b) reports the corresponding crude cohort effects, calculated by averaging the participation rates from the diagonals of the age × period cross-tabulation.

4.1(a) Rates of participation in organised sporting activity, by age and period

	Period (Survey year)			
Age	*2000*	*2003*	*2006*	*Total*
6–8 years	62.8	64.7	67.4	**64.9**
9–11 years	71.0	72.5	74.1	**72.5**
12–14 years	66.1	68.1	67.8	**67.3**
Total	**66.7**	**68.5**	**69.8**	**68.2**

4.1(b) Rates of participation in organised sporting activity, by cohort

Cohort	*Birth-years*	*Total*
Cohort 1	1986–1988	**66.1**
Cohort 2	1989–1991	**69.6**
Cohort 3	1992–1994	**67.5**
Cohort 4	1995–1997	**69.4**
Cohort 5	1998–2000	**67.4**

In forming participation rates, we have employed modified observation weights assigned to each individual in the pooled dataset. These weights represent a rescaled version of the original sampling weights, designed to minimise distortions arising from varying sample sizes and growth in the underlying reference population.[4]

Specifically, the original sampling weights have been adjusted so that

(a) the sum of all observation weights in the pooled dataset is equal to the average of the reference populations for the three surveys – i.e. $\frac{1}{3} \times (2{,}382{,}301 + 2{,}393{,}432 + 2{,}408{,}729) = 2{,}394{,}821$ children, and

(b) the *average* observation weights per respondent, calculated from each of the three surveys, are constrained to be equal.

Before proceeding to estimate more refined measures of the age, period and cohort effects on children's participation in organised sport, it is appropriate to briefly review these concepts:

- Age effects are essentially related to the ageing process – children become increasingly involved in sport as their physical skills and abilities increase and some will disengage as the demands of continued participation rise (in terms of required skills and commitment) or as alternative recreational or vocational activities consume more of their leisure time.
- Period effects are those temporal influences which impact on all children, irrespective of their age – for example, changes in funding may impact on the range and quality of sporting activities on offer, or growing awareness of the health consequences of childhood obesity may prompt parents to encourage their children to become more physically active. Changing economic conditions might alter the affordability of engaging in organised sporting activities.
- Cohort effects are most clearly evident between different generations of children, and are often attributed to the social, political and cultural influences which dominated their formative years. For example, we might be interested to know whether children who are born into the *Information Age* are more or less likely to engage in sport than their predecessors. If the age-specific participation rates for one cohort of children differ significantly from those of another, then a cohort effect may be present – but it is necessary to first discount the possibility that the observed difference is attributable to period effects.

The current study is constrained by the relatively short time interval spanned by the three surveys, 2000–2006. This has implications for the likely detection of these three effects.

[4] The original weights reflect the design features (e.g. strata, clusters and sample sizes) employed in the surveys. Appendix A contains a brief mathematical description of the modified weights used in the pooled dataset.

For example, we expect the age effect to be pronounced and broadly consistent over all time periods.

Between 2000 and 2006, considerable public attention was directed towards the physical well-being of children (and Australians generally), and it is plausible that public health campaigns and policy initiatives may have had a measurable impact on participation rates, producing a period effect.

The oldest and youngest children reported in the surveys were born 14–15 years apart (1985–86 and 1999–2000), and might well display cohort-specific differences in participation rates over their lifetimes. However we do not have the data to compare these cohorts age-for-age. In fact, we observe only a single cohort (Cohort 3) over all age-groups. That is, all of the age-for-age comparisons in our dataset relate to cohorts which are so close that we would not expect *a priori* to find a significant cohort effect.

In the preceding discussion we have hinted at the issue of disentangling the age, period and cohort effects from the raw data. This can be appreciated by examining the crude measures represented by the marginal rates in table 4.1. The age-specific participation rates are calculated from data from all periods, but the estimates for older age-groups are dominated by the older cohorts. Likewise all age-groups are represented in the crude period-specific participation rates, but the more recent periods are characterised by younger cohorts. As indicated earlier, only one of the five cohorts is observed over all age-groups and all periods; four out of five cohorts are thus only observed over restricted age and period ranges.

The general Age-Period-Cohort (APC) accounting model we employ in this section may be written:

$$M_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk} \tag{1}$$

where the $$M_{ijk} = \ln\left(\frac{R_{ijk}/100}{1 - R_{ijk}/100}\right)$$

are the natural logarithms of the odds of participation corresponding to the cells of the age × period (× cohort) cross-tabulation represented in table 4.1(a); the R_{ijk} are the nine participation rates reported in table 4.1(a); μ is the average log-odds pertaining to the complete target population; ε_{ijk} is the discrepancy between the observed value of M_{ijk} and the value predicted by the model; and the age, period and cohort effects (α_i, β_j and γ_k respectively) are constrained such that

$$\sum_{i=1}^{3} \alpha_i = \sum_{j=1}^{3} \beta_j = \sum_{k=1}^{5} \gamma_k = 0$$

By modelling M_{ijk} rather than directly modelling R_{ijk} we are able to avert situations where the model predicts outcomes that fall outside the range of 0–100% for R_{ijk}.

From the form of our model, it is clear that we wish to decompose participation rates into independent contributions arising from age, period and cohort effects. However, the problem with our model, as stated, is that it does not permit a unique solution – or rather there are an infinite number of solutions which satisfy the model and the associated zero-sum constraints.

An expedient solution to this problem – which is frequently used in practice – is to impose an extra constraint on the parameters. If chosen intelligently, this constraint may ultimately prove harmless, but it remains nonetheless arbitrary. A more satisfactory solution to this long-standing statistical dilemma has been proposed by Yang *et al.* (2008). Their 'intrinsic' estimator, based on the Moore–Penrose generalised inverse, can be shown to have certain optimality properties with respect to all other estimable solutions – but most importantly it restores objectivity to the analysis.

In figure 4.3, we illustrate age, period and cohort effects computed via the 'intrinsic' estimator approach. For exposition, we have combined the average underlying log-odds of participation (captured by the μ parameter) separately with the individual age (α_i), period (β_j) and cohort (γ_k) effects and applied the inverse transformation to recover fitted rates of participation.

For example, in the absence of any age, period or cohort effects, the fitted underlying participation rate would be given by

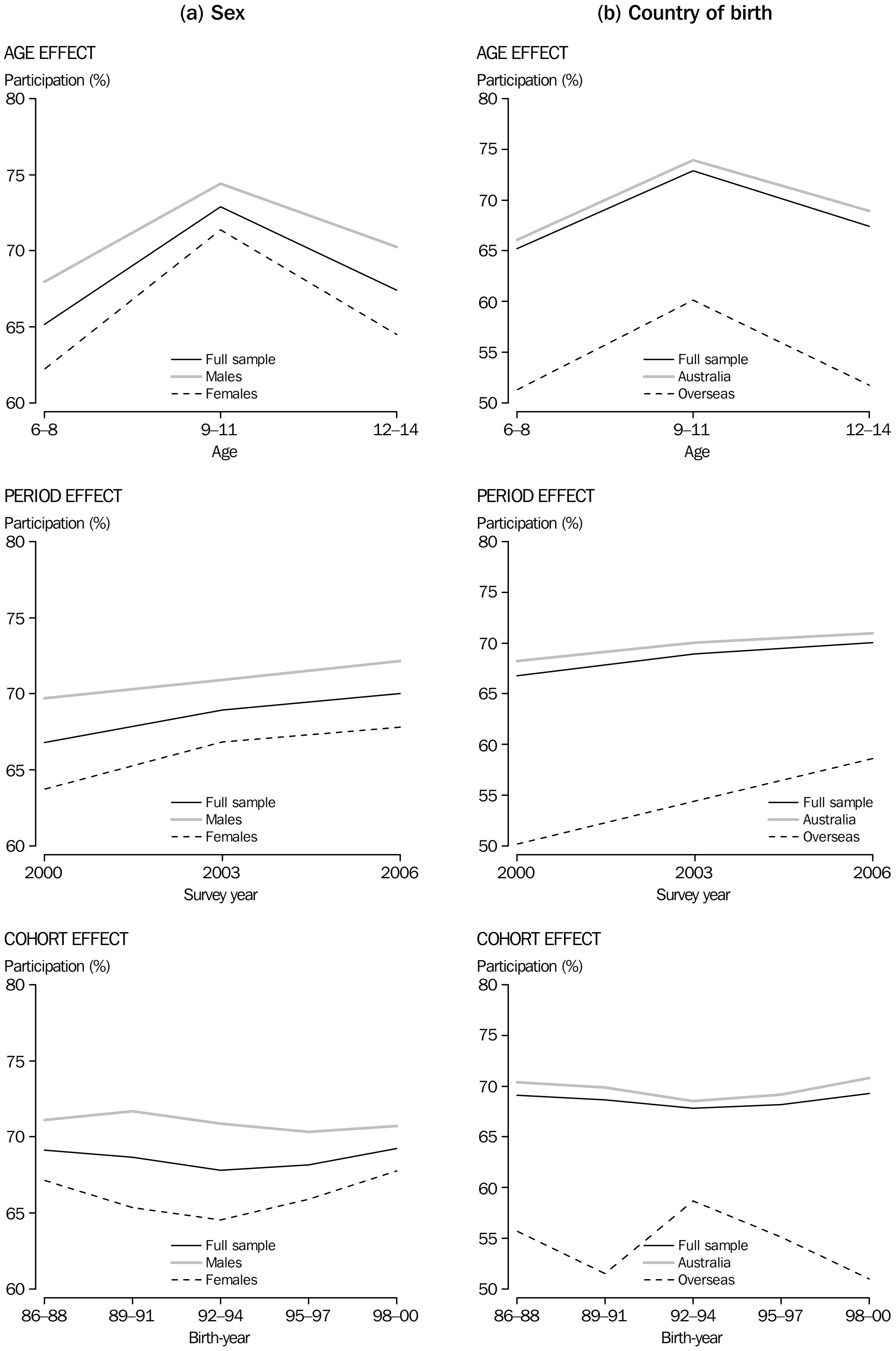
$$\frac{\exp(\mu)}{1+\exp(\mu)} \times 100\% \qquad (2)$$

Table 4.2 reports the fitted underlying participation rates for various subsets of the pseudo-longitudinal dataset. The fitted age effects graphed in figure 4.3 are obtained by substituting $(\mu + \alpha_i)$ for μ in equation (2). Likewise, the period effects are obtained by substituting $(\mu + \beta_j)$, and the cohort effects by substituting $(\mu + \gamma_k)$.

4.2 Average rates of participation estimated by the APC model

Subpopulation	*Participation rate*	*Subpopulation*	*Participation rate*
Full sample	68.6%	Lives in capital city	67.6%
Males	70.9%	Does not live in capital city	70.1%
Females	66.1%	Lives in high SEIFA area	81.4%
Australian-born	69.8%	Lives in average SEIFA area	69.1%
Overseas-born	54.4%	Lives in low SEIFA area	51.9%

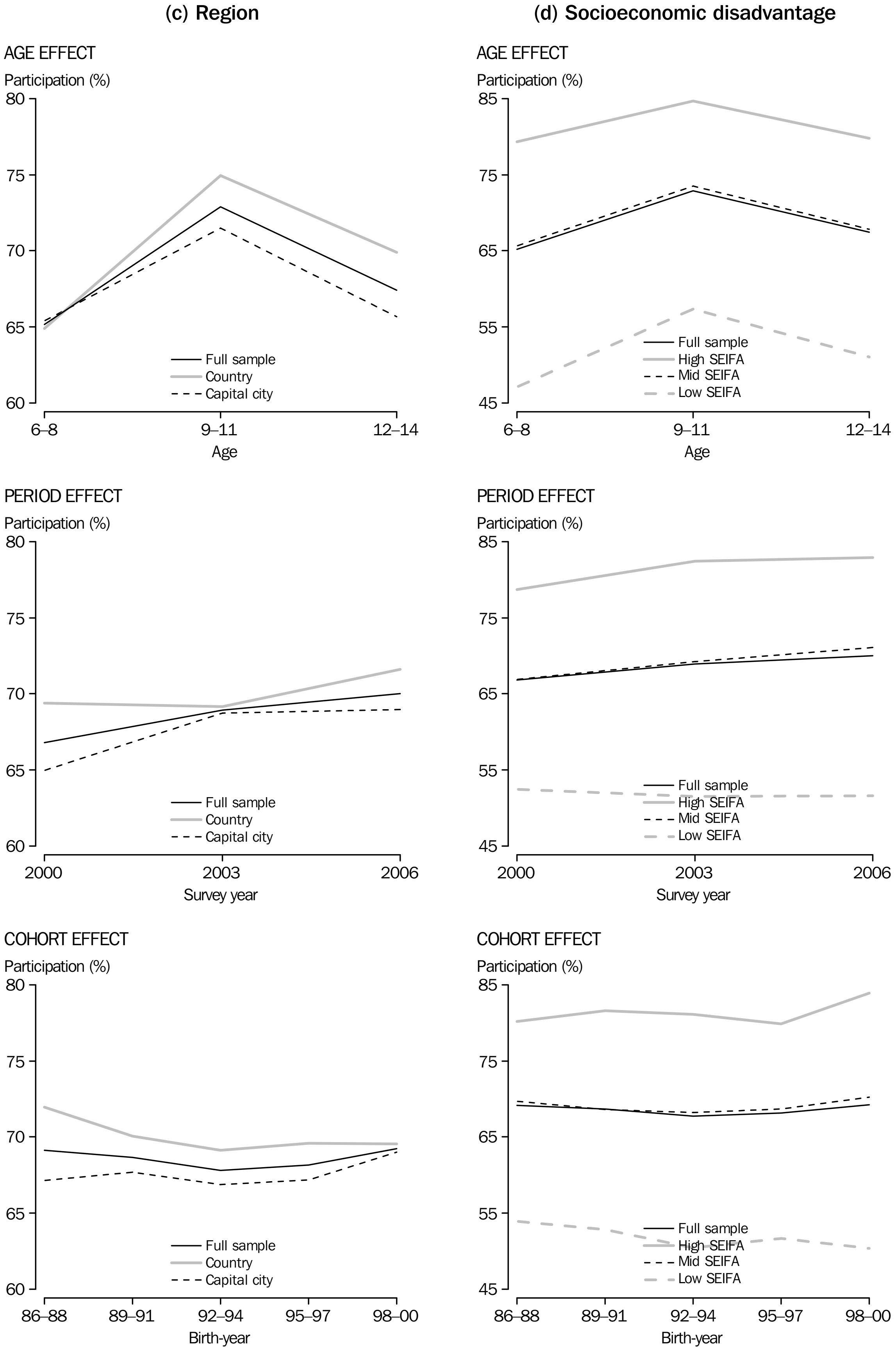
4.3 Age, period and cohort effects for specified subsets of the pseudo-longitudinal dataset



(a) Sex

(b) Country of birth

4.3 Age, period and cohort effects for specified subsets of the pseudo-longitudinal dataset – cont.



(c) Region

(d) Socioeconomic disadvantage

From table 4.2 we note that sports participation rates are slightly higher for boys than girls, and there is little difference in participation rates between children living in a capital city and children living in the country.[5] However, children born overseas are reported to have significantly lower participation rates than Australian-born children, and the relative socioeconomic status[6] of the area in which a child lives can have a substantial impact on the likelihood that they will participate in organised sport. These differences are also evident from the vertical separation of the lines plotted in figure 4.3.

For all groups of children analysed by our APC model, the age effect peaks in the 9–11 year age-group. On average, participation rises by about eight percentage points between the 6–8 and 9–11 year age-groups, and then falls by about six percentage points between the 9–11 and 12–14 year age-groups.

At age 6–8 years, city and country children are reported to have similar rates of participation in organised sport. As they grow older, children living outside of the capital cities have higher participation rates than those living in a capital city, and the gap is observed to widen with increasing age.

On average, and in all but one specific case, the period effect on participation rates is observed to rise over time. That is, participation rates are observed to be trending upwards. Children from the most socioeconomically disadvantaged areas of Australia form the notable exception.

On average, participation rates rose by 3.2 percentage points between 2000 and 2006. For girls, participation rose by 4.0 percentage points, compared with 2.5 percentage points for boys. For overseas-born children, participation rates rose by a sizeable 8.4 percentage points. Participation rates rose by more in the capital cities than outside the capital cities, with the main impact being observed earlier in the cities. Children from well-off neighbourhoods increased their participation by 4.2 percentage points between 2000 and 2006, while those from the most disadvantaged areas made no progress.

If present, cohort effects might be expected to demonstrate monotonic trends (i.e. increasing or decreasing with the birth-year of the cohort). They would also be expected to display consistency for all or most subsamples of the population. Neither feature is evident in our estimates. In addition, the magnitudes of the fitted effects are generally so small that it is doubtful whether any would pass a test of statistical significance.

[5] We use ‘country’ to loosely describe and encompass all geographical regions of Australia which lie outside the metropolitan areas of the nine capital cities.

[6] Socioeconomic status is measured by reference to the SEIFA Index of Relative Socioeconomic Disadvantage (see ABS, 2006b). ‘High SEIFA’ areas are those which form the top quintile of the distribution of the Index, while ‘Low SEIFA’ areas are those in the bottom quintile.

To summarise,

- Age effects are found to be sizeable in magnitude and consistent over all subpopulations;
- There is strong evidence that participation rates rose overall and in most sections of the population between 2000 and 2006, but the magnitude of the period effect varies from strong to weak across different subpopulations;
- No compelling evidence was found to confirm the existence of a statistically significant cohort effect; and
- Sex, country of birth, regional differences and socioeconomic status have been found individually to be significant influences on participation rates.

In Section 6 we shall build upon these observations to develop a more comprehensive model of the socio-demographic characteristics which impact on children's participation in organised sport. In that model we shall not be using the 'intrinsic' estimator to identify the age, period and cohort effects. Instead, we shall collapse the specification of the 'insignificant' cohort effect, to focus on estimating the stronger age and period effects.

5. FACTORS INFLUENCING PARTICIPATION IN SPORT

The CPCLA surveys collect a range of socio-demographic characteristics on both children and their family situations. In this section, we identify those characteristics we believe will prove most informative in explaining differences in children's sporting participation rates. In Section 6, our selected attributes will be tested within a logistic modelling framework.

Table 5.1 provides a list of selected characteristics which are available for consideration. In fact, each characteristic in the table (except for the SEIFA items) represents the dominant response for a dichotomous data field. These data fields (*Sex*, *Country of birth*, etc. will be reviewed in greater detail later in this section.

5.1 Comparison of the composition of the subpopulations of children aged 6–14 years who do and do not participate in organised sport

	Proportion of children who share this characteristic –		
Characteristic	*Children who participate in organised sport*	*Children who do not participate in organised sport*	*All children*
Child is male*	0.5327	0.4708	0.5130
Child was born in Australia	0.9414	0.8971	0.9273
Child has at least one Australian-born parent*	0.8446	0.7045	0.8001
Child lives in a capital city*	0.5820	0.6094	0.5907
Child lives in an area with low SEIFA	0.1302	0.2616	0.1719
Child lives in an area with average SEIFA*	0.6234	0.6133	0.6202
Child lives in an area with high SEIFA	0.2464	0.1251	0.2079
Child lives in a couple family*	0.8258	0.7335	0.7965
Child has at least one parent employed*	0.8904	0.7273	0.8386
Child spends ≤ 25 hours per week watching television and/or using computers*	0.5613	0.5187	0.5477

Note: Proportions have been calculated using the modified observation weights described in Section 4.

* denotes characteristics which are shared by a majority of respondents.

The primary purpose of table 5.1 is to compare and contrast the subpopulations of children who do and do not participate in organised sport. Significant compositional differences suggest that the identified characteristic(s) may influence the probability of participation in organised sport.

For example, children with at least one Australian-born parent, and children with at least one parent in employment are better represented within the subpopulation of sports participants than they are in the general population. Hence the children of these parents are apparently more likely to participate in organised sport than the children who do not have an Australian-born parent or do not have at least one parent in employment.

When modelling participation rates and odds ratios it is salutory to note that apparently significant results can sometimes be observed for rather insignificant subgroups – hence the compositional data in table 5.1 should provide a useful perspective for reviewing the results of our logistic model.

The characteristics denoted by '*' in table 5.1 are all shared by a majority of children, and the subset of the population which shares ALL of these common characteristics might be viewed as a 'typical' reference group. In fact, this subset contains only 4.6% of all children aged 6–14 years, and the participation rate for children in this group is 79.7% – considerably higher than the population average of 68.2% (see table 4.1). This suggests that there are perhaps a number of different negative influences on participation, each one affecting only a minority of children.

In the remainder of this section we shall present and discuss our selected attributes for inclusion in the logistic model. We do not presume to have full explanations for the differences in participation we observe between the groups of children who are captured by our dichotomous variables (e.g. boys vs girls, Australian-born vs overseas-born), but we provide some speculative suggestions. Without some understanding of the underlying drivers of participation it is difficult to appreciate the significance of change.

Sex

In Section 4 we clearly established that, age-for-age, boys are more likely to participate in organised sport than girls. The gap between participation rates would be even more pronounced if we had not opted to include 'dancing' within the scope of sporting activities. While some sports (e.g. swimming and athletics) cater equally to male and female participants, the majority of organised sports tend to be very gender-specific. It is probably fair to observe that the range of team sports open to female-only participation is limited compared to the male-dominated sports. This may be largely the result of an historical legacy where much of the infrastructure support for organised sport was developed for male sporting codes. Today, media coverage of women's sport remains limited, perpetuating a dearth of female sporting role models.

Country of birth

In Section 4, we observed a considerable difference (15 percentage points) in the participation rates of Australian-born children and overseas-born children. We also noted a strong upward trend in participation rates between 2000 and 2006 for overseas-born children. While these effects are certainly significant, they apply to only a very small proportion of the population. By alternatively focussing on the country of birth of the parent(s) we attempt to broaden our focus. The parents of most overseas-born children will themselves have been born overseas. Also, we might expect the Australian-born children of migrants to face similar obstacles and cultural biases to their overseas-born compatriots.

While sporting activities retain a prominent place in Australian culture, migrants and migrant communities may differ in their assessment of the value of sport in society. These differences may reflect conflicting cultural or religious views, or they may simply arise from a lack of familiarity with Australian sports and sporting institutions. Poor English language skills may hinder participation, as might a reluctance (on the part of the child or the parent) to associate with children from different cultural or ethnic backgrounds.

Capital city

We naturally assume that children living in the larger cities will have access to a wider range of sporting venues and facilities than children living in small towns or rural areas. This is undoubtedly true, but it may also be true that children living in large cities have a wider choice of alternatives to sporting activity. This may explain why children living outside the capital cities are actually more likely to participate in organised sport.

Family structure

This variable recognises the role of parents in encouraging and facilitating children's sporting activities. In a family where two parents share household and family responsibilities, it is more likely that at least one parent will be able to meet the commitments which often accompany children's engagement in organised sport – transporting or accompanying the child to training and competition, active participation in team or club affairs (e.g. coaching, attending committee meetings, fund-raising) or officiating in the conduct of the sport.

Parents' employment status

Most organised sporting activities incur financial costs. These may include the cost of specialised equipment and clothing, membership or registration fees, and travel and accommodation costs. This may be beyond the financial means of families that do not have a secure source of income.

Socioeconomic status

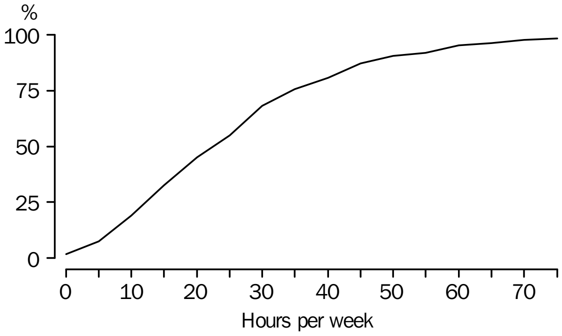
The CPCLA surveys do not collect information on the health or education of children or parents, nor do they collect family income data. The SEIFA Index of Relative Socioeconomic Disadvantage measures the relative socioeconomic status of an area by combining income, education, occupation, unemployment etc. statistics on the residents of the area. While the SEIFA index does not specifically reflect the individual circumstances of the child in the survey, it nevertheless provides an informative guide to the circumstances of most children living in the same neighbourhood. High SEIFA areas are generally characterised by high levels of income and education, and convenient access to facilities and services. Low SEIFA areas are characterised by the converse.

In Section 4 we observed that children from different socioeconomic areas differed significantly in their rates of sporting participation. We cannot tell whether this is a consequence of financial circumstances, access to facilities or 'neighbourhood' effects such as peer influences or self-esteem. Perhaps all play a role.

Television and computer usage

Perhaps the most pervasive alternative use of children's leisure time involves watching television or DVDs and using computers (including game consoles). While most children who participate in sport will also participate in these activities, children who indulge in high levels of television and computer usage are perhaps more likely to have made a lifestyle decision which does not include regular physical activity.

5.2 Cumulative distribution of television and computer usage



From figure 5.2 we observe that, for children aged 6–14 years, median usage of television and computers is approximately 25 hours per week. 25% of children are reported to have usage in excess of 35 hours per week.

We recognise that our list of attributes affecting participation in organised sport is incomplete. Other possible influences might include

- support and encouragement from schools;
- peer support and encouragement;
- the establishment of development squads to encourage and retain talented participants;
- changing fads and fashions in both sporting and non-sporting activities, e.g. skateboards; and
- the insurance risk/premium crisis that raised fees during this period.

Ideally we would include these factors in our analysis of children's sports participation, but difficulties in conceptualising and measuring these effects and often lack of data have precluded doing so.

6. MODEL-BASED ANALYSIS

In this section we present a model-based analysis of the factors associated with children's participation in organised sport. The advantage of the model-based analysis is that we are able to examine the effect of individual attributes after controlling for the effects of all other factors. We are also able to assess whether those effects are statistically significant. In addition to identifying significant factors associated with participation, we are also interested to establish whether the period effect observed in Section 4 is statistically significant – and hence evidence of the increased participation of children aged 6–14 years in organised sport.

6.1 Logistic regression model

In Section 4, we developed an age-period-cohort (APC) model to explain differences in the average participation rates reported in table 4.2. The logistic regression model developed in this section is concerned with estimating the probability, P_i, that an individual child, identified by i, participates in organised sport. For every child in our sample we have observed k categorical variables $(X_{i1},\ldots,X_{ik})$ that we believe may influence the likelihood of participation (including age, period and cohort effects). We have also observed whether or not the child does indeed participate in organised sport.

Our logistic regression model may thus be expressed as follows:

$$\ln\left[\frac{P_i}{1-P_i}\right] = \alpha + \beta_1 X_{i1} + \ldots + \beta_k X_{ik} \ ,$$

where α is an intercept parameter representing the log-odds of participation for children sharing selected *reference* characteristics, and the β's are k linear regression parameters corresponding to the k explanatory variables. The $(k+1)$ parameters of the model can be estimated using standard maximum likelihood techniques.

6.2 Modelling age, period and cohort effects

We have indicated already that we wish to include age, period and cohort variables in our model to test whether there are statistically significant effects on sports participation that are peculiar to the age group the child belongs to, the period or year in which the survey was conducted and/or the birth-year cohort to which the child belongs.

As in the APC model in Section 4, the inclusion of all three variables in the model poses a problem for estimation, since they are not independent of one other. For any given age and period, we can uniquely determine cohort. This gives rise to perfect collinearity and thus an 'identification' problem. This problem can only be resolved by imposing additional constraints on the relevant parameters. Although several methods have been proposed to resolve this problem there does not seem to be any consensus as to which if any is the most appropriate, and the identification problem remains contentious.[7]

For this analysis we shall be using the dummy variable method, as discussed in Glenn (2003). Whenever dummy variables are used in a regression model, it is always necessary to nominate one of the response categories from each variable as a reference category, which is subsequently omitted from the estimation process. The APC model requires, in addition, that one more category (from any one of the age, period or cohort variables) be dropped. This implies equating the effects of the two excluded categories from the chosen variable. While it is common to exclude adjacent categories, it is suggested that the decision be justified on the basis of theory or on what one already knows about the phenomena being studied.[8]

Our preliminary analysis in Section 4 identified evidence of strong and consistent age effects, and period effects which are consistent in direction but not necessarily strength, but no discernible evidence of cohort effects. We prefer to keep the cohort variables in our model to see whether they may have an effect on participation, after controlling for other variables in the model. However, in order to resolve the identification problem, we have chosen to redefine the cohort structure by merging cohorts 1 and 2, and cohorts 4 and 5 to form single cohorts – defined by birth-years 1986–1991 and 1995–2000, respectively. Cohort 3 (1992–1994) remains unchanged. In addition to removing the collinearity problem, this approach also produces three cohorts which are more or less equally represented (numerically) in the sample. (Previously, cohorts 1 and 5 were only observed in one survey. Now all cohorts are observed in at least two surveys.)

We have arbitrarily selected the reference categories for the age, period and cohort variables to be 6–8 years (age), 2000 (period) and birth-years 1992–1994 (cohort). (From table 4.1(a) we observe that this reference has the lowest rate of participation for all age × period cells.) This leaves six independent age, period and cohort parameters to be estimated.

[7] Some researchers have even argued that the identification problem is intrinsic to the APC data and thus cannot be resolved with statistical methods. A thorough review of the identification problem and proposed methodologies can be found in Heuer (1997), O'Brien (2000), and Glenn (2003). See also Holford (1991) and Fu (2008).

[8] If one can be confident that either age, period, or cohort has no effect on the dependent variable, then that variable can be omitted from the analysis and the age–period–cohort conundrum is resolved.

6.3 Additional model variables

In Section 5, we introduced seven additional explanatory variables for our model, and provided some rationalisations for their inclusion. We also noted in Section 5 that our list of variables is restricted to data collected by the CPCLA surveys, and hence several key determinants of sporting participation (e.g. health, education, income) are either missing or proxied.

Six of the seven selected variables are dichotomous (0–1) variables, and the seventh (Socioeconomic status) has three categorical responses. After omitting reference categories, there are eight parameters to be estimated.

In Section 4, we applied our APC model to separate subpopulations of the pseudo-longitudinal dataset, and observed some variation in the age, period and cohort effects. In the logistic regression model, we shall be estimating age, period and cohort effects for the entire dataset. Using figure 4.3 for reference, the parameter estimates for the seven additional explanatory variables will effectively determine the vertical separation of the age, period and cohort effects (for boys vs girls or Australian-born vs overseas born, etc.) – but not the shape or slope of the curve.

We could produce age, period or cohort effects for different subpopulations by including interaction terms in our regression model. However, rather than complicate the analysis unnecessarily, we have chosen to model main effects only.

6.4 Model estimation

The logistic regression model has been estimated using cohort-based weights and the estimates of the standard errors take into account the underlying survey design.

Cohort weights are derived by dividing the average population (over three years) for each gender-specific cohort by the total sample size (over three years) for that cohort and then rescaling these weights back to sum to total sample size. Where cohorts are combined the appropriate combined average population and combined total sample size are used. Using cohort-based weights ensures that each cohort has the same unchanged weight irrespective of which year or period that cohort is from.

The *Monthly Population Survey* on which the CPCLA survey is based is a stratified cluster survey. As such, information on stratification and clustering has to be used to obtain correct estimates of the standard errors. PROC SURVEYLOGISTIC in SAS has been used to estimate the model parameters as this allows for incorporation of survey design (i.e. strata and clusters) in the estimation process. While PROC LOGISTIC will give identical estimates for the coefficients, it will produce different and incorrect estimates of the standard errors, as it assumes a simple random sampling survey design. Standard errors determine whether a particular variable is significant or not under alternative methods of estimation.

6.1 Logistic regression model results

Parameter	*Estimate*	*Standard error*	*p-value*	*Odds ratio*
Intercept	0.9436	0.0529	<0.0001	
Age effects				
Aged 6–8 years*	0.0000			
Aged 9–11 years	0.4107	0.0501	<0.0001	1.508
Aged 12–14 years	0.1646	0.0796	0.0386	1.179
Period effects				
2000 Survey*	0.0000			
2003 Survey	0.0954	0.0546	0.0806	1.100
2006 Survey	0.1733	0.0807	0.0317	1.189
Cohort effects				
Cohorts 1&2 – Born 1986–1991	0.0530	0.0617	0.3901	1.054
Cohort 3 – Born 1992–1994*	0.0000			
Cohorts 4&5 – Born 1995–2000	0.0244	0.0618	0.6928	1.025
Sex				
Boys*	0.0000			
Girls	–0.2677	0.0304	<0.0001	0.765
Country of birth				
At least one parent born in Australia*	0.0000			
Both parents born overseas	–0.6548	0.0442	<0.0001	0.520
Capital city				
Living in capital city*	0.0000			
Living in rest of the state	0.0632	0.0380	0.0961	1.065
Family structure				
Couple family*	0.0000			
Single parent family	–0.0817	0.0454	0.0720	0.922
Parents' employment status				
At least one parent employed*	0.0000			
No parent(s) in employment	–0.8224	0.0496	<0.0001	0.439
Socioeconomic status				
Highest SEIFA quintile	0.6028	0.0499	<0.0001	1.827
Middle three SEIFA quintiles*	0.0000			
Lowest SEIFA quintile	–0.4986	0.0475	<0.0001	0.607
Television and computer usage				
Above average television and computer usage	–0.1688	0.0339	<0.0001	0.845
Below average television and computer usage*	0.0000			
N	24,561			
Likelihood Ratio test p-value	<.0001			
Hosmer–Lemeshow goodness-of-fit p-value	0.5811			
Max-rescaled R-squared	0.1146			
% Concordant	67.0			

* denotes the reference category for each characteristic.

Table 6.1 reports the parameter estimates for the logistic regression model, and the model diagnostic tests presented at the bottom of the table suggest a reasonably good fit to the data.

The highly significant value for the Likelihood Ratio test ($p<0.0001$) implies that including all these variables in the model gives better estimates than simply using the average to derive the estimate for the dependent variable. The Hosmer–Lemeshow goodness-of-fit test *p-value* of 0.5811 suggests the model is a 'good fit' as this value is greater than 0.05 (i.e. the 5% significance level). Although the Max-rescaled R^2 is only 0.1146, which is below the acceptable rule-of-thumb figure of 0.2, this figure is acceptable for social science research. The %Concordant statistic of 67.0% implies that slightly over two-thirds of the respondents are correctly predicted by the model as either participating or not participating in organised sport.

6.5 Model results

Table 6.1 presents the parameter estimates for the explanatory variables, and their corresponding standard errors, p-values and odds ratios.

Since all variables in our model are categorical variables, the parameter estimates and odds ratios for each variable response are expressed relative to the reference category. For example, the odds ratio for girls reports the odds of participation for girls relative to the odds of participation for boys (the reference category). An odds ratio of more than 1 indicates that children with that characteristic are more likely to participate in organised sport than children in the reference category (controlling for other variables), while odds ratios of less than 1 identify children who are less likely to participate.

The intercept term in our model is an estimate of the log-odds of participation for children who report ALL of the reference characteristics (denoted by '*'). The value of the intercept parameter equates to an estimated participation rate of 72%.[9] As for the 'typical' reference group discussed in Section 5, this is higher than the population average.[10]

The model further predicts (for example) that the odds of participation for a girl who shares the remaining reference characteristics are 0.765 times those of a boy having the same characteristics. This equates to an estimated participation rate of 66% – a difference of six percentage points.[11]

[9] The estimated odds of participation are $exp(0.9436) = 2.569$, and the estimated participation rate is therefore $100 \times \frac{2.569}{1+2.569} = 72.0\%$.

[10] In fact most reference characteristics in table 6.1 are identical to those in table 5.1. The estimated participation rate is only lowered by the selection of the APC reference categories.

[11] $100 \times \frac{0.765 \times 2.569}{1+0.765 \times 2.569} = 66.3\%$.

Of the fourteen β-parameters estimated, only the two cohort parameters proved insignificant at the 10% level. Three were significant at the 10% level only, two at the 5% level only, and seven at the 1% level.

As expected, age effects were found to be strong. Children aged 6–8 years and 12–14 years are less likely to participate in sports than those aged 9–11 years.

With respect to period effects, participation rates in 2006 were found to be significantly higher (3.4 percentage points) than participation rates in 2000, at the 5% level, with most of the increase apparently occurring between 2000 and 2003.

As already noted, the cohort parameters did not prove statistically significant at the 10% level. Furthermore, the estimates displayed no monotonicity, as might be expected if such an effect were present.

Parents' employment status proved to be the attribute associated with the largest impact on children's sports participation. The estimated reduction in participation rates for children who do not have at least one parent in employment is 19 percentage points (from 72% to 53%).

Children whose parents were born overseas are also estimated to have significantly lower participation rates (15 percentage points lower) compared to those with at least one parent born in Australia.

Children living in the most disadvantaged socioeconomic areas are predicted to have much lower rates of participation than those in areas of average socioeconomic status, while children in the most well-off areas are predicted to have participation rates significantly higher than the average.

The probability of participation in organised sport is also significantly lower (at the 1% level) for girls compared to boys.

There is strong evidence to suggest that time spent on television and computer usage is effectively a substitute for participation in organised sport. Children who spend above average amounts of time watching television and playing on computers are less likely to participate in sports than those who spend a shorter amount of time on these pastimes.

The influence of family structure was found to be weakly significant. To the extent that single parent households may be disproportionately concentrated in areas of low socioeconomic conditions, and single parents may be disproportionately represented in the labour force, some confounding of these effects might be anticipated.

The difference in participation rates between the children living in the capital cities and those living elsewhere also proved to be weakly significant. In Section 4 we observed no difference in participation rates for children aged 6–8 years, but by age 12–14 years the gap had grown to about four percentage points. This is roughly consistent with the average gap of 1.3 percentage points predicted by the logistic model, while the observed heterogeneity provides a plausible explanation for the weak significance of the estimate.

7. CONCLUSIONS

In this paper we have looked at how the data from three repeated cross-sectional surveys of children's sports participation may be pooled to construct a 'pseudo-longitudinal' dataset. The advantage of such a dataset is that it permits us to review the contributions of age, period and cohort effects in explaining the changes in participation rates reported by the three surveys.

There remain, of course, questions that can only be answered by true longitudinal data. For example, we cannot observe whether young children who play sport will continue to play sport, or what factors may prompt individual children to engage or disengage in a range of sporting or alternative activities.

We initially applied a simple age-period-cohort accounting model to the full dataset and to selected subpopulations of interest. This provided useful insights into the relative importance of the age, period and cohort effects. We found evidence of strong age effects, indicating that children's participation in organised sporting activity reaches a peak in the 9–11 years age-group, and then declines. This behaviour was consistently observed, even for groups of children that report significantly different rates of participation. We also observed a rising trend in participation rates between 2000 and 2006, but this was not uniformly observed over all subpopulations. In particular, no increase in participation was reported among children from more disadvantaged areas. No discernible evidence of consistent cohort effects was found.

To obtain further insights into the factors associated with participation in organised sporting activity, especially at the individual level, we developed a logistic regression model. In this model we supplemented the age, period and cohort effects with a range of personal, family and neighbourhood socio-demographic characteristics. Not only did this modelling framework allow us to identify the relative importance of key factors associated with children's participation in sport, it also provided a means of verifying our preliminary findings on age, period and cohort effects.

Results from the logistic regression model suggest that factors such as age, sex, parents' country of birth, parents' employment status, the relative socioeconomic status of the neighbourhood, and time allocated to television and computer usage are strongly associated with the decision to participate in organised sport.

The logistic model results suggest that participation rates rose, on average, by 3.4 percentage points between 2000 and 2006. Although we cannot ascertain the true underlying reasons for this positive trend, we can be confident that the model has accounted for a range of compositional and sample design features which might give rise to spurious period/survey effects.

The absence of a cohort effect in the model is not totally unexpected – given the short period spanned by the study and the minimal age separation of the defined cohorts. We would have liked to provide insight into whether or not the younger cohorts of children are participating less in sport than their predecessors. The limitations of our study, however, would suggest that it is premature to form a strong view on this question. The main shortcoming of the current analysis is that most cohorts have not yet been tracked over all age-groups, and hence we cannot contrast their age-specific participation rates. The inclusion of data from future surveys (e.g. the 2009 CPCLA) should lead to improvements in the estimation of cohort effects.

As a final caveat, we note that the decision to participate in organised sporting activity may be influenced by a host of other variables which are not available for inclusion in our analysis – such as school or community support, peer influences, public policy changes and the costs associated with participation in sport.

ACKNOWLEDGEMENTS

The authors would like to thank Professor Ray Chambers of the Centre for Statistical and Survey Methodology, University of Wollongong for his advice and comments, especially in relation to the weighting method and the calculation of standard errors in the modelling section of the paper. Responsibility for any errors or omissions remains solely with the authors.

REFERENCES

Australian Bureau of Statistics (2000, 2003, 2006a) *Children's Participation in Cultural and Leisure Activities, Australia*, cat. no. 4901.0, ABS, Canberra.

—— (2006b) *Information Paper: An Introduction to Socio-Economic Indexes for Areas (SEIFA)*, cat. no. 2039.0, ABS, Canberra.

Fu, W.J. (2008) "A Smoothing Cohort Model in Age Period Cohort Analysis with Applications to Homicide Arrest Rates and Lung Cancer Mortality Rates", *Sociological Methods Research*, 36, pp. 327–361.

Glenn, N.D. (2003) "Distinguishing Age, Period and Cohort Effects", in J.T. Mortimer and M.J. Shannan (eds), *Handbook of the Life Course*, pp. 465–476, Kluwer Academic/Plenum, New York.

Heuer, C. (1997) "Modelling of Time Trends and Interactions in Vital Rates using Restricted Regression Splines", *Biometrics*, 53, pp. 161–177.

Holford, T.R. (1991) "Understanding the Effects of Age, Period and Cohort on Incidence and Mortality Rates", *Annual Review of Public Health*, 12, pp. 425–457.

O'Brien, R.M. (2000) "Age Period Cohort Characteristic Models", *Social Science Research*, 29, pp. 123–139.

Yang, Y.; Schulhofer-Wohl, S.; Fu, W.J. and Land, K.C. (2008) "The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It", *American Journal of Sociology*, 113(6), pp. 1697–1736.

APPENDIX

A. MODIFIED SAMPLE WEIGHTS FOR CALCULATING PARTICIPATION RATES

In this Appendix, we provide a general mathematical derivation of the modified sample weights used in Section 4 to calculate the participation rates, which are subsequently modelled by our general Age-Period-Cohort accounting model.

Let

n_i $(i = 1,2,3)$ denote the sample sizes for the three CPCLA surveys; and

w_{ij} $(i = 1,2,3\,; j = 1,\ldots,n_i)$ denote the original sampling weights.

Define $$P_i = \sum_{j=1}^{n_i} w_{ij} \qquad (i = 1,2,3).$$

P_i is conceptually the target population for the i-th survey. This may or may not be strictly true, but it is certain that P_1, P_2 and P_3 are not identical.

Our modified observation weights take the form

$$w_{ij}^{*} = k_i w_{ij}$$

for some k_i $(i = 1,2,3)$.

If P_1, P_2 and P_3 were identical, it would be natural to define

$$k_i = \frac{n_i}{n_1 + n_2 + n_3}.$$

This would ensure the pooled sample weights sum to the desired population total, while compensating for variations in the sampling fractions employed in each survey.

Acknowledging that P_1, P_2 and P_3 are not identical, we nonetheless wish to impose a meaningful constraint upon the sum of the modified weights in the pooled sample. We choose to constrain the weights to sum to the mean of the three population totals:

$$\overline{P} = \frac{1}{3}\left(P_1 + P_2 + P_3\right).$$

That is, we choose to apply the constraint:

$$\sum_{i=1}^{3}\sum_{j=1}^{n_i} w_{ij}^{*} = \sum_{i=1}^{3} k_i \sum_{j=1}^{n_i} w_{ij} = \sum_{i=i}^{3} k_i P_i = \overline{P}.$$

While this condition constrains the choice of the scaling factors k_i $(i = 1, 2, 3)$, it does not uniquely determine them.

Perhaps the simplest solution to our imposed constraint is given by

$$k_i = \frac{1}{3} \times \frac{\overline{P}}{P_i},$$

which defines the case where the sum of the (modified) weights for each of the three surveys will equal $\overline{P}/3$.

That is, $$\sum_{j=1}^{n_1} w_{1j}^* = \sum_{j=1}^{n_2} w_{2j}^* = \sum_{j=1}^{n_3} w_{3j}^* = \frac{\overline{P}}{3}.$$

Our preferred solution is

$$k_i = \frac{n_i}{(n_1 + n_2 + n_3)} \times \frac{\overline{P}}{P_i},$$

which equates the average (modified) observation weights calculated from each survey sample:

$$\frac{1}{n_1} \sum_{j=1}^{n_1} w_{1j}^* = \frac{1}{n_2} \sum_{j=1}^{n_2} w_{2j}^* = \frac{1}{n_3} \sum_{j=1}^{n_3} w_{3j}^* .$$

Intuitively, this restores approximately equal weighting to all respondents within the pooled dataset, regardless of which survey they replied to.

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au** the ABS website is the best place for data from our publications and information about the ABS.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE	1300 135 070
EMAIL	client.services@abs.gov.au
FAX	1300 135 211
POST	Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS www.abs.gov.au

Produced by the Australian Bureau of Statistics